Blind Deconvolution on Graphs with Noisy Observations

Jinhong Du[1]
Advisor: Lihua Yang[2]

***Abstract*— Graph convolution can be viewed as a network diffusion determined by a *graph-shift operator* S and *filter coefficients* h. This paper deals with the *graph deconvolution* problem where the unknown input signal is *sparse* and the observed output signal is noisy, i.e., to recover the sparse input signal x and the vector of filter coefficients h from the noisy output signal y with a known graph-shift operator, which extends the scope of classical blind deconvolution of temporal or spatial signals. While y is a bilinear function of x and h, it is also a linear function of the vectorization of the rank-one and row-sparse matrix $\mathbf{x}\mathbf{h}^\top$. Thus, we can use rank and sparsity minimization to handle the original problem by the *lifting* method. We prove the equivalence of the two optimization problems under some conditions. By convex relaxation, we further simplify the optimization problem and obtain a probabilistic guarantee for entry-wise l_1 minimization to recover $\mathbf{x}\mathbf{h}^\top$ of the lifting problem with bounded errors when the observation of y contains noise. Besides, we also provide the lower bound of the number of observations N required for the theorem to hold. Numerical results show the effectiveness of our proposed algorithms in both generated and real-world graphs.**

Keywords: graph deconvolution, graph signal processing, sparsity, lifting method, convex relaxation

I. Introduction

With the rise of big data, how to deal with all kinds of data has become a problem. Real-world data are often structured, and graphs are powerful to describe most kinds of structured data such as social networks, traffic networks and so on. The edges of a graph can represent the relationship structure of data, and the values at the vertices can represent the signals. A one-dimension graph signal is a vector of which the i-th entry equals the value on the i-th node. The signal in a node can be transmitted through the edges to another node with or without weight. One challenge in *graph signal processing* is to generalize traditional signal processing tools on graphs, which belongs to irregular domains. The primary approach would be focusing on the vertex (spatial) domain or the frequency (graph spectral) domain.

Blind identification of graph filters is the extension of the classical signal deconvolution problem. Classical signal deconvolution problem aims to recover two unknown vectors from their circular convolution[1], while we extend the deconvolution process to the graph setting. Graph signal convolution relies on the operator called *graph-shift operator*[1], [2]. Graph-shift operator is a useful tool to describe the graph network diffusion, which demonstrates the local transitivity of the graph signals. By choosing the *filter coefficients* $\mathbf{h}$, one can form a *graph filter* $\mathbf{H}$ as polynomials of the graph-shift operator $\mathbf{S}$[3]. The output signals would be given by $\mathbf{y} = \mathbf{H}\mathbf{x}$, and it can also be explained by $\mathbf{y} = \mathbf{x} *_{\mathcal{G},\mathbf{S}} \mathbf{h}$ as shown in the following sections. The problem we are dealing with is that in the case we only observe the noisy output signals while the input signals and the filter coefficients are unknown, we want to recover $\mathbf{x}$ and $\mathbf{h}$, i.e., *blind deconvolution*. Since this is an ill-posed inverse problem, we assume that the length of $\mathbf{h}$ is small and the input signal $\mathbf{x}$ is sparse.

The assumption of the sparsity of input signals is reasonable for many real-world problems. For example, in epidemiology study, the sources of one disease may be limited to only some cities. Assuming that the disease spreads abiding by the local connectivity with some specific probability, we can describe it by a weighted directed graph. After a few time steps, we observed the number of patients in different cities and hope to identify the sources of the disease. Besides, in social network analysis, the behavior of several individuals will influence their neighbors with some specific probability. Then it can be models as a graph diffusion process, and can also be discretized and approximated by a linear graph convolution process[4]. Last but not least, in many applications like multi-agent systems, graph filters are used distributedly to perform linear transformations such as distributed average consensus [5] and so on. In this scenario, our algorithm can be used to locate the influential agents, i.e., the non-zero entries in the input signal $\mathbf{x}$. Beyond what have been mentioned above, our proposed algorithm also applies to many other practical problems related to graph signal processing.

There is only limited research in the field of graph signals deconvolution. Related work has been done in [6], [7], [8], where the authors provide the probabilistic guarantee for accurate blind identification of both the sparse input signals and the filter coefficients under some specific conditions. Their conclusions rely on a novel technique in the field of compressive sensing called SparseLift[9] and apply to the graph spectral domain. Our work follows the idea of [8], refines some of the proof in both [8] and [9] and extends the probability guarantee for recovery of sparse inputs and filter coefficient with noisy observations, which is more common in real-world problems. The improvement can also be adapted to achieve a tighter bound in a noiseless setting. Furthermore, we provide an instructive lower bound of the number of nodes N, which ensures the probabilistic bound of our main theorem is valid.

The rest of the paper is organized as follows. In Section 2, we introduce the settings of related problems in detail and list different optimization problems in the presence of noise

[1]Jinhong Du is with the School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China. dujh7@mail2.sysu.edu.cn
[3]Lihua Yang is with the School of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China. mcsylh@mail.sysu.edu.cn

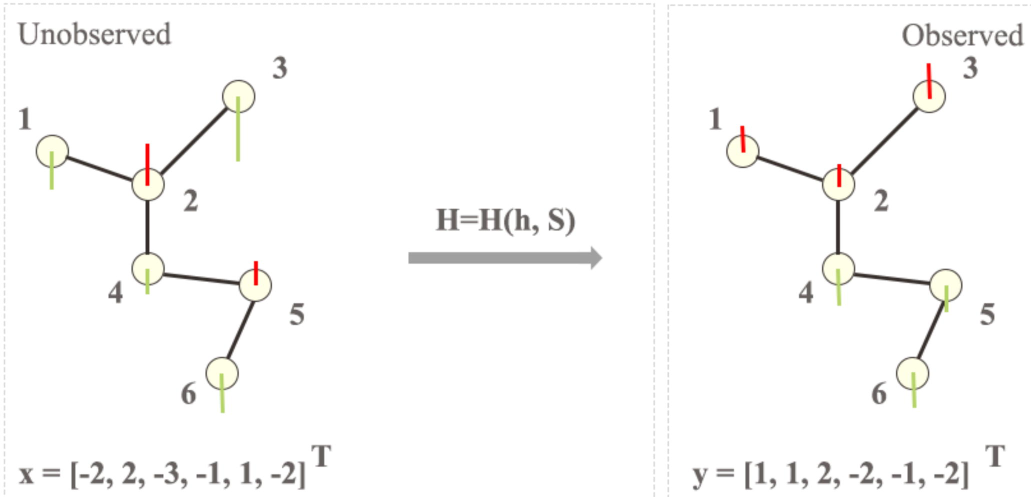

Fig. 1: Figure explanation for the graph model.

and no noise. In Section 3, we present our main results with detailed proof in Section 4. Section 5 provides numerical results of our proposed algorithms in both generated data and real-world data, and finally, Section 6 gives conclusions of the whole paper.

