ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10448021

JOINT SIGNAL RECOVERY AND GRAPH LEARNING FROM INCOMPLETE TIME-SERIES

Amirhossein Javaheri ⋆† *Arash Amini* ⋆ *Farokh Marvasti* ⋆ *Daniel P. Palomar* †

⋆ Sharif University of Technology
† Hong Kong University of Science and Technology

ABSTRACT

Learning a graph from data is the key to taking advantage of graph signal processing tools. Most of the conventional algorithms for graph learning require complete data statistics, which might not be available in some scenarios. In this work, we aim to learn a graph from incomplete time-series observations. From another viewpoint, we consider the problem of semi-blind recovery of time-varying graph signals where the underlying graph model is unknown. We propose an algorithm based on the method of block successive upperbound minimization (BSUM), for simultaneous inference of the signal and the graph from incomplete data. Simulation results on synthetic and real time-series demonstrate the performance of the proposed method for graph learning and signal recovery.

Index Terms— Graph signal, graph learning, incomplete data, missing sample recovery, time-series.

1. INTRODUCTION

Graph signal processing is an emerging field of research with numerous applications in social networks [1], neural networks [2], communications [3], and even financial markets [4].

Finding a graphical model for efficient signal representation is a prerequisite for graph signal processing. An undirected graph models bilateral correlations (e.g., factor graphs [5]), whereas for unilateral dependencies, a directed graph is utilized (e.g., Bayesian networks [6]). There are numerous approaches to learning the topology of the graph that best represents the data, including stochastic approaches via a Gaussian Markov random field (GMRF) model [7–9] or deterministic approaches incorporating measures of smoothness or stationarity for signal representation [10]. Another category of methods aims to recover a signal from corrupted (noisy or incomplete) measurements using a known graph model. Many of these algorithms exploit a graph-induced fidelity criterion, e.g., spatio-temporal smoothness [11] or total variation [12, 13] to help recover the graph signal.

Graph learning algorithms rely on the integrity (completeness) of the data, whereas for graph signal recovery methods, the graph model of the data needs to be given a priori. In practical applications, however, the data may be incomplete, or the graphical model may be unavailable. In this work, we study the problem of graph learning from incomplete data or semi-blind graph signal recovery, in which the underlying graph structure is unknown. Our method, applied to both i.i.d. and time-dependent data, is shown to improve the reconstruction quality of missing entries in the signal by joint estimation of the signal and graph.

The organization of the paper is as follows. In Section 1.1, we formulate the problem and in Section 2, we propose an iterative solution method. The simulation results are provided in Section 3. For the notations, we reserve bold lower-case letters for vectors (e.g., $\mathbf{x}$) and bold upper-case letters for matrices (e.g., $\mathbf{X}$). The Kronecker and the Hadamard products are respectively denoted with $\otimes$ and $\odot$. In addition, $^\circ$ and $\oslash$, are respectively used for element-wise power and division.

1.1. Problem Statement

An undirected graph with n vertices can be represented by $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{w}\}$, with $\mathcal{V} = \{1, \dots, n\}$ being the set of vertices, $\mathcal{E} \subseteq \{\{i, j\} | \, i, j \in \mathcal{V}\}$ the set of edges, and $\mathbf{w} \in \mathbb{R}^{n(n-1)/2}$ the edge weights. Let $\mathbf{X}^* = [\mathbf{x}_1^*, \dots, \mathbf{x}_T^*] \in \mathbb{R}^{n \times T}$ denote the matrix of the original signal for T time-stamps. Assume we have observations with missing entries as $\mathbf{Y} = \mathbf{M} \odot \mathbf{X}^*$, where $\mathbf{M}$ denotes the binary sampling mask matrix. Now, given $\mathbf{Y}$ and $\mathbf{M}$, the problem considered in this paper is to estimate the original signal $\mathbf{X}^*$ and the weights $\mathbf{w}^*$ of an undirected graph model that encode how similarly the signal elements vary with time (assuming spatial smoothness for the temporal variations of the signal).

