Network Topology Inference with Sparsity and Laplacian Constraints

Jiaxi Ying
Hong Kong University of Science and Technology
Email: jx.ying@connect.ust.hk

Xi Han*
North China University of Technology
Email: hanxi@ncut.edu.cn
* Corresponding author

Rui Zhou
Shenzhen Research Institute of Big Data
Email: rui.zhou@sribd.cn

Xiwen Wang
Hong Kong University of Science and Technology
Email: xwangew@connect.ust.hk

Hing Cheung So
City University of Hong Kong
Email: hcso@ee.cityu.edu.hk

Abstract—**We tackle the network topology inference problem by utilizing Laplacian constrained Gaussian graphical models, which recast the task as estimating a precision matrix in the form of a graph Laplacian. Recent research [1] has uncovered the limitations of the widely used ℓ_1-norm in learning sparse graphs under this model: empirically, the number of nonzero entries in the solution grows with the regularization parameter of the ℓ_1-norm; theoretically, a large regularization parameter leads to a fully connected (densest) graph. To overcome these challenges, we propose a graph Laplacian estimation method incorporating the ℓ_0-norm constraint. An efficient gradient projection algorithm is developed to solve the resulting optimization problem, characterized by sparsity and Laplacian constraints. Through numerical experiments with synthetic and financial time-series datasets, we demonstrate the effectiveness of the proposed method in network topology inference.**

Keywords-**Network topology inference; Graph Laplacian; Gradient projection; Sparsity constraint;**

I. INTRODUCTION

In modern signal processing applications, the analysis of signals residing on networks, often referred to as graph signals, has become increasingly important [2–7]. These graph signals emerge in a variety of fields, such as data gathered from wireless sensor networks and electroencephalography (EEG) signals recorded in brain connectivity networks. Laplacian constrained Gaussian graphical models (GGMs) [8, 9] provide a powerful tool for characterizing these signals on smooth graphs [10], where a substantial edge weight between two vertices indicates a high similarity in their signal values. In this paper, we tackle the network topology inference problem under Laplacian constrained GGMs, which recasts the task as estimating the precision matrix (i.e., inverse covariance matrix) as a graph Laplacian in a multivariate Gaussian distribution. The zero pattern of the precision matrix reveals the network topology, offering valuable insights into

how these sensors interact, which can be used to optimize the network's performance and reliability.

GGMs have been widely explored in the literature, with the graphical lasso [11, 12] serving as a prominent estimation method. This approach leverages an ℓ_1-norm regularized Gaussian maximum likelihood estimation, which has proven effective in imposing sparsity on the solution. With larger ℓ_1-norm regularization parameters, the solution becomes increasingly sparse. In this paper, our focus lies on Laplacian constrained GGMs, wherein the precision matrix takes the form of a graph Laplacian. Interestingly, recent studies [1, 8] have revealed that applying the ℓ_1-norm to learn Laplacian constrained GGMs results in an increased number of nonzero entries as the regularization parameter grows, yielding dense graphs rather than sparse ones. While nonconvex regularization overcomes this issue [1, 8], it necessitates tuning multiple parameters. We introduce a graph Laplacian estimation method incorporating the ℓ_0-norm constraint, which is the most intuitive and natural approach to control the sparsity of solutions.

Laplacian constrained GGMs have attracted growing interest in the fields of signal processing and machine learning over graphs [13–15]. Recent work [13] has established that the maximum likelihood estimator (MLE) under Laplacian constraints exists with as few as one observation, irrespective of the underlying dimension. This finding significantly reduces the sample size requirement from $n \geq p$ in general GGM cases, where n and p denote the sample size and problem dimension, respectively. Precision matrices in Laplacian constrained GGMs take the form of a graph Laplacian, which enables the interpretation of the eigenvalues and eigenvectors as spectral frequencies and Fourier bases [16]. Structured graph learning has been explored by leveraging spectral graph theory [9, 17–20].

Generalized Laplacian constrained GGMs have also garnered increasing attention [14, 21]. These models feature nonpositive off-diagonal entries in the precision matrix, while the zero-sum condition for rows/columns is not upheld. The

resulting precision matrix is a symmetric *M*-matrix, and such models satisfy the total positivity property [22–24], a strong form of positive dependence. The works [23, 25] have demonstrated that the MLE for these models exists if the sample size meets the condition $n \geq 2$. One approach to estimating a generalized graph Laplacian is the MLE [23, 25], which implicitly promotes sparsity through the *M*-matrix constraint. The (weighted) ℓ_1-norm regularized MLE [14, 26] provides improved sparsity control, and several algorithms have been developed to tackle it, such as block coordinate descent [14, 27], proximal point algorithm [28], and projected Newton-like methods [29]. The estimation of diagonally dominant *M*-matrices as precision matrices has been studied in [14, 30, 31].