II. Problem Formulation

A. Notation

Some basic notations are defined below to facilitate our problem formulation.

Entries of a matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ and a vector $\mathbf{a} \in \mathbb{C}^N$ are denoted as A_{ij} and a_j for $1 \le i \le M$ and $1 \le j \le N$, respectively. The i-th row of $\mathbf{A}$ will be denoted by $\mathbf{a}_i^\top$.

For any vector $\mathbf{a}$, $\bar{\mathbf{a}}$ is its complex conjugate, and $\operatorname{diag}(\mathbf{a})$ is a diagonal matrix with i-th diagonal entry equals to a_i. The notations $\|\mathbf{a}\|_2$ and $\|\mathbf{a}\|_0$ stand for the vector l_2 norm and l_0 pseudo-norm, respectively

For matrices, the symbols $(\cdot)^\top$, $(\cdot)^H$, $\mathbb{P}$, $\mathbb{E}$, $\operatorname{vec}(\cdot)$, $\circ$, $\otimes$ and $*$ signify matrix transpose, conjugate transpose, probability, expectation, vectorization, Hadamard product (entry-wise product), Kronecker product and Khatri-Rao product (column-wise Kronecker product). We use $\mathbf{I}_N$ and $\mathbf{0}_N$ to denote a $N\times N$ identity matrix and a $N\times N$ zero matrix, respectively. The $M\times N$ zero matrix is represented by $\mathbf{0}_{M\times N}$. The notations $\|\mathbf{A}\|_\infty$, $\|\mathbf{A}\|_2$, $\|\mathbf{A}\|_F$, $\|\mathbf{A}\|$, $\|\mathbf{A}\|_0$ and $\|\mathbf{A}\|_*$ stand for the entrywise largest absolute value, the matrix l_2 norm, the Frobenius norm, the spectral norm (the largest singular value) of matrix $\mathbf{A}$, the number of non-zero rows of $\mathbf{A}$, and the nuclear norm (the sum of all singular values).

For a linear mapping $\mathcal{A}: \mathbb{C}^{N\times L} \mapsto \mathbb{C}^{P\times Q}$, $\|\mathcal{A}\|$ is defined as $\|\mathbf{A}\|$ where $\mathbf{A} \in \mathbb{C}^{PQ\times LN}$ is the matrix representation of $\mathcal{A}$ such that $\operatorname{vec}(\mathcal{A}(\mathbf{X})) = \mathbf{A}\operatorname{vec}(\mathbf{X})$, $\forall\ \mathbf{X} \in \mathbb{C}^{N\times L}$.

B. The Graph Model

We denote a simple undirected graph by $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ containing no graph loops or multiple edges, where $\mathcal{N}$ is the set of N nodes and $\mathcal{E}$ is the set of edges. If an edge from node i to node j exists, then $(i,j) \in \mathcal{E}$. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ of the graph has entries $A_{ji} = 1$ when $(i,j) \in \mathcal{E}$ and zero otherwise.

The *graph-shift operator* $\mathbf{S} \in \mathbb{R}^{N\times N}$ is a kind of generalization of the shift or delay filter in discrete signal processing[10], [2], [3]. $\mathbf{S}$ can represent the local connection of the graph with non-zero entries S_{ji} only when $(i,j) \in \mathcal{E}$ or $i = j$. In particular, the adjacency matrix $\mathbf{A}$ and the Laplacian matrix are graph-shift operators. Given the *coefficients of filters* $\mathbf{h} = \begin{pmatrix} h_1 & h_2 & \cdots & h_L \end{pmatrix}^\top$, the *graph filter* is defined by $\mathbf{H} = \sum_{l=1}^{L} h_l \mathbf{S}^{l-1}$. If $\mathbf{x} \in \mathbb{R}^N$ is the input graph signals, then the output signals would be $\mathbf{y} = \mathbf{H}\mathbf{x} \in \mathbb{R}^N$. One may define this model as $\mathbf{y} = \mathbf{x} *_{\mathcal{G},\mathbf{S}} \mathbf{h}$, the convolution of $\mathbf{x}$ and $\mathbf{h}$ with the graph-shift operator $\mathbf{S}$. The length of $\mathbf{h}$ and the local connectivity of $\mathbf{S}$ determine all the L-hop neighbors to be convoluted for one node. For example, in Fig. 1, nodes 1, 3 and 4 are the 1-hop neighbors of node 2, and node 5 is the 2-hop neighbor of node 2. If we choose $L = 2$, then the convolution operation at node 2 will only be affected by its 1-hop and 2-hop neighbors. In particular, when $\mathbf{S}$ is chosen as the Laplacian matrix, it turns out to be a very common setup in deep convolutional networks on graph-structured data[11], [12].

For a sparse input with the dimension of support being $S \ll N$, without loss of generality, we will assume the first S entries of $\mathbf{x}$ are non-zero. When $\mathbf{S}$ is normal, the spectral decomposition of $\mathbf{S}$ is given by $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ where $\mathbf{V}$ is unitary, i.e., $\mathbf{V}^H\mathbf{V} = \mathbf{I}_N$. Then

$$\begin{aligned} \mathbf{y} &= \mathbf{H}\mathbf{x} \\ &= \sum_{l=1}^{L} h_l \mathbf{V}\mathbf{\Lambda}^{l-1}\mathbf{V}^H\mathbf{x} \\ &= \mathbf{V}\left(\sum_{l=1}^{L} h_l \mathbf{\Lambda}^{l-1}\right)\mathbf{V}^H\mathbf{x} \\ &= \mathbf{V}\operatorname{diag}(\mathbf{\Psi}\mathbf{h})\mathbf{V}^H\mathbf{x} \end{aligned}$$

where $\mathbf{\Psi} \in \mathbb{C}^{N\times L}$ is the Vandermonde matrix where $\Psi_{ij} = \Lambda_{ii}^{j-1}$. By setting $\mathbf{U} = \sqrt{N}\mathbf{V}^H$, we have $\mathbf{U}\mathbf{y} = \operatorname{diag}(\mathbf{\Psi}\mathbf{h})\mathbf{U}\mathbf{x}$

where the scale constant $\sqrt{N}$ is used to ensure that $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$, making $\mathbf{U}$ as an analogy of the unnormalized discrete Fourier matrix. The frequency representation of the input signal $\mathbf{x}$, of the output signal $\mathbf{y}$ and the filter $\mathbf{h}$, is given by $\hat{\mathbf{x}} = \mathbf{U}\mathbf{x}$, $\hat{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$ and $\hat{\mathbf{y}} = \mathbf{U}\mathbf{y}$ respectively. Thus, the model can be expressed as $\hat{\mathbf{y}} = \mathrm{diag}(\hat{\mathbf{h}})\hat{\mathbf{x}} = \hat{\mathbf{h}} \circ \hat{\mathbf{x}}$ in the spectral domain.