2. PROPOSED METHOD

Assume the graph is connected. Using the graph learning framework in [8, 9] and assuming a Laplacian GMRF model for the temporal difference of the signal inspired by the notion of spatio-temporal smoothness [11], we propose to solve the following problem for joint estimation of $\mathbf{X}^*$ and $\mathbf{w}^*$

$$\mathbf{X}^*, \mathbf{w}^* = \underset{\mathbf{X}, \mathbf{w} \geq \mathbf{0}}{\operatorname{argmin}} \, f(\mathbf{X}, \mathbf{w}) \tag{1}$$

$$f(\mathbf{X}, \mathbf{w}) \triangleq \|\mathbf{Y} - \mathbf{M} \odot \mathbf{X}\|_F^2 + \alpha \operatorname{Tr}\left(\mathcal{L}(\mathbf{w}) \Delta(\mathbf{X}) \Delta(\mathbf{X})^\top\right) - \beta \log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J}) + \gamma \|\mathbf{w}\|_1$$

THIS WORK WAS SUPPORTED BY THE HONG KONG GRF 16207820 AND GRF 16206123 RESEARCH GRANTS.

where $\Delta(\mathbf{X}) = \mathbf{X} - \mathbf{X}\mathbf{D} = [\mathbf{x}_1 - \mathbf{x}_0, \ldots, \mathbf{x}_T - \mathbf{x}_{T-1}]$ represents the first-order temporal differences of the signal, and $\mathbf{D} \in \mathbb{R}^{T\times T}$ is an upper triangular matrix with only 1s as the first upper diagonal elements (assuming $\mathbf{x}_0 = \mathbf{0}$). Also, $\mathbf{J} = (1/n)\mathbf{1}\mathbf{1}^\top$, and $\mathcal{L} : \mathbb{R}^{n(n-1)/2} \to \mathbb{R}^{n\times n}$ denotes the Laplacian operator [14] that maps $\mathbf{w}$ to the Laplacian matrix. The adjoint of this operator is also denoted with $\mathcal{L}^*$. The first term in (1) is in fact the similarity criterion, and the second term quantifies the spatio-temporal smoothness, which measures the smoothness of temporal variations in a graph signal. For temporally i.i.d. data, this simply leads to the conventional graph signal smoothness [10], making our model applicable for both i.i.d. and time-dependent data. The third term, partially referred to as data-fidelity in [8], appears in the maximum likelihood estimation of the Laplacian in a GMRF model. The last term also acts as a sparsity-promoting regularization function on $\mathbf{w}$.

To solve problem (1), we use the block successive upperbound minimization (BSUM) [15, 16] method which is actually an extension of the block-coordinate-descent (BCD) [17]. Here, we minimize an uppperbound or a majorizer (see the definition in [15]) of the original cost with respect to a block variable in each iteration, resulting in unique and closed-form solutions to the subproblems (with guaranteed convergence [15]). These update steps are as follows:

2.1. X-update step

Let $\mathbf{w}$ be fixed. Then $f(\mathbf{X}, \mathbf{w})$ in (1), would be only a function of $\mathbf{X}$ as

$$f_{\mathbf{X}}(\mathbf{X}) = \operatorname{Tr}\left((\mathbf{Y} - \mathbf{M}\odot\mathbf{X})(\mathbf{Y} - \mathbf{M}\odot\mathbf{X})^\top\right) + \alpha \operatorname{Tr}\left(\mathcal{L}(\mathbf{w})\Delta(\mathbf{X})\Delta(\mathbf{X})^\top\right) + \text{const.} \tag{2}$$

Using vectorial representation, one obtains

$$f_{\mathbf{X}}(\mathbf{X}) = \operatorname{vec}(\mathbf{X})^\top \mathbf{G}\operatorname{vec}(\mathbf{X}) - 2\operatorname{vec}(\mathbf{X})^\top \mathbf{b} + \text{const}$$

where $\mathbf{G} = \operatorname{Diag}(\operatorname{vec}(\mathbf{M})) + \alpha\mathbf{H}^\top(\mathbf{I}_T \otimes \mathcal{L}(\mathbf{w}))\mathbf{H}$, $\mathbf{H} = \mathbf{I}_{nT} - \mathbf{D}^\top \otimes \mathbf{I}_n$, and $\mathbf{b} = \operatorname{vec}(\mathbf{Y})$.