In this paper, we investigate the network topology inference problem by estimating the precision matrix as a graph Laplacian under a sparsity constraint. It is important to note that conventional estimation methods for general GGMs, like graphical lasso, typically utilize sparsity-promoting regularization to learn sparse graphs, as the ℓ_0-constrained formulation does not yield optimal solutions when the sample size is smaller than the dimension (i.e., $n < p$). Our paper presents three main contributions:

- We propose a graph Laplacian estimation method that incorporates the ℓ_0-norm constraint, addressing the shortcomings of the ℓ_1-norm regularization when estimating Laplacian constrained GGMs. We establish that the existence of optimal solutions can be guaranteed under Laplacian constraints, even when $n = 1$.
- We devise an efficient gradient projection algorithm to solve the resulting estimation problem with sparsity and Laplacian constraints.
- We conduct numerical experiments on both synthetic and real-world datasets, demonstrating the effectiveness of our proposed method in inferring network topologies.

Notation: $\|x\|$ and $\|x\|_0$ denote Euclidean norm and the number of nonzero entries, respectively. $\mathbb{S}_+^p$ and $\mathbb{S}_{++}^p$ denote the sets of positive semi-definite and positive definite matrices with the dimensions $p \times p$, respectively. $\mathbb{R}_+^p$ represents the set of all p-dimensional vectors with non-negative real-valued components. $[p]$ denotes the set $\{1, \ldots, p\}$.

II. BACKGROUND AND PROBLEM FORMULATION

In this section, we first provide an introduction to Laplacian constrained GGMs, and then present the problem formulation.

A. Laplacian Constrained Graphical Models

We define a weighted, undirected graph $\mathcal{G} = (V, E, W)$, where V denotes the set of vertices, E represents the set of edges, and $W \in \mathbb{R}_+^{p \times p}$ is the weighted adjacency matrix

with W_{ij} denoting the graph weight between vertex i and vertex j. The graph Laplacian L is defined as:

$$L = D - W, \tag{1}$$

where D is a diagonal matrix where $D_{ii} = \sum_{j=1}^{p} W_{ij}$. Throughout this paper, we focus on connected graphs, wherein the graph comprises a single component. Following from spectral graph theory [32], the Laplacian matrix of a connected graph with p vertices has a rank of $p - 1$. Consequently, the set of Laplacian matrices for connected graphs can be formulated as:

$$\mathcal{S}_L := \big\{ X \in \mathbb{S}_+^p \mid X \cdot \mathbf{1} = \mathbf{0}, \ X = X^\top,$$
$$X_{ij} \leq 0, \forall\, i \neq j, \ \mathrm{rank}(X) = p - 1 \big\}. \tag{2}$$

where $\mathbf{0}$ and $\mathbf{1}$ denote the zero and one vectors, respectively.

Let $y = [y_1, \ldots, y_p]^\top$ be a zero-mean p-dimensional random vector following $y \sim \mathcal{N}(\mathbf{0}, \Sigma)$, where Σ is the covariance matrix. We associate the random vector y with a graph $\mathcal{G} = (V, E, W)$. As a result, y forms a GGM with respect to the graph $\mathcal{G}$. When the inverse covariance matrix, also called precision matrix, is a graph Laplacian, the random vector forms a Laplacian constrained GGM. Consequently, the problem of topology inference can be transformed into graph Laplacian estimation.