Consider the economy-sized singular value decomposition $\mathbf{\Psi} = \mathbf{P}\mathbf{\Sigma}\mathbf{R}^H$ where $\mathbf{\Sigma} \in \mathbb{R}^{L\times L}$ is a diagonal matrix, and, $\mathbf{P} \in \mathbb{C}^{N\times L}$ and $\mathbf{R} \in \mathbb{C}^{L\times L}$ are matrices satisfying $\mathbf{P}^H\mathbf{P} = \mathbf{R}^H\mathbf{R} = \mathbf{I}_L$, we have

$$\mathbf{U}\mathbf{y} = \mathrm{diag}(\mathbf{P}\mathbf{\Sigma}\mathbf{R}\mathbf{h})\mathbf{U}\mathbf{x} \xlongequal{\mathbf{h}'=\mathbf{\Sigma}\mathbf{R}\mathbf{h}} \mathrm{diag}(\mathbf{P}\mathbf{h}')\mathbf{U}\mathbf{x}.$$

If $\mathbf{\Psi}$ has full rank and $\mathbf{h}'$ is available, then we can recover $\mathbf{h}$ by $\mathbf{R}\mathbf{\Sigma}^{-1}\mathbf{h}$. However, when $L' = \mathrm{rank}(\mathbf{\Psi}) \leq L$, we can only partially recover $\mathbf{h}$ from $\mathbf{h}' \in \mathbb{R}^{L'}$. Therefore, we just need to figure out how to recover $\mathbf{h}'$ in the above setting.

C. Optimization Problem without Noise

We shall first consider the noiseless setting since the equivalence of some optimization problems is built on such deterministic conditions. Then in Section 2.4, we will extend our scope to the noisy setting.

1) Original Problem: Our optimization problem is given by

$$\begin{aligned} \text{find} \quad & \mathbf{h}', \mathbf{x} \\ s.t. \quad & \mathbf{U}\mathbf{y} = \mathrm{diag}(\mathbf{P}\mathbf{h}')\mathbf{U}\mathbf{x} \\ & \|\mathbf{x}\|_0 \leq S \\ & \mathbf{h}' \neq \mathbf{0}_{L\times 1} \end{aligned} \tag{P.1}$$

Notice that this problem is identifiable only up to a scale factor. As long as we find out a pair of vectors $(\mathbf{x}, \mathbf{h}')$ which is one of the minimizers of Problem (P.1), we consider it a successful recovery.

The first constraint of the above optimization problem is a bilinear constraint, which means that if we fix one of the two variables $\mathbf{h}'$ and $\mathbf{x}$, it is a linear constraint concerning the other variable since $\mathrm{diag}(\mathbf{P}\mathbf{h}')\mathbf{U}\mathbf{x} = (\mathbf{P}^\top * \mathbf{U}^\top)^\top \mathrm{vec}(\mathbf{x}\mathbf{h}'^\top)$.

2) Lifting the Bilinear Constraints: The bilinear constraint is hard to deal with. As suggested by [1], [9], by setting $\mathbf{Z} = \mathbf{x}\mathbf{h}'^\top$ we can change the vector-valued problem into a matrix-valued problem that

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \mathrm{rank}(\mathbf{Z}) \\ \text{s.t.} \quad & \mathbf{U}\mathbf{y} = (\mathbf{P}^\top * \mathbf{U}^\top)^\top \mathrm{vec}(\mathbf{Z}) \\ & \|\mathbf{Z}\|_0 \leq S \end{aligned} \tag{P.2}$$

where $\|\mathbf{Z}\|_0$ denotes the number of non-zero rows of $\mathbf{Z}$. Now our goal is to find a rank-one row-sparse matrix $\mathbf{Z}$ satisfying the first constraint. Hopefully, after recovering $\mathbf{Z}$, we can take a left and a right singular vector of $\mathbf{Z}$ as $\mathbf{x}$ and $\mathbf{h}'$ respectively when $\mathrm{rank}(\mathbf{Z}) = 1$. When $\mathrm{rank}(\mathbf{Z}) = 0$, i.e., $\mathbf{Z} = \mathbf{0}_{N\times L}$, this is a special case when $\mathbf{x} = 0$. When $\mathrm{rank}(\mathbf{Z}) > 1$, we can recover $\mathbf{x}$ and $\mathbf{h}'$ by computing the left and right singular vectors of $\mathbf{Z}$ concerning the largest singular values.

There is a problem that naturally arises - whether this problem is equivalent to the original problem? In [8], the author gives a proof of the equivalence under some conditions. The equivalence of Problem (P.1) and Problem (P.2) means that every solution to one problem gives a solution to the other and vice versa. First, we introduce some notations. Let $\mathcal{I} \subset \{1, 2, \cdots, N\}$ be row indices and $\mathcal{I}^C = \{1, 2, \cdots, N\} \setminus \mathcal{I}$ be the complementary set of indices. For any matrix $\mathbf{A}$, $\mathbf{A}_\mathcal{I}$ denotes a matrix formed by concatenating the rows of $\mathbf{A}$ indexed by $\mathcal{I}$, and $\mathrm{spark}(\mathbf{A})$ denotes the smallest possible number of $\mathbf{A}$'s different columns to be linear dependent (when $\mathbf{A}$'s columns are linear independent, $\mathrm{spark}(\mathbf{A}) = \infty$)[13]. We slightly modify the statement of Proposition 1 in [8] and present it in Proposition 1.

Proposition 1: Let Problem (P.1) be feasible and $S \geq 2$, then Problem (P.1) and Problem (P.2) are equivalent if and only if there exists a set of row indices $\mathcal{I}$ with $\mathrm{spark}(\mathbf{U}_\mathcal{I}) \leq S$ such that $\mathrm{rank}(\mathbf{\Psi}_{\mathcal{I}^C}) \geq L$.

Notice that the condition that Problem (P.1) is feasible is essential since it ensures the minimizers of Problem (P.2) have at most rank one so that $\mathbf{x}$ and $\mathbf{h}'$ can be recovered from $\mathbf{Z}$ without loss of information regardless of a scale factor. Also, we require the condition $S \geq 2$ in Proposition 1, since when $\mathrm{spark}(\mathbf{U}_\mathcal{I}) \leq S = 1$, $\mathrm{spark}(\mathbf{U}_\mathcal{I})$ is precisely one, which means that at least one column of $\mathbf{U}_\mathcal{I}$ is a zero vector and however, this will not hold when $\mathbf{U}$ has all non-zero entries. Although Proposition 1 excludes the case when $S = 1$, we have to point out that when $S = 1$ and Problem (P.1) is feasible, Problem (P.1) and Problem (P.2) are equivalent naturally. Furthermore, as shown in Corollary 1 in [8], if all the eigenvalues of $\mathbf{S}$ be distinct and $N \geq L + S - 1$, then Problem (P.1) and Problem (P.2) are equivalent, though the minimizer of (P.2) is unwarrantable to be unique.