Lemma 1. *Assume $\mathbf{X}_0$ to be constant. The function*

$$f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0) = f_{\mathbf{X}}(\mathbf{X}) + \operatorname{vec}(\mathbf{X} - \mathbf{X}_0)^\top(\theta\mathbf{I} - \mathbf{G})\operatorname{vec}(\mathbf{X} - \mathbf{X}_0)$$

is a strictly convex majorizer for $f_{\mathbf{X}}(\mathbf{X})$ if $\theta > 1 + 4\alpha\left\|\mathcal{L}(\mathbf{w})\right\|$.

Proof. For $\theta > \lambda_{\max}(\mathbf{G}) = \|\mathbf{G}\|$, the matrix $\theta\mathbf{I} - \mathbf{G}$ is positive definite. This implies that $f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0) \geq f_{\mathbf{X}}(\mathbf{X})$, and the equality is only occurred at $\mathbf{X} = \mathbf{X}_0$. In addition, we can obtain an upper-bound for $\|\mathbf{G}\|$ as

$$\begin{aligned}
\|\mathbf{G}\| &= \max_{\|\mathbf{x}\|=1} \left\|\operatorname{Diag}(\operatorname{vec}(\mathbf{M}))\mathbf{x} + \alpha\mathbf{H}^\top(\mathbf{I}\otimes\mathcal{L}(\mathbf{w}))\mathbf{H}\mathbf{x}\right\| \\
&\leq \max_{\|\mathbf{x}\|=1} \|\operatorname{Diag}(\operatorname{vec}(\mathbf{M}))\mathbf{x}\| + \alpha\left\|\mathbf{H}^\top(\mathbf{I}\otimes\mathcal{L}(\mathbf{w}))\mathbf{H}\mathbf{x}\right\| \\
&\leq \max_{\|\mathbf{x}\|=1} \|\mathbf{x}\| + \alpha\|\mathbf{H}\|^2\|\mathbf{I}\otimes\mathcal{L}(\mathbf{w})\|\|\mathbf{x}\| \\
&= 1 + \alpha\|\mathbf{H}\|^2\|\mathcal{L}(\mathbf{w})\|
\end{aligned}$$

It is also easy to show that $\|\mathbf{H}\| \leq 2$. Therefore, $f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0)$ would be a majorization function for $f_{\mathbf{X}}(\mathbf{X})$ if $\theta > 1 + 4\alpha\left\|\mathcal{L}(\mathbf{w})\right\| \geq \|\mathbf{G}\|$. We also obtain

$$f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0) = \theta\left\|\operatorname{vec}(\mathbf{X} - \mathbf{X}_0) + \frac{\mathbf{G}\operatorname{vec}(\mathbf{X}_0) - \mathbf{b}}{\theta}\right\|^2 + \text{const}$$

Hence, $f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}_0)$ is trivially a strictly convex (quadratic) function whose unique minimizer is given by $\operatorname{vec}(\mathbf{X}_0) - \frac{1}{\theta}\left(\mathbf{G}\operatorname{vec}(\mathbf{X}_0) - \mathbf{b}\right) = \operatorname{vec}\left(\mathbf{X}_0 - \frac{1}{2\theta}\frac{\partial}{\partial\mathbf{X}} f_{\mathbf{X}}(\mathbf{X})|_{\mathbf{X}_0}\right)$. □

Setting $\mathbf{X}_0 = \mathbf{X}^{(j)}$, the $\mathbf{X}$-update step yields as follows

$$\begin{aligned}
\mathbf{X}^{(j+1)} &= \operatorname*{argmin}_{\mathbf{X}} f^S_{\mathbf{X}}(\mathbf{X};\mathbf{X}^{(j)}) = \mathbf{X}^{(j)} - \tfrac{1}{2\theta}\tfrac{\partial}{\partial\mathbf{X}} f_{\mathbf{X}}(\mathbf{X}^{(j)}) \\
\tfrac{\partial}{\partial\mathbf{X}} f_{\mathbf{X}}(\mathbf{X}) &= 2\left(\alpha\mathcal{L}(\mathbf{w})\Delta(\mathbf{X})(\mathbf{I} - \mathbf{D}^\top) + \mathbf{M}\odot\mathbf{X} - \mathbf{Y}\right)
\end{aligned} \tag{3}$$