B. Problem Formulation

We consider the problem of estimating the precision matrix as a graph Laplacian, given n independent and identically distributed observations $\{y^{(k)}\}_{k=1}^n$. The maximum likelihood estimation can be formulated as the following Laplacian constrained log-determinant program:

$$\underset{X}{\text{minimize}} \quad -\log \det^\star(X) + \mathrm{tr}(SX),$$
$$\text{subject to} \quad X \in \mathcal{S}_L, \tag{3}$$

where $\det^\star$ denotes the pseudo determinant defined by the product of nonzero eigenvalues [33]. It has been demonstrated in [1] that the Laplacian set $\mathcal{S}_L$ defined in (2) can be equivalently written as

$$\mathcal{S}_L = \big\{ X \in \mathbb{R}^{p \times p} \mid (X + J) \in \mathbb{S}_{++}^p, \ X \in \mathcal{S}_Z, \big\}, \tag{4}$$

where $J = \frac{1}{p} \mathbf{1}_{p \times p}$ is a constant matrix with each element equal to $\frac{1}{p}$, and the set $\mathcal{S}_Z$ is defined as:

$$\mathcal{S}_Z := \big\{ X \in \mathbb{R}^{p \times p} \mid X \cdot \mathbf{1} = \mathbf{0}, X = X^\top, X_{ij} \leq 0, \forall\, i \neq j \big\}.$$

Then Problem (3) can be equivalently reformulated as:

$$\underset{X}{\text{minimize}} \quad -\log \det(X + J) + \mathrm{tr}(SX),$$
$$\text{subject to} \quad X + J \in \mathbb{S}_{++}^p, \ X \in \mathcal{S}_Z, \tag{5}$$

where S is the sample covariance matrix, constructed as $S = \frac{1}{n} \sum_{k=1}^{n} y^{(k)} (y^{(k)})^\top$. It is worth noting that we replace $\det^\star(X)$ with $\det(X + J)$ in (5), as done in [14], because the function $\det^\star$ is not continuous, which poses difficulty in developing algorithms.

III. PROPOSED METHOD

In this section, a new formulation incorporating sparsity and Laplacian constraints is introduced. We develop a gradient projection algorithm to solve the resulting problem.

A. Sparsity Constrained Formulation

The ℓ_1-norm regularization in graphical lasso has been widely acknowledged for its effectiveness across various fields. However, the ℓ_1-norm has been shown to be less effective in Laplacian constrained GGMs [1, 8]. Consequently, we propose a novel formulation incorporating the sparsity constraint to address this limitation.

We introduce a sparsity-constrained maximum likelihood estimation for Laplacian constrained GGMs:

$$
\begin{aligned}
& \underset{\boldsymbol{X}}{\text{minimize}} && -\log\det(\boldsymbol{X}+\boldsymbol{J}) + \text{tr}(\boldsymbol{SX}), \\
& \text{subject to} && \boldsymbol{X}+\boldsymbol{J} \in \mathbb{S}_{++}^p, \boldsymbol{X} \in \mathcal{S}_Z, \|\boldsymbol{X}\|_{0,\text{off}} \le 2s,
\end{aligned}
\tag{6}
$$

where $\|\boldsymbol{X}\|_{0,\text{off}}$ denotes the number of nonzero off-diagonal entries in $\boldsymbol{X}$. The sparsity constraint $\|\boldsymbol{X}\|_{0,\text{off}} \le 2s$ ensures that the learned graph is sparse, with the number of edges not exceeding a predetermined value s. The sparsity level s can be estimated in certain tasks. For instance, for learning structured graphs with p vertices, a connected tree graph has $p-1$ edges, while a circular graph has p edges.

We highlight that traditional estimation methods for general GGMs, such as graphical lasso, often employ sparsity-promoting regularization rather than the ℓ_0-norm constraint. This is because the ℓ_0-norm constrained formulation fails to provide optimal solutions when the number of samples is less than the dimension (i.e., $n < p$).

Theorem 1. *The set of global minimizers of Problem* (6) *is nonempty and compact almost surely as long as the number of observations $n \ge 1$.*

Theorem 1 establishes that the set of optimal solutions for the ℓ_0-norm constrained problem in (6) is guaranteed to be nonempty almost surely, even when $n = 1$. This crucial insight forms the basis of our proposed ℓ_0-norm approach. It is worth noting that the existence of optimal solutions for Problem (6) is implicitly assumed in the remainder of this paper. As demonstrated in Theorem 1, this assumption holds with probability one.

B. Gradient Projection Algorithm

Problem (6) is a nonconvex optimization problem with multiple constraints, where the constraints $X_{ij} = X_{ji}$ and $\boldsymbol{X} \cdot \mathbf{1} = \mathbf{0}$ are linear, resulting in only $p(p-1)/2$ variables in $\boldsymbol{X}$ being free. To address these constraints, we employ a linear operator, as defined in [9], which maps a vector $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ to a matrix $\mathcal{L}\boldsymbol{x} \in \mathbb{R}^{p \times p}$.