3) Convex Relaxation: Even though we get rid of the bilinear constraints by lifting the original problem, Problem (P.2) is still hard to deal with. Convex relaxation is a method to convert a nonconvex problem into a convex problem which is relatively easy to solve and has near-optimal solutions of the original problem[14]. A convex surrogate to $\mathrm{rank}(\mathbf{Z})$ is the nuclear norm $\|\mathbf{Z}\|_* = \sum_{i=1}^{\min\{N,L\}} \sigma_i(\mathbf{Z})$ where $\sigma_i(\mathbf{Z})$ is the i-th singular value of $\mathbf{Z}$[15]. Similarly, a convex surrogate to $\|\mathbf{Z}\|_0$ can be the mixed $l_{2,1}$ norm $\|\mathbf{Z}\|_{2,1} = \sum_{i=1}^N \|\mathbf{z}_i^\top\|_2 = \sum_{i=1}^N \left(\sum_{j=1}^L Z_{ij}^2\right)^{1/2}$, i.e., the sum of the Euclidean norms of the rows of the matrix[16]. Then the relaxed convex optimization problem is given by

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{Z}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{U}\mathbf{y} = (\mathbf{P}^\top * \mathbf{U}^\top)^\top \mathrm{vec}(\mathbf{Z}) \end{aligned} \tag{P.3}$$

where λ is a hyperparameter to control the low rank and row-sparsity tradeoff. Also, one can further refine Problem (P.3) by iteratively reweighting[8], but this is beyond the scope of our discussion.

Also, one can use the entrywise l_1 matrix norm $\|\mathbf{Z}\|_1 = \sum_{i=1}^N \sum_{j=1}^L |Z_{ij}|$ to replace the term of $\|\mathbf{Z}\|_0$ in Problem (P.2). However, in [9], the author suggests only minimizing the entrywise l_1 matrix norm because for any matrix $\mathbf{A} \in \mathbb{R}^{N\times L}$, $L^{-\frac{1}{2}}\|\mathbf{A}\|_1 \leq \|\mathbf{A}\|_F \leq \|\mathbf{A}\|_* \leq \sqrt{\mathrm{rank}(A)}\|\mathbf{A}\|_F \leq \sqrt{L}\|\mathbf{A}\|_1$. This convex relaxation method called SparseLift, which gives rise to the following optimization problem.

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_1 \\ \text{s.t.} \quad & \mathbf{U}\mathbf{y} = (\mathbf{P}^\top * \mathbf{U}^\top)^\top \operatorname{vec}(\mathbf{Z}) \end{aligned} \tag{P.4}$$

The reason behind this is that in [17] and [18], it is shown that minimizing only the l_1 norm of the matrix is sufficient to recover simultaneously low-rank and row-sparse matrix. We only pursue the entrywise sparsity of $\mathbf{Z}$ without any structure constraints. This problem can be transformed to a linear programming problem and is easier to compute than Problem (P.3) since one need not to compute the singular value decomposition for the nuclear norm and tune the hyperparameter λ. Besides, it is shown that entrywise l_1 minimization is comparable with $l_* + \lambda l_{2,1}$ minimization.

D. Optimization Problem with Noise

If our observation $\mathbf{y}$ contains noise, i.e., $\omega = \mathbf{y} - \mathbf{y}_{\text{true}}$ and $\|\omega\|_2 \le \eta$, then the original optimization problem is given by

$$\begin{aligned} \text{find} \quad & \mathbf{h}', \mathbf{x} \\ s.t. \quad & \|\mathbf{U}\mathbf{y} - \operatorname{diag}(\mathbf{P}\mathbf{h}')\mathbf{U}\mathbf{x}\|_2 \le \eta \\ & \|\mathbf{x}\|_0 \le S \\ & \mathbf{h}' \ne \mathbf{0}_{L\times 1} \end{aligned} \tag{P.5}$$

Problem (P.3) and Problem (P.5) become Problem (P.6) and Problem (P.7), respectively.

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{Z}\|_{2,1} \\ \text{s.t.} \quad & \|\mathbf{U}\mathbf{y} - (\mathbf{P}^\top * \mathbf{U}^\top)^\top \operatorname{vec}(\mathbf{Z})\|_2 \le \eta \end{aligned} \tag{P.6}$$

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_1 \\ \text{s.t.} \quad & \|\mathbf{U}\mathbf{y} - (\mathbf{P}^\top * \mathbf{U}^\top)^\top \operatorname{vec}(\mathbf{Z})\|_2 \le \eta \end{aligned} \tag{P.7}$$

III. Main Results

A. Preliminary Definitions

Before presenting our main results, we first give the following assumptions.

Assumption 1: The graph-shift operator $\mathbf{S}$ has N distinct eigenvalues and is normal, i.e., it satisfies $\mathbf{S}\mathbf{S}^H = \mathbf{S}^H\mathbf{S}$.

Assumption 2: The frequency representation of the observed graph signal $\mathbf{y}$ adheres to the model $\hat{\mathbf{y}} = \operatorname{diag}(\mathbf{P}\mathbf{h}')\widetilde{\mathbf{U}}\mathbf{x} + \mathbf{w}$, where $\widetilde{\mathbf{U}}$ is a random $N \times N$ matrix obtained by concatenating N rows sampled independently and uniformly with replacement from $\mathbf{U}$, and $\mathbf{w}$ is the random noise with $\|\mathbf{w}\|_2 \le \eta$. Also, $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$.

Assumption 3: There exists a disjoint partition of the index set $\{1,2,\cdots,L\}$ into $\{\Gamma_p\}_{p=1}^P$ with $|\Gamma_p| = Q$, $\bigcup_{p=1}^P \Gamma_p = \{1,2,\cdots,L\}$ and $\Gamma_p \cap \Gamma_{p'} = \varnothing$, such that $N = PQ$ and

$$\max_{1\le p\le P} \left\| \mathbf{T}_p - \frac{Q}{N}\mathbf{I}_L \right\| \le \frac{Q}{4N}$$

where $\mathbf{T}_p = \sum_{i\in\Gamma_p} \overline{\psi_i}\psi_i^\top$ and $\psi_i^\top$ is the i-th row of $\mathbf{P}$.