2.2. w-update step

Since $\mathbf{w} \geq \mathbf{0}$, we may write $2\|\mathbf{w}\|_1 = \operatorname{Tr}(\mathcal{L}(\mathbf{w})\mathbf{H}_{\text{off}})$, where $\mathbf{H}_{\text{off}} = \mathbf{I} - \mathbf{1}\mathbf{1}^\top$ Hence, assuming $\mathbf{X}$ to be fixed, after division by β, the cost function in (1) reduces to

$$f_{\mathbf{w}}(\mathbf{w}) = \operatorname{Tr}(\mathcal{L}(\mathbf{w})\mathbf{K}) - \log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J}) \tag{4}$$

where $\mathbf{K} = \frac{1}{\beta}\left(\alpha\Delta(\mathbf{X})\Delta(\mathbf{X})^\top + \gamma/2\mathbf{H}_{\text{off}}\right)$.

Lemma 2. *Let $\mathbf{q} = \mathcal{L}^*\left((\mathcal{L}(\mathbf{w}^{(j)}) + \mathbf{J})^{-1}\right)$, $\mathbf{r} = \mathcal{L}^*(\mathbf{K})$, and $\tau > 0$ be a constant. Also assume $\mathbf{w}_0 \geq 0$. Then, the following is a strictly convex majorization function for $f_{\mathbf{w}}(\mathbf{w})$.*

$$\begin{aligned}
f^S_{\mathbf{w}}(\mathbf{w};\mathbf{w}_0) = \tau\langle\mathbf{q}\odot\mathbf{w}_0^{\circ 2}, \mathbf{w}\oslash\mathbf{w}_0 + (\mathbf{w}_0 + 1/\tau)\oslash\mathbf{w} - 2\rangle + \\
\langle\mathbf{w},\mathbf{r}\rangle + \operatorname{Tr}\left((\mathcal{L}(\mathbf{w}_0) + \mathbf{J})^{-1}\mathbf{J}\right) - \log\det(\mathcal{L}(\mathbf{w}_0) + \mathbf{J}) - n
\end{aligned} \tag{5}$$

Proof. Taking advantage of the notion of the Laplacian operator [14], we have

$$\mathcal{L}(\mathbf{w}) + \mathbf{J} = \mathbf{E}\operatorname{Diag}(\mathbf{w})\mathbf{E}^\top + \mathbf{J} = \mathbf{G}\operatorname{Diag}(\tilde{\mathbf{w}})\mathbf{G}^\top \tag{6}$$

where $\tilde{\mathbf{w}} = [\mathbf{w}^\top 1/n]^\top$ and $\mathbf{G} = [\mathbf{E}, \mathbf{1}]$. The matrix $\mathbf{E} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n(n-1)/2}] \in \mathbb{R}^{n\times n(n-1)/2}$, consists of vectors $\boldsymbol{\xi}_k$ for $k = i - j + \frac{j-1}{2}(2n - j),\ i > j$, each of which has a $+1$ at the j-th position, a -1 at the i-th position, and zeros elsewhere. Since the $\log\det$ function is concave, an upperbound for $-\log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J})$ can be constructed via the following inequality [9]

$$\begin{aligned}
-\log\det(\mathcal{L}(\mathbf{w}) + \mathbf{J}) \leq \operatorname{Tr}\left(\mathbf{F}_0(\mathbf{G}\operatorname{Diag}(\tilde{\mathbf{w}})\mathbf{G}^\top)^{-1}\right) \\
- \log\det(\mathcal{L}(\mathbf{w}_0) + \mathbf{J}) - n
\end{aligned} \tag{7}$$

where $\mathbf{F}_0 = \mathcal{L}(\mathbf{w}_0) + \mathbf{J}$, and the equality is only achieved at $\mathbf{w} = \mathbf{w}_0$. Moreover, using Lemma 4 in [9], one can obtain

another majorizer as follows:

$$\begin{aligned}&\text{Tr}\left(\mathbf{F}_0(\mathbf{G}\text{Diag}(\tilde{\mathbf{w}})\mathbf{G}^\top)^{-1}\right) \\ &\leq \text{Tr}\left(\mathbf{F}_0^{-1}\mathbf{G}\text{Diag}(\tilde{\mathbf{w}}_0^{\circ 2} \oslash \tilde{\mathbf{w}})\mathbf{G}^\top\right) \\ &= \langle \mathbf{w}_0^{\circ 2} \oslash \mathbf{w}, \mathcal{L}^*(\mathbf{F}_0^{-1})\rangle + \text{Tr}\left(\mathbf{F}_0^{-1}\mathbf{J}\right)\end{aligned} \tag{8}$$