Definition 2. 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

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

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

The adjoint operator $\mathcal{L}^*$ of $\mathcal{L}$ is defined to fulfill the condition $\langle \mathcal{L}\boldsymbol{x}, \boldsymbol{Y} \rangle = \langle \boldsymbol{x}, \mathcal{L}^*\boldsymbol{Y} \rangle$, for all $\boldsymbol{x} \in \mathbb{R}^{p(p-1)/2}$ and $\boldsymbol{Y} \in \mathbb{R}^{p \times p}$. Furthermore, it is a linear operator that maps a matrix $\boldsymbol{Y} \in \mathbb{R}^{p \times p}$ back to a vector $\mathcal{L}^*\boldsymbol{Y} \in \mathbb{R}^{p(p-1)/2}$.

Definition 3. The adjoint operator $\mathcal{L}^* : \mathbb{R}^{p \times p} \to \mathbb{R}^{p(p-1)/2}$, $\boldsymbol{Y} \mapsto \mathcal{L}^*\boldsymbol{Y}$, is defined by

$$
[\mathcal{L}^*\boldsymbol{Y}]_k = Y_{ii} - Y_{ij} - Y_{ji} + Y_{jj},
\tag{8}
$$

where $i, j \in [p]$ obeying $k = i - j + \frac{j-1}{2}(2p - j)$ and $i > j$.

For notational simplicity, we represent $p(p-1)/2$ as d. By incorporating the linear operator $\mathcal{L}$, we can effectively simplify Problem (6) as:

$$
\begin{aligned}
& \underset{\boldsymbol{x}}{\text{minimize}} && -\log\det(\mathcal{L}\boldsymbol{x}+\boldsymbol{J}) + \text{tr}(\boldsymbol{S}\mathcal{L}\boldsymbol{x}), \\
& \text{subject to} && \boldsymbol{x} \in \Omega_s \cap \mathbb{R}_+^d \cap \mathbb{V}_{++}^d,
\end{aligned}
\tag{9}
$$

where the sets Ω_s and $\mathbb{V}_{++}^d$ are defined as:

$$
\Omega_s := \big\{ \boldsymbol{x} \in \mathbb{R}^d \,|\, \|\boldsymbol{x}\|_0 \le s \big\},
$$

and

$$
\mathbb{V}_{++}^d := \big\{ \boldsymbol{x} \in \mathbb{R}^d \,|\, \mathcal{L}\boldsymbol{x} + \boldsymbol{J} \in \mathbb{S}_{++}^p \big\}.
$$

Both the sets Ω_s and $\mathbb{R}_+^p$ are closed and can be addressed using a projection $\mathcal{P}_{\Omega_s \cap \mathbb{R}_+^p}$ onto their intersection with respect to the Euclidean norm. The projection $\mathcal{P}_{\Omega_s \cap \mathbb{R}_+^p}(\boldsymbol{z})$ can be computed efficiently by sorting the p entries of $\mathcal{P}_{\mathbb{R}_+^p}(\boldsymbol{z})$ and retaining only the s largest values, while setting the remaining ones to zero. Since $\mathbb{V}_{++}$ is not closed, we employ a backtracking line search method to tackle this constraint.

Let f represent the objective function of Problem (9). A gradient projection step at $\boldsymbol{x}_k$ can be constructed as:

$$
\boldsymbol{x}_k(\eta_k) \in \mathcal{P}_{\mathbb{R}_+^p \cap \Omega_s}(\boldsymbol{x}_k - \eta_k \nabla f(\boldsymbol{x}_k)),
\tag{10}
$$

where η_k is the step size, and $\nabla f(\boldsymbol{x}_k)$ denotes the gradient of f at $\boldsymbol{x}_k$:

$$
\nabla f(\boldsymbol{x}_k) = \mathcal{L}^*\big(-(\mathcal{L}\boldsymbol{x}_k + \boldsymbol{J})^{-1} + \boldsymbol{S}\big).
\tag{11}
$$

Define the gradient mapping at $\boldsymbol{x}_k$ as follows:

$$
G_{\frac{1}{\eta_k}}(\boldsymbol{x}_k) := \frac{1}{\eta_k}(\boldsymbol{x}_k - \boldsymbol{x}_k(\eta_k)).
\tag{12}
$$

We note that, in the unconstrained case, $G_{\frac{1}{\eta_k}}(\boldsymbol{x}_k)$ simplifies to $\nabla f(\boldsymbol{x}_k)$. As such, the gradient mapping can be considered an extension of the standard gradient operation.