In Assumption 1, the distinct eigenvalues of $\mathbf{S}$ ensure that $\mathbf{P}$ has full rank so that we can assume $\mathbf{P}^H\mathbf{P} = \mathbf{I}_L$ without loss of generality in Assumption 2. To see this, we first look at the singular value decomposition of $\mathbf{P}$ is given by $\mathbf{P} = \mathbf{P}\boldsymbol{\Sigma}\mathbf{R}$. Since $\mathbf{P}\mathbf{h}' = \mathbf{P}\boldsymbol{\Sigma}\mathbf{R}\mathbf{h} = \mathbf{P}\mathbf{h}''$, we can choose $\mathbf{P} \in \mathbb{R}^{N\times L'}$ as $\mathbf{P}$ and partially recover $\mathbf{h}'$ from $\mathbf{h}'' \in \mathbb{R}^{L'}$, where $L' = \operatorname{rank}(\mathbf{P}) \le L$. The normality of $\mathbf{S}$ leads to the spectral decomposition of $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ so that the model in Assumption 2 can be set up. However, Assumption 1 can be weakened to only require that $\mathbf{S}$ is a symmetric matrix.

In Assumption 2, the use of $\widetilde{\mathbf{U}}$ tends to generalize the case of the random Fourier model.

In Assumption 3, it follows that

$$\max_{1\le p\le P} \|\mathbf{T}_p\| \le \frac{5Q}{4N} \tag{1}$$

In the case when N cannot be divisible by Q, we can choose $\lfloor \frac{N}{Q} \rfloor$ partitions with size Q and one with size less than Q. Discussing the case of equal partitions is convenient. However, the sizes of all partitions can be different.

Meanwhile, we also introduce some useful notations and definitions. For a given matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ and a positive integer $k \in \mathbb{N}$ and $k \le N$, we define the *extended mutual coherence*, a function $\rho_{\mathbf{A}}(k)$ as

$$\rho_{\mathbf{A}}(k) = \max_{i\in\{1,2,\cdots,M\}} \max_{\Omega\in\Omega_k^N} \|\mathbf{a}_{i,\Omega}\|_2^2$$

where Ω_k^N is the set of all k-subsets of $\{1,2,\cdots,N\}$ and $\mathbf{a}_{i,\Omega}$ is the orthogonal projection of the i-th row of $\mathbf{A}$ onto the index set Ω. The extended mutual coherence can be used to bound the l_2 norm of $\mathbf{A}$'s rows projected onto any index set Ω with $|\Omega| \le k$.

Let $\mathcal{M} : \mathbb{C}^{N\times L} \mapsto \mathbb{C}^N$ be the operator such that $\hat{\mathbf{y}} = \mathcal{M}(\mathbf{Z}) = \{\mathbf{u}_i^\top \mathbf{Z}\psi_i\}_{i=1}^N$ where $\{\cdot\}$ is the elementwise definition of a vector. By using the inner product defined as $\langle \mathbf{X}, \mathbf{Y} \rangle = \operatorname{trace}(\mathbf{X}\mathbf{Y}^H)$ for all $\mathbf{X}, \mathbf{Y} \in \mathbb{C}^{N\times L}$, the adjoint operator $\mathcal{M}^* : \mathbb{C}^N \mapsto \mathbb{C}^{N\times L}$ is given by

$$\mathcal{M}^*(\mathbf{y}) = \sum_{i=1}^N y_i \overline{\mathbf{u}}_i \psi_i^H, \qquad \mathbf{y} \in \mathbb{C}^N,$$

and the operator $\mathcal{M}^*\mathcal{M} : \mathbb{C}^{N\times L} \mapsto \mathbb{C}^{N\times L}$ is given by

$$\mathcal{M}^*\mathcal{M}(\mathbf{Z}) = \sum_{i=1}^N \overline{\mathbf{u}}_i \mathbf{u}_i^\top \mathbf{Z} \psi_i \psi_i^H, \qquad \mathbf{Z} \in \mathbb{C}^{N\times L}.$$

Let $\mathbf{M} = (\mathbf{P}^\top * \mathbf{U}^\top)^\top \in \mathbb{C}^{N\times NL}$ be the linear transformation matrix such that $\hat{\mathbf{y}} = \mathbf{M}\operatorname{vec}(\mathbf{Z})$. Then we have

$$\begin{aligned} \operatorname{vec}(\mathcal{M}(\mathbf{Z})) &= \mathbf{M}\operatorname{vec}(\mathbf{Z}) \\ \operatorname{vec}(\mathcal{M}^*\mathcal{M}(\mathbf{Z})) &= \mathbf{M}^H\mathbf{M}\operatorname{vec}(\mathbf{Z}) \\ &= \sum_{i=1}^N (\overline{\psi}_i \psi_i^\top \otimes \overline{\mathbf{u}}_i \mathbf{u}_i^\top) \operatorname{vec}(\mathbf{Z}) \end{aligned}$$

For Γ_p defined in Assumption 3, we use the notations $\mathcal{M}_p$ and $\mathbf{M}_p$ to denote the projection of $\mathcal{M}$ and $\mathbf{M}$ to the index set Γ_p respectively, so that $\mathcal{M}_p(\mathbf{Z}) = \{\mathbf{u}_i^\top \mathbf{Z}\psi_i\}_{i\in\Gamma_p}$ and $\mathbf{M}_p \in \mathbb{C}^{Q\times LN}$ is the matrix representation of $\mathcal{M}_p$.

Since $\mathbf{x}$ is S-sparse, without loss of generality, we can assume that only the first S entries of $\mathbf{x}$ are non-zero. We use the notation Ω to denote the support of $\mathbf{x}$ and the indexes of $\mathbf{Z}$'s non-zero rows as well. $\mathbf{x}_\Omega$ and $\mathbf{Z}_\Omega$ represent the orthogonal projections of $\mathbf{x}$ and $\mathbf{Z}$ onto Ω respectively. Therefore, we have

$$\begin{aligned}\mathcal{M}_\Omega(\mathbf{Z}) &= \{\mathbf{u}_i^\top \mathbf{Z}_\Omega \psi_i\}_{i=1}^N \\ &= \{\mathbf{u}_{i,\Omega}^\top \mathbf{Z}_\Omega \psi_i\}_{i=1}^N \\ \mathcal{M}_\Omega^*(\mathbf{y}) &= \sum_{i=1}^N y_i \bar{\mathbf{u}}_{i,\Omega} \psi_i^H\end{aligned}$$

Then the operators $\mathcal{M}_{p,\Omega}$ and $\mathcal{M}_{p,\Omega}^*$ can be defined based on the above definitions.

B. Theorem and Corollary

In [8], the authors have provided the probabilistic recovery guarantee for Problem (P.4). First, their conclusion is based on a noiseless setting. So the assumptions are slightly different from ours. Assumption 2 is replaced by Assumption 4 to express a different underlying model.

Assumption 4: The frequency representation of the observed graph signal y adheres to the model $\hat{y} = \mathrm{diag}(\mathbf{P}\mathbf{h}')\widetilde{\mathbf{U}}\mathbf{x}$, where $\widetilde{\mathbf{U}}$ is a random $N \times N$ matrix obtained by concatenating N rows sampled independently and uniformly with replacement from $\mathbf{U}$. Also, $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$ and $\mathbf{P}^H\mathbf{P} = \mathbf{I}_L$.
We restate their main result as follows.