Now, let $g_{\mathbf{w}}(\mathbf{w};\mathbf{w}_0) = \text{Tr}(\mathcal{L}(\mathbf{w})\mathbf{K}) + \langle \mathbf{w}_0^{\circ 2} \oslash \mathbf{w}, \mathcal{L}^*(\mathbf{F}_0^{-1})\rangle$. Also, define $\mathbf{r} = \mathcal{L}^*(\mathbf{K})$ and $\mathbf{q} = \mathcal{L}^*(\mathbf{F}_0^{-1}) = \mathcal{L}^*(\mathcal{L}(\mathbf{w}_0 + \mathbf{J})^{-1})$. Then, $g_{\mathbf{w}}(\mathbf{w};\mathbf{w}_0)$ can be decomposed to scalar functions of w_i as

$$g_{\mathbf{w}}(\mathbf{w};\mathbf{w}_0) = \sum_{i=1}^{n(n-1)/2} g_{w_i}(w_i;w_{0i}) = r_i w_i + q_i \frac{w_{0i}^2}{w_i} \tag{9}$$

with $r_i = [\mathbf{r}]_i, \quad q_i = [\mathbf{q}]_i$. Now, let $h(x) = x + \frac{1}{x} - 2$. It can be easily shown that $h(x)$ is always non-negative for $x > 0$, and only achieves zero at $x = 1$. In addition, since $\mathcal{L}(\mathbf{w}^{(j)}) + \mathbf{J} \succ 0$ and $\mathbf{K} \succeq 0$, one can conclude that $q_i > 0$ and $r_i \geq 0$ (using the property of the adjoint Laplacian operator). Thus, one may suggest a majorizer for $g_{w_i}(w_i;w_{0i})$ as

$$\begin{aligned}g^S_{w_i}(w_i;w_{0i}) &= g_{w_i}(w_i;w_{0i}) + \tau q_i w_{0i}^2\ h(w_i/w_{0i}) \\ &= r_i w_i + \tau q_i w_{0i}^2 \left(\frac{w_i}{w_{0i}} + \frac{w_{0i} + 1/\tau}{w_i} - 2\right)\end{aligned}$$

with τ being a positive constant. Finally, after simplifications, one obtains the majorizer in (5) for $f_{\mathbf{w}}(\mathbf{w})$, using (7) and (8). It is easy to verify that $f^S_{\mathbf{w}}(\mathbf{w};\mathbf{w}^{(j)})$ is strictly convex for $\mathbf{w} \geq \mathbf{0}$ (via second order derivatives). □

Using the above lemma, the $\mathbf{w}$-update step yields as

$$\begin{aligned}\mathbf{w}^{(j+1)} &= \underset{\mathbf{w}}{\text{argmin}}\ f^S_{\mathbf{w}}(\mathbf{w};\mathbf{w}^{(j)}) \\ &= \mathbf{w}^{(j)} \odot \sqrt{(\tau\mathbf{w}^{(j)} \odot \mathbf{q} + \mathbf{q}) \oslash (\tau\mathbf{w}^{(j)} \odot \mathbf{q} + \mathbf{r})}.\end{aligned} \tag{10}$$

2.3. Main algorithm

The proposed method with all steps can be summarized in Algorithm 1. We choose $\mathbf{X}^{(0)} = \mathbf{Y}$ and $\mathbf{w}^{(0)} = \mathcal{P}_{\mathbf{w}\geq\mathbf{0}}(\mathbf{S}_Y^\dagger)$ for initialization, where $\mathbf{S}_Y = \frac{1}{T}\mathbf{Y}\mathbf{Y}^\top$ and $\mathcal{P}_{\mathbf{w}\geq\mathbf{0}}$ denotes the projection onto the set $\mathbf{w} \geq \mathbf{0}$. The stopping criterion is met when the relative error between consecutive iterations becomes smaller than a threshold or when the number of iterations exceeds a limit.