We determine the step size using an Armijo-like rule, ensuring global convergence for our algorithm. Specifically, we examine step sizes $\eta_k \in \sigma \left\{\beta^0, \beta^1, \beta^2, \ldots\right\}$, where $\sigma > 0$ and $\beta \in (0, 1)$. We seek the smallest integer $m \in \mathbb{N}$ such that the iterate $\boldsymbol{x}_k(\eta_k) \in \mathbb{V}_{++}$ with $\eta_k = \sigma\beta^m$, and leads to a sufficient reduction of the objective function value:

$$f\left(\boldsymbol{x}_k(\eta_k)\right) \leq f\left(\boldsymbol{x}_k\right) - \alpha\eta_k\big\|G_{\frac{1}{\eta_k}}(\boldsymbol{x}_k)\big\|^2, \qquad (13)$$

where $\alpha \in (0, 1)$. The backtracking line search condition (13) is a variant of the Armijo rule.

To ensure $\boldsymbol{x}_k(\eta_k) \in \mathbb{V}_{++}$, we need to verify the positive definiteness of $\mathcal{L}\boldsymbol{x}_k(\eta_k) + \boldsymbol{J}$. This verification can be conducted during the computation of the Cholesky factorization for objective function evaluation. A summary of our algorithm is provided in Algorithm 1. Furthermore, our method is adaptable for estimating various structured matrices, including Hankel matrices, through the application of the Hankel linear operator [34–36].

Algorithm 1 Proposed algorithm

1: **Input:** Sample covariance matrix $\boldsymbol{S}$, sparsity level s, $\sigma > 0$, $\alpha \in (0, 1)$, and $\beta \in (0, 1)$;

2: **for** $k = 0, 1, 2, \ldots$ **do**

3: $\quad m \leftarrow 0$;

4: $\quad$ **repeat**

5: $\qquad$ Update $\boldsymbol{x}_{k+1} \in \mathcal{P}_{\Omega_s \cap \mathbb{R}_+^p}\left(\boldsymbol{x}_k - \sigma\beta^m \nabla f\left(\boldsymbol{x}_k\right)\right)$,

6: $\qquad m \leftarrow m + 1$;

7: $\quad$ **until** $\boldsymbol{x}_{k+1} \in \mathbb{V}_{++}$, and

$$f\left(\boldsymbol{x}_{k+1}\right) \leq f\left(\boldsymbol{x}_k\right) - \alpha\sigma\beta^m\big\|G_{\frac{1}{\sigma\beta^m}}(\boldsymbol{x}_k)\big\|^2;$$

8: **end for**

IV. EXPERIMENTAL RESULTS

We conduct numerical experiments on synthetic and real-world data to verify the performance of the proposed method.

A. Synthetic Data

We carry out numerical simulations to evaluate the estimation performance of the GLE-ADMM [15], NGL [8], and our proposed method. The GLE-ADMM and NGL methods use the ℓ_1-norm and MCP penalties to estimate Laplacian-constrained precision matrices, respectively. In contrast, our method employs the ℓ_0-norm constraint.

We randomly generate an Erdos-Renyi graph, a well-known random graph model, to serve as the underlying ground-truth graph. The graph contains 100 vertices, and the graph weights associated with the edges are uniformly sampled from $U(2, 5)$. We independently draw samples $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(n)}$ under the Laplacian-constrained GGM.
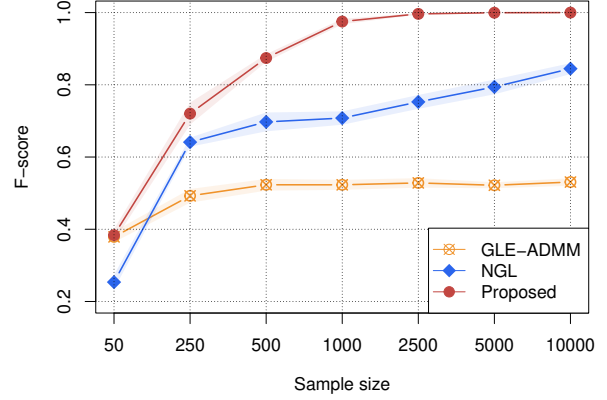

Figure 1: F-score as a function of sample size for learning Erdos-Renyi graphs consisting of 100 nodes.