Theorem 1: For a given graph-shift operator $\mathbf{S}$, assume that an S-sparse graph signal $\mathbf{x} \in \mathbb{R}^N$ when passed through a filter with coefficients $\mathbf{h}' \in \mathbb{R}^L$ results in a signal with frequency representation $\hat{\mathbf{y}} \in \mathbb{C}^N$ adhering to the model in Assumption 4. Also, denote by $\mathbf{U} \in \mathbb{C}^{N\times N}$ and $\mathbf{P} \in \mathbb{C}^{N\times L}$ the GFT for signals and filters associated with $\mathbf{S}$, respectively, where $\mathbf{U}$ is normalized such that $\mathbf{U}^H\mathbf{U} = N\mathbf{I}_N$. Define

$$\alpha = \frac{3\ln 2\left(120\dfrac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{P}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{P}}(L)} + 8\sqrt{\dfrac{\rho_{\mathbf{U}}(1)\rho_{\mathbf{P}}(1)LS}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{P}}(L)}}\right)^{-1}}{\rho_{\mathbf{U}}(S)\rho_{\mathbf{P}}(L)\ln(4\gamma\sqrt{2LS})\ln(2SN^2)}$$

where $\gamma = \sqrt{2N(\ln(2LN)+1)+1}$. Under Assumptions 1, 3 and 4, if $\alpha \geq 1$ then the unique solution to Problem (P.4) is the rank-one matrix $\mathbf{Z} = \mathbf{x}\mathbf{h}'^\top$, with probability at least $1 - N^{1-\alpha}$.

However, the above bound is not elegant enough and relatively loose, which cannot guarantee that $\alpha \geq 1$. Also, in most real-world problems, the observations often contain noise. Therefore, we want to find out a recovery guarantee for Problem (P.7) with an elegant and tight probabilistic lower bound. Meanwhile, we want to know when the lower bound is valid. These two ideas give rise to our main results - the following theorem and corollary.

Theorem 2: For a given graph-shift operator $\mathbf{S}$, assume that $\mathbf{x} \in \mathbb{R}^N$ is an S-sparse graph signal, $\mathbf{h}' \in \mathbb{R}^L$ is a filter coefficient vector and $\hat{\mathbf{y}} \in \mathbb{C}^N$ is the frequency representation of the output signal adhering to the model in Assumption 2. Also, $\mathbf{U} \in \mathbb{C}^{N\times N}$ and $\mathbf{P} \in \mathbb{C}^{N\times L}$ are the graph Fourier transformation for signals and filters associated with $\mathbf{S}$, respectively. Under Assumptions 1, 2 and 3, the solution $\hat{\mathbf{Z}}$ given by Problem (P.7) satisfies

$$\|\hat{\mathbf{Z}} - \mathbf{x}\mathbf{h}'^\top\|_F \leq \left(\frac{12(\sqrt{6}\gamma + \tau\sqrt{LS})}{\frac{1}{2} - \frac{1}{2\sqrt{2}\gamma}} + 2\sqrt{6}\right)\eta$$

where $\gamma = \sqrt{2N(\ln(2LN)+1)+1}$ and $\tau = 2\sqrt{\frac{3\ln(4\gamma\sqrt{2LS})}{2\ln 2}}$, with probability at least $1-\alpha$ and

$$\alpha = 8PL^2S^2\exp\left(-\frac{5}{4P\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}h\left(\frac{1}{5\sqrt{LS}}\right)\right)$$

where $h(u) = (1+u)\ln(1+u) - u$ for $u \geq 0$, if $\alpha \leq 1$ and $P \geq \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}$.

Corollary 1: The conditions $\alpha \leq 1$ and $P \geq \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}$ in Theorem 2 hold, if the following inequalities hold,

$$\begin{aligned}N &\geq \max\left\{\frac{1}{\ln 2}\left(C_1 + \sqrt{2(C_1+1)}\right),\right. \\ &\qquad \left.\frac{C_3}{\ln(C_2C_3) - \ln[\ln(C_2C_3)]}\right\} \\ C_2C_3 &\geq e\end{aligned}$$

where

$$\begin{aligned}C_1 &= -\ln(\ln 2) + \frac{3}{2}\ln 2 + \frac{1}{2}\ln(2LS) \\ C_2 &= 8L^2S^2 \\ C_3 &= \frac{5}{4\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}h\left(\frac{1}{5\sqrt{LS}}\right).\end{aligned}$$

Notice, similar results can also be obtained in the absence of noise, but this paper focuses only on noisy situations. The proof of Theorem 2 and Corollary 2 will be presented in Section 4. To prove Theorem 2, we need the following ingredients.

C. Main Ingredients

1) Exact Recovery of Deterministic Matrix: The first ingredient is Proposition 2, which gives deterministic conditions for the error bound between the solution $\hat{\mathbf{Z}}$ of the proposed algorithm and the ground truth $\mathbf{Z}$ (Theorem 4.33 in [19]).

Proposition 2: Let $\widetilde{\mathbf{m}}_{ij}$ $(1 \leq i \leq L,\ 1 \leq j \leq N)$ be the columns of $\mathbf{M} \in \mathbb{C}^{N\times NL}$ such that

$$\mathbf{M} = \begin{pmatrix}\widetilde{\mathbf{m}}_{11} & \cdots & \widetilde{\mathbf{m}}_{L1} & \cdots & \widetilde{\mathbf{m}}_{1N} & \cdots & \widetilde{\mathbf{m}}_{LN}\end{pmatrix},$$

$\mathbf{Z} \in \mathbb{C}^{N\times L}$ with support on Ω and $\mathbf{y} = \mathcal{M}(\mathbf{Z}) + \omega$ with $\|\omega\|_2 \leq \eta$. For $\gamma, \tau \geq 0$,

$$\|\mathcal{M}_\Omega^*\mathcal{M}_\Omega - \mathbf{I}_\Omega\| \leq \frac{1}{2} \tag{a}$$

$$\max_{\substack{1\leq i\leq L\\ j>s}} \|\mathbf{M}_\Omega^H \widetilde{\mathbf{m}}_{ij}\|_2 \leq 1 \tag{b}$$

and that there exists a $\mathbf{z} = \mathcal{M}^*(\mathbf{y}) \in \mathrm{range}(\mathcal{M}^*)$ with $\mathbf{y} \in \mathbb{C}^{L\times 1}$ such that

$$\|\mathbf{y}\|_2 \leq \tau\sqrt{LS} \tag{c}$$

$$\|\mathbf{z}_\Omega - \mathrm{sign}(\mathbf{Z})\|_F \leq \frac{1}{4\sqrt{2}\gamma} \tag{d}$$

$$\|\mathbf{z}_{\Omega^\perp}\|_\infty \leq \frac{1}{2} \tag{e}$$

then the minimizer $\hat{\mathbf{Z}}$ to the problem that

$$\begin{aligned} \min_{\mathbf{Z}} \quad & \|\mathbf{Z}\|_1 \\ \text{s.t.} \quad & \|\mathbf{y} - \mathcal{M}(\mathbf{Z})\|_2 \le \eta \end{aligned}$$

satisfies

$$\|\hat{\mathbf{Z}} - \mathbf{Z}\|_F \le \left(\frac{12(\sqrt{6}\gamma + \tau\sqrt{LS})}{\frac{1}{2} - \frac{1}{2\sqrt{2}\gamma}} + 2\sqrt{6} \right) \eta.$$

We will later use $\gamma = \sqrt{2N(\ln(2LN)+1)+1}$ and $\tau = 2\sqrt{\frac{3\ln(4\gamma\sqrt{2LS})}{2\ln 2}}$ in the following content.