2.4. Computational complexity

The update step in (3) is $\mathcal{O}(n^2T + T^2n + n^3)$ computationally complex. Furthermore, given $\mathbf{K}$, the complexity of the $\mathbf{w}$ update step (10) is controlled by $(\mathcal{L}(\mathbf{w}^{(k)}) + \mathbf{J})^{-1}$, which needs $\mathcal{O}(n^3)$ operations. It also costs $\mathcal{O}(n^2T + T^2n)$ operations to compute $\mathbf{K}$. Hence, each iteration of Algorithm 1 is $\mathcal{O}(n^3 + n^2T + T^2n)$ computationally complex.

Algorithm 1 Proposed algorithm to solve problem (1)

Input: $\mathbf{Y}, \mathbf{M}$. **Parameters:** α, β, γ, and τ.
Output: $\mathbf{X}^{(j)}, \mathbf{L}^{(j)} = \mathcal{L}(\mathbf{w}^{(j)})$.
Initialization: $\mathbf{X}^{(0)} = \mathbf{Y}$, $\mathbf{w}^{(0)} = \mathcal{P}_{\mathbf{w}\geq\mathbf{0}}(\mathbf{S}_Y^\dagger)$, $j = 0$
repeat
  Obtain $\mathbf{X}^{(j+1)}$ using (3) with $\mathbf{w} = \mathbf{w}^{(j)}$.
  Obtain $\mathbf{w}^{(j+1)}$ via (10) with $\mathbf{X} = \mathbf{X}^{(j+1)}$.
  Set $j \leftarrow j + 1$
until a stopping condition is met

3. SIMULATION RESULTS

Here, we present the simulation results of our proposed algorithm for graph learning and missing sample recovery on synthetic and real data.

3.1. Synthetic data

For generation of synthetic data, we consider $n = 64$ and $T = 640$. We use the Stochastic Block Model for the underlying graph comprising of 4 clusters (blocks), with inter-cluster and intra-cluster edge probabilities of 0.7 and 0.075, respectively. The Laplacian matrix of the graph is then scaled to have $\text{Tr}(\mathbf{L}^*) = n$. We then generate random samples of the signal via $\mathbf{x}_t^* = \sqrt{\mathbf{L}^{*\dagger}}\boldsymbol{\nu}_t, \quad \boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The original data matrix $\mathbf{X}^*$ is then constructed as $\mathbf{X}^* = [\mathbf{x}_1^*, \ldots, \mathbf{x}_T^*]$. Next, we normalize $\mathbf{X}^*$, so that each row has zero mean and unit standard deviation. Finally, the observations are obtained as $\mathbf{Y} = \mathbf{M} \odot \mathbf{X}^*$, where $\mathbf{M}$ is the binary (sampling) mask matrix. Now, we provide the matrices $\mathbf{Y}$ and $\mathbf{M}$ as inputs, for signal and graph inference. The hyper-parameters of our method are chosen as $\alpha = 0.02$, $\beta = 0.02T$, $\gamma = 0.002T$, and $\tau = 100$. To measure the performance of the graph learning algorithms, we use the relative error (RelErr) and the F-score criteria. Let $\mathbf{L}^* \in \mathbb{R}^{n\times n}$ be the ground-truth Laplacian, and $\widehat{\mathbf{L}} \in \mathbb{R}^{n\times n}$ be the estimated one, the relative error and the F-score values are defined as

$$\text{RelErr} = \frac{\|\mathbf{L}^* - \widehat{\mathbf{L}}\|_F}{\|\mathbf{L}^*\|_F}, \qquad \text{F-score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}.$$

The terms TP, FP, and FN, respectively denote the true positive, false positive, and false negative connections in the inferred graph. The signal recovery performance is also measured via SNR and the normalized mean squared error (NMSE) criteria, defined as

$$\text{SNR} = 20\log_{10}\left(\frac{\|\mathbf{X}^*\|_F}{\|\mathbf{X}^* - \widehat{\mathbf{X}}\|_F}\right), \quad \text{NMSE} = \frac{1}{T}\sum_{i=1}^{T}\frac{\|\mathbf{x}_i^* - \hat{\mathbf{x}}_i\|^2}{\|\mathbf{x}_i^*\|^2}$$

where $\mathbf{X}^*$ and $\widehat{\mathbf{X}}$, denote the ground-truth and the estimated data matrices, with $\mathbf{x}_i^*$ and $\hat{\mathbf{x}}_i$ being their i-th columns, respectively.