To evaluate the performance of edge recovery, we utilize the F-score (FS) metric, which is defined as:

$$\mathsf{FS} = \frac{2\mathsf{tp}}{2\mathsf{tp} + \mathsf{fp} + \mathsf{fn}}, \qquad (14)$$

where tp, fp, and fn represents true positives, false positives, and false negatives, respectively. The F-score values range between 0 and 1, with a score of 1 indicating perfect identification of all connections and non-connections in the ground-truth graphs. The curves displayed in Figure 1 are the average results of 10 Monte Carlo simulations.

Figure 1 demonstrates that our proposed method achieves an F-score of 1, indicating the correct identification of all graph edges when the sample size is sufficiently large. Moreover, our method generally requires fewer samples than NGL and GLE-ADMM to attain a specific F-score value.

B. Real-world Data

We perform numerical experiments on a financial time-series dataset to evaluate the effectiveness of edge recovery. The dataset comprises 485 stocks that constitute the S&P 500 index, with data spanning from January 5, 2016, to July 1, 2020. This results in 1142 observations per stock. We construct the log-returns data matrix as:

$$X_{ij} = \log \pi_{i,j} - \log \pi_{i-1,j},$$

where $\pi_{i,j}$ denotes the closing price of the j-th stock on the i-th day. The stocks are categorized into 11 sectors by the global industry classification standard (GICS) system.

To measure the performance of edge recovery for financial time-series data, we employ the *modularity* metric [37]. A stock graph with high *modularity* exhibits dense connections among stocks belonging to the same sector and sparse connections between stocks from different sectors. Thus, a higher *modularity* value suggests a more accurate representation of the actual stock network.
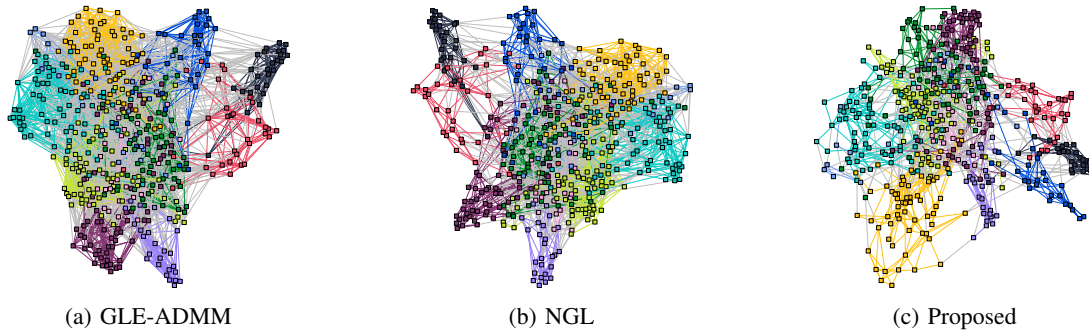
(a) GLE-ADMM   (b) NGL   (c) Proposed

Figure 2: Stock graphs learned via (a) GLE-ADMM, (b) NGL, and (c) Proposed method. The *modularity* values for GLE-ADMM, NGL, and our proposed method are 0.4253, 0.4956, and 0.6207, respectively.

Figure 2 presents the outstanding performance of our proposed method in edge recovery when compared to GLE-ADMM and NGL. This superiority is demonstrated by the fact that our method's graph predominantly features connections between vertices within the same sector, while maintaining a minimal number of connections (represented by gray-colored edges) between vertices across different sectors. With *modularity* values of 0.4253, 0.4956, and 0.6207 for GLE-ADMM, NGL, and our method, respectively, our approach showcases improved interpretability and more precise edge recovery than the competing techniques.

V. CONCLUSIONS

In this paper, we have proposed a novel formulation incorporating sparsity and Laplacian constraints for inferring network topology. Our approach addresses the limitations of the ℓ_1-norm regularization in Laplacian constrained Gaussian graphical models while effectively promoting sparsity. We have introduced an efficient gradient projection algorithm to solve the resulting problem. The efficacy of our method has been demonstrated through numerical experiments on both synthetic and financial time-series datasets.

REFERENCES

[1] J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "Nonconvex sparse graph learning under Laplacian constrained graphical model," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 7101–7113.

[2] 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*, vol. 30, no. 3, pp. 83–98, 2013.