2) Noncommutative Bernstein Inequality for Matrices: By leveraging the following well-known theorem - noncommutative Bernstein inequality for matrices, we can prove that the conditions in Proposition 2 can be satisfied with a certain probability under Assumptions 1, 2 and 3. Notice that, here we use a more tight bound (Theorem 6.1 in [20]) than the bound used in [8], [9].

Proposition 3: Consider a finite sequence $\{\mathbf{Z}_i\}_{i\in\Gamma}$ of independent centered random matrices with dimension $M \times M$. Assume that $\|\mathbf{Z}_i\| \le R$ almost surely for all i and introduce the random matrix $\mathbf{\Sigma} = \sum_{i\in\Gamma} \mathbf{Z}_i$ with variance parameter

$$\sigma^2 = \max\left\{ \left\| \sum_{i\in\Gamma} \mathbb{E}(\mathbf{Z}_i\mathbf{Z}_i^H) \right\|, \left\| \sum_{i\in\Gamma} \mathbb{E}(\mathbf{Z}_i^H\mathbf{Z}_i) \right\| \right\}.$$

Then for all $t \ge 0$,

$$\mathbb{P}(\|\mathbf{\Sigma}\| \ge t) \le 2M \exp\left(-\frac{\sigma^2}{R^2} h\left(\frac{Rt}{\sigma^2}\right) \right).$$

where $h(u) = (1+u)\ln(1+u) - u$ for $u \ge 0$.

Notice that in Proposition 3, the upper bound of $\mathbb{P}(\|\mathbf{\Sigma}\| \ge t)$ is increasing with respect to the variable σ^2. Thus, one can achieve a similar result for the sequence $\{\mathbf{Z}_i\}_{i\in\Gamma}$ by finding the upper bounds of $\max_{i\in\Gamma}\|\mathbf{Z}_i\|$ and σ^2.

3) Extended Mutual Coherence: The last ingredient is the inequality related to the extended mutual coherence (Lemma 1 in [8]). The inequalities about $\rho_{\mathbf{U}}(S)$ and $\rho_{\mathbf{P}}(L)$ will be used implicitly in the proof of our main results.

Proposition 4: For $\mathbf{U}$ and $\mathbf{P}$ defined in the Assumption 2, we have

$$\begin{aligned} S \le & \rho_{\mathbf{U}}(S) \le S\rho_{\mathbf{U}}(1) \\ \frac{L}{N} \le & \rho_{\mathbf{P}}(L) \le L\rho_{\mathbf{P}}(1) \end{aligned}$$

From the assumptions that $\mathbf{U}^H\mathbf{U} = \mathbf{I}_N$ and $\mathbf{P}^H\mathbf{P} = \mathbf{I}_L$, we also have $\rho_{\mathbf{U}}(S) \le N$ and $\rho_{\mathbf{P}}(L) \le 1$.

IV. Proof of the Main Results

A. Techinical Lemma

Here we first prove two useful lemmas. Lemma 1 and Lemma 2 are more general results of condition (a) in Proposition 2.

Lemma 1: For any fixed $\delta \in (0,1]$ and partition $\{\Gamma_p\}_{p=1}^P$ of $\{1,2,\cdots,N\}$ with $|\Gamma_p| = Q$ for all p, and defining $\mathbf{T}_p = \sum_{i\in\Gamma_p} \overline{\psi_i\psi_i^\top}$, then we have

$$\max_{1\le p\le P} \sup_{\|\mathbf{Z}_\Omega\|_F=1} \|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z}) - \mathbf{Z}_\Omega\mathbf{T}_p\|_F \le \frac{\delta Q}{N}$$

with probability at least $1-\alpha$ where

$$\alpha = 2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)} h\left(\frac{4}{5}\delta\right) \right)$$

if $\alpha \le 1$.

Proof: Since

$$\begin{aligned} & \mathrm{vec}(\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z}) - \mathbf{Z}_\Omega\mathbf{T}_p) \\ &= \mathrm{vec}(\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z})) - \mathrm{vec}(\mathbf{Z}_\Omega\mathbf{T}_p) \\ &= \mathbf{M}_{p,\Omega}^H\mathbf{M}_{p,\Omega}\,\mathrm{vec}(\mathbf{Z}) - \sum_{i\in\Gamma_p} (\overline{\psi_i\psi_i^\top} \otimes \mathbf{I}_{N,\Omega})\,\mathrm{vec}(\mathbf{Z}) \\ &= \sum_{i\in\Gamma_p} (\overline{\psi_i\psi_i^\top} \otimes \bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top)\,\mathrm{vec}(\mathbf{Z}) - \sum_{i\in\Gamma_p} (\overline{\psi_i\psi_i^\top} \otimes \mathbf{I}_{N,\Omega})\,\mathrm{vec}(\mathbf{Z}) \\ &= \sum_{i\in\Gamma_p} [\overline{\psi_i\psi_i^\top} \otimes (\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega} - \mathbf{I}_{N,\Omega})]\,\mathrm{vec}(\mathbf{Z}) \end{aligned}$$

we define $\mathbf{\Upsilon}_i = (\overline{\psi_i\psi_i^\top}) \otimes (\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top - \mathbf{I}_{N,\Omega}) \in \mathbb{C}^{LN\times LN}$. It follows that the spectral norm is given by

$$\sup_{\|\mathbf{Z}_\Omega\|_F=1} \|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z}) - \mathbf{Z}_\Omega\mathbf{T}_p\|_F = \left\| \sum_{i\in\Gamma_p} \mathbf{\Upsilon}_i \right\|$$

Notice that $\mathbf{\Upsilon}_i$, $i = 1,2,\cdots,N$ are centered matrices and Hermitian matrices since

$$\begin{aligned} \mathbb{E}\mathbf{\Upsilon}_i &= (\overline{\psi_i\psi_i^\top}) \otimes \mathbb{E}(\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top - \mathbf{I}_{N,\Omega}) \\ &= (\overline{\psi_i\psi_i^\top}) \otimes \mathbf{0}_N \\ &= \mathbf{0}_{LN} \\ \mathbf{\Upsilon}_i^H &= [(\overline{\psi_i\psi_i^\top}) \otimes (\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top - \mathbf{I}_{N,\Omega})]^H \\ &= (\overline{\psi_i\psi_i^\top})^H \otimes (\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top - \mathbf{I}_{N,\Omega})^H \\ &= (\overline{\psi_i\psi_i^\top}) \otimes (\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top - \mathbf{I}_{N,\Omega}) \\ &= \mathbf{\Upsilon}_i \end{aligned}$$

Therefore, we just need to find suitable $R \ge \max_i \|\mathbf{\Upsilon}_i\|$ and $\hat{\sigma}^2 \ge \sigma^2$ and then use Proposition 3 to find the probabilistic bound on $\left\|\sum_{i\in\Gamma_p} \mathbf{\Upsilon}_i\right\|$.