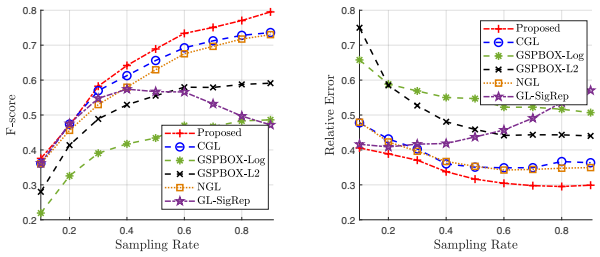

Fig. 1: Performance results of the Laplacian matrix estimation from synthetic data (in terms of F-score and relative error) versus the sampling rate (SR).

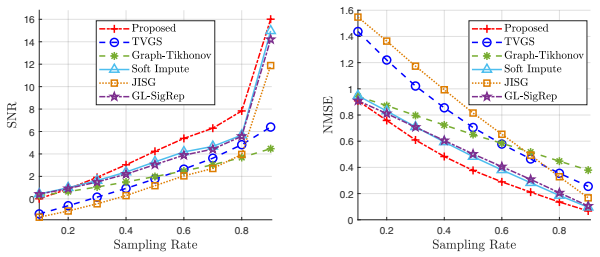

Fig. 2: Performance results of the original synthetic data matrix reconstruction (in terms of SNR and NMSE) versus the sampling rate.

3.1.1. Graph learning

Here, we evaluate the performance of the proposed method for learning the Laplacian matrix from incomplete data. We compare the results of our method with several benchmark algorithms for undirected graph learning. These include the CGL [8], the GSP toolbox graph learning methods [10], namely the GSPBOX-Log and the GSPBOX-L2, the GL-SigRep method in [18] for both graph learning and signal recovery assuming smooth signal representation, and the nonconvex graph learning algorithm in [19] called NGL[1]. For fair comparison, the estimated Laplacian matrix is scaled such that $\mathrm{Tr}(\widehat{\mathbf{L}}) = n$. Figure 1 shows the estimation results in terms of RelErr and F-score versus different values of the sampling rate (SR). As it is implied from the figure, the proposed algorithm has superior performance in estimating the graph Laplacian matrix, especially at higher sampling rates.

3.1.2. Data matrix recovery

In this part, we evaluate the performance of the proposed method for recovery of the original signal $\mathbf{X}^*$. We compare our method with several benchmark algorithms, some of which are also based on graphical modeling. These include the SOFT-IMPUTE algorithm[2] for matrix completion using nuclear norm regularization [20], the GL-SigRep method [18], the method for joint inference of signals and graphs (JISG) [21], the time-varying graph signal reconstruction method (TVGS) [22] and the method in [23] named as Graph-Tikhonov. For the last two methods, the underlying graph Laplacian matrix must be given a priori. For this, we use the CGL algorithm to infer the Laplacian matrix from the incomplete observations $\mathbf{Y}$. Figure 2 depicts the performance of the proposed algorithm compared to the benchmark methods for recovery of the data matrix, at different values of the sampling rate. The proposed method is shown to outperform the other methods for recovery of the missing data.

[1] https://github.com/mirca/sparseGraph

[2] https://cran.r-project.org/web/packages/softImpute/index.html

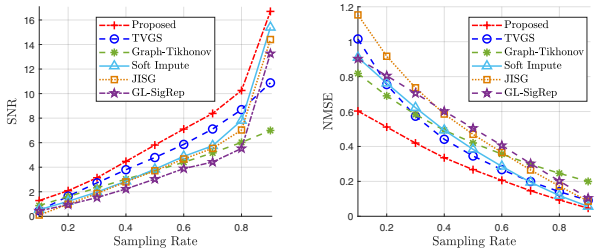

Fig. 3: Performance results of the recovery algorithms (in terms of SNR and NMSE) for reconstruction of PM2.5 data matrix from incomplete measurements, at different sampling rates.