[3] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[4] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.

[5] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[6] A. G. Marques, S. Segarra, and G. Mateos, "Signal processing on directed graphs: The role of edge directionality when processing and learning from network data," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 99–116, 2020.

[7] G. Leus, A. G. Marques, J. M. Moura, A. Ortega, and D. I. Shuman, "Graph signal processing: History, development, impact, and outlook," *IEEE Signal Processing Magazine*, vol. 40, no. 4, pp. 49–60, 2023.

[8] 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*, 2020.

[9] S. Kumar, J. Ying, J. V. d. M. 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.

[10] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[11] 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*, vol. 9, pp. 485–516, 2008.

[12] A. d'Aspremont, O. Banerjee, and L. El Ghaoui, "First-order methods for sparse covariance selection," *SIAM*

Journal on Matrix Analysis and Applications, vol. 30, no. 1, pp. 56–66, 2008.

[13] J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "Minimax estimation of Laplacian constrained precision matrices," in *International Conference on Artificial Intelligence and Statistics*, 2021, pp. 3736–3744.

[14] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under Laplacian and structural constrints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

[15] 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*, vol. 67, no. 16, pp. 4231–4244, 2019.

[16] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

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

[18] ——, "Learning bipartite graphs: Heavy tails and multiple components," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 14 044–14 057.

[19] S. Kumar, J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "Bipartite structured Gaussian graphical modeling via adjacency spectral priors," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 322–326.

[20] J. V. d. M. Cardoso, J. Ying, and D. P. Palomar, "Nonconvex graph learning: sparsity, heavy tails, and clustering," in *Signal Processing and Machine Learning Theory*. Elsevier, 2024, pp. 1049–1072.

[21] Y. Yuan, D. W. Soh, K. Guo, Z. Xiong, and T. Q. S. Quek, "Joint network topology inference via structural fusion regularization," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–18, 2023.

[22] S. Fallat, S. Lauritzen, K. Sadeghi, C. Uhler, N. Wermuth, and P. Zwiernik, "Total positivity in Markov structures," *The Annals of Statistics*, vol. 45, no. 3, pp. 1152–1184, 2017.

[23] S. Lauritzen, C. Uhler, and P. Zwiernik, "Maximum likelihood estimation in Gaussian models under total positivity," *The Annals of Statistics*, vol. 47, no. 4, pp. 1835–1863, 2019.

[24] D. Rossell and P. Zwiernik, "Dependence in elliptical partial correlation graphs," *Electronic Journal of Statistics*, vol. 15, no. 2, pp. 4236–4263, 2021.

[25] 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*, vol. 473, pp. 145–179, 2015.

[26] J. Ying, J. V. d. M. Cardoso, and D. P. Palomar, "A fast algorithm for graph learning under attractive Gaussian Markov random fields," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021, pp. 1520–1524.

[27] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 6350–6354.

[28] Z. Deng and A. M.-C. So, "A fast proximal point algorithm for generalized graph Laplacian learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5425–5429.

[29] J.-F. Cai, J. V. D. M. Cardoso, D. P. Palomar, and J. Ying, "Fast projected Newton-like method for precision matrix estimation under total positivity," *arXiv preprint arXiv:2112.01939*, 2021.

[30] J. Ying, J. V. D. M. Cardoso, and D. P. Palomar, "Adaptive estimation of graphical models under total positivity," in *International Conference on Machine Learning*, 2023, pp. 40 054–40 074.

[31] M. Truell, J.-C. Hütter, C. Squires, P. Zwiernik, and C. Uhler, "Maximum likelihood estimation for brownian motion tree models based on one sample," *arXiv preprint arXiv:2112.00816*, 2021.

[32] F. R. Chung, *Spectral graph theory*. American Mathematical Soc., 1997, no. 92.

[33] A. Holbrook, "Differentiating the pseudo determinant," *Linear Algebra and its Applications*, vol. 548, pp. 293 – 304, 2018.

[34] J. Ying, H. Lu, Q. Wei, J.-F. Cai, D. Guo, J. Wu, Z. Chen, and X. Qu, "Hankel matrix nuclear norm regularized tensor completion for n-dimensional exponential signals," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3702–3717, 2017.

[35] 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*, vol. 66, no. 21, pp. 5520–5533, 2018.

[36] 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*, vol. 41, no. 2, pp. 470–490, 2016.

[37] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.