3.2. Real data

This part provides numerical results on real data. For this purpose, we use the data corresponding to the PM2.5 concentration in the air of the state of California. The data matrix of dimension 93x300, contains concentration values of PM2.5 particulate matters, measured at 93 stations (in California) for 300 days starting from January 1, 2015. We then normalize the data as explained in the previous part and construct $\mathbf{Y} = \mathbf{M} \odot \mathbf{X}^*$ (with random $\mathbf{M}$). Here, we only provide the results of graph signal recovery (since there is no ground-truth graph model). Figure 3 is an example to show that our proposed method has better performance compared to the benchmark algorithms, in recovery of real-world signals.

4. CONCLUSION

In this paper, we examined the problem of learning a graph from incomplete data, which can also be considered as semi-blind recovery of missing samples of a time-varying graph signal. We proposed an algorithm to jointly estimate the underlying graph model and the signal based on the block successive upperbound minimization method. We further analyzed the computational complexity of our proposed method in Section 2.4. The results of simulations on synthetic and real data provided in Section 3, also demonstrate the efficiency of the proposed method for both signal recovery and graph inference from incomplete time-series observations.

5. REFERENCES

[1] W. Campbell, C. Dagli, and C. Weinstein, "Social Network Analysis with Content and Graphs," *Lincoln Laboratory Journal*, vol. 20, no. 1, pp. 61–81, 2013.

[2] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, Jan. 2021, arXiv: 1901.00596.

[3] H. Tamura, K. Nakano, M. Sengoku, and S. Shinoda, "On Applications of Graph/Network Theory to Problems in Communication Systems," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 5, no. 1, pp. 15–21, Jan. 1970.

[4] J. V. de M. Cardoso and D. P. Palomar, "Learning undirected graphs in financial markets," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 741–745.

[5] H. A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 28–41, Jan. 2004.

[6] A. Lalitha, X. Wang, O. Kilinc, Y. Lu, T. Javidi, and F. Koushanfar, "Decentralized Bayesian Learning over Graphs," *arXiv:1905.10466 [cs, stat]*, May 2019.

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

[8] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph Learning from Data under Structural and Laplacian Constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

[9] 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, Aug. 2019.

[10] V. Kalofolias, "How to learn a graph from smooth signals," in *Artificial intelligence and statistics*. PMLR, 2016, pp. 920–929.

[11] X. Mao, K. Qiu, T. Li, and Y. Gu, "Spatio-Temporal Signal Recovery Based on Low Rank and Differential Smoothness," *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6281–6296, Dec. 2018.

[12] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, "Signal Recovery on Graphs: Variation Minimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 17, pp. 4609–4624, Sept. 2015.

[13] P. Berger, G. Hannak, and G. Matz, "Graph Signal Recovery via Primal-Dual Algorithms for Total Variation Minimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 842–855, Sept. 2017.

[14] S. Kumar, J. Ying, J. V. de 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.

[15] M. Razaviyayn, M. Hong, and Z. Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.

[16] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[17] P. Tseng, "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization," *Journal of Optimization Theory and Applications*, vol. 109, no. 3, pp. 475–494, June 2001.

[18] 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, Dec. 2016.

[19] J. Ying, J. V. de M. Cardoso, and D. Palomar, "Nonconvex Sparse Graph Learning under Laplacian Constrained Graphical Model," in *Advances in Neural Information Processing Systems*. 2020, vol. 33, pp. 7101–7113, Curran Associates, Inc.

[20] R. Mazumder, T. Hastie, and R. Tibshirani, "Spectral Regularization Algorithms for Learning Large Incomplete Matrices," *Journal of Machine Learning Research*, vol. 11, no. 80, pp. 2287–2322, 2010.

[21] V. N. Ioannidis, Y. Shen, and G. B. Giannakis, "Semi-Blind Inference of Topologies and Dynamical Processes Over Dynamic Graphs," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2263–2274, May 2019.

[22] K. Qiu, X. Mao, X. Shen, X. Wang, T. Li, and Y. Gu, "Time-Varying Graph Signal Reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 870–883, Sept. 2017.

[23] N. Perraudin, A. Loukas, F. Grassi, and P. Vandergheynst, "Towards stationary time-vertex signal processing," in *2017 ICASSP*, 2017, pp. 3914–3918.