The spectral norm of Kronecker product is given by

$$\begin{aligned} \|\mathbf{\Upsilon}_i\| &= |\overline{\psi_i\psi_i^\top}|\|\bar{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top - \mathbf{I}_{N,\Omega}\| \\ &\le \rho_{\mathbf{P}}(L)\max\{|\bar{\mathbf{u}}_{i,\Omega}^\top\mathbf{u}_{i,\Omega} - 1|, 1\} \\ &\le \rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1) \end{aligned}$$

because of the fact that $\rho_{\mathbf{U}}(S) \ge S \ge 1$. Therefore, we can choose $R = \rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)$.

Next, we find a bound of σ^2,

$$\begin{aligned}\sigma^2 &= \left\|\sum_{i\in\Gamma_p}\mathbb{E}(\boldsymbol{\Upsilon}_i\boldsymbol{\Upsilon}_i^H)\right\| \\ &= \left\|\sum_{i\in\Gamma_p}\mathbb{E}\left[(\overline{\psi_i}\psi_i^\top\overline{\psi_i}\psi_i^\top)\otimes[(\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top-\mathbf{I}_{N,\Omega})^2]\right]\right\| \\ &\le \rho_{\mathbf{P}}(L) \\ &\quad \left\|\sum_{i\in\Gamma_p}(\overline{\psi_i}\psi_i^\top)\otimes\mathbb{E}[\rho_{\mathbf{U}}(S)\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top-2\overline{\mathbf{u}}_{i,\Omega}\mathbf{u}_{i,\Omega}^\top+\mathbf{I}_{N,\Omega}]\right\| \\ &= \rho_{\mathbf{P}}(L)\left\|\sum_{i\in\Gamma_p}(\overline{\psi_i}\psi_i^\top)\otimes[(\rho_{\mathbf{U}}(S)-1)\mathbf{I}_{N,\Omega}]\right\| \\ &= \rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)\|\mathbf{T}_p\| \\ &\le \rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)\frac{5Q}{4N}\end{aligned}$$

where the last inequality comes from (2).

Notice that $\boldsymbol{\Upsilon}_i$ can be expressed as a $LS\times LS$ matrix since $|\Omega|=S$. We obtain

$$\begin{aligned}&\mathbb{P}\left(\left\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\right\|\ge\frac{\delta Q}{N}\right) \\ \le& 2LS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{4}{5}\delta\right)\right)\end{aligned}$$

thus,

$$\begin{aligned}&\mathbb{P}\left(\max_{1\le p\le P}\left\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\right\|\le\frac{\delta Q}{N}\right) \\ &\ge 1-P\mathbb{P}\left(\left\|\sum_{i\in\Gamma_p}\boldsymbol{\Upsilon}_i\right\|\ge\frac{\delta Q}{N}\right) \\ &\ge 1-2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{4}{5}\delta\right)\right)\end{aligned}$$

By setting $\alpha=2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{4}{5}\delta\right)\right)$, if $\alpha\in[0,1]$, the conclusion holds naturaly. ■

Lemma 2: Under Assumptions 1, 2 and 3, we have

$$\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_F=1}\|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z})-\frac{Q}{N}\mathbf{Z}_\Omega\|_F\le\frac{Q}{2N}$$

with probability at least $1-\alpha$ where

$$\alpha=2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{1}{5}\right)\right)$$

if $\alpha\le 1$.

Proof: From Lemma 1, by setting $\delta=\frac{1}{4}$, we have

$$\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_F=1}\|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z})-\mathbf{Z}_\Omega\mathbf{T}_p\|_F\le\frac{Q}{4N}$$

with probability at least $1-\alpha$ where

$$\alpha=2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{1}{5}\right)\right)$$

if $\alpha\le 1$. ByAssumption 3, we have

$$\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_F=1}\|\mathbf{Z}_\Omega\mathbf{T}_p-\frac{Q}{N}\mathbf{Z}_\Omega\|_F=\max_{1\le p\le P}\|\mathbf{T}_p-\frac{Q}{N}\mathbf{I}_L\|\le\frac{Q}{4N}$$

Thereore,

$$\begin{aligned}&\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_F=1}\|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z})-\frac{Q}{N}\mathbf{Z}_\Omega\|_F \\ \le&\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_F=1}\|\mathcal{M}_{p,\Omega}^*\mathcal{M}_{p,\Omega}(\mathbf{Z})-\mathbf{Z}_\Omega\mathbf{T}_p\|_F \\ &+\max_{1\le p\le P}\sup_{\|\mathbf{Z}_\Omega\|_F=1}\|\mathbf{Z}_\Omega\mathbf{T}_p-\frac{Q}{N}\mathbf{Z}_\Omega\|_F \\ \le&\frac{Q}{4N}+\frac{Q}{4N} \\ =&\frac{Q}{2N}\end{aligned}$$

with probability at least $1-\alpha$ where

$$\alpha=2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{1}{5}\right)\right)$$

if $\alpha\le 1$. ■

B. Local isometry property

Condition (a) and (b) in Proposition 2 corresponse to the local isometry preoperty, which means that the operator $\mathcal{M}_\Omega^*\mathcal{M}_\Omega$ differs only a little from the identity operator.

Lemma 3: Under Assumptions 1, 2 and 3, condition (a) holds with probability at least $1-\alpha_a$ where

$$\alpha_a=2LS\exp\left(-\frac{5}{4\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{2}{5}\right)\right)\quad(2)$$

if $\alpha_a\le 1$.

Proof: We choose $\delta=\frac{1}{2}$ and $P=1$ in Lemma 1. $P=1$ implies $Q=N$ and $\mathbf{T}_1=\mathbf{I}_L$. Therefore,

$$\|\mathcal{M}_\Omega^*\mathcal{M}_\Omega-\mathbf{I}_\Omega\|\le\frac{1}{2}$$

with probability at least $1-\alpha_a$ if $\alpha_a\le 1$. ■

To consider condition (b), we need the following proposition.

Proposition 5: For any matrix $\mathbf{X}\in\mathbb{C}^{M\times N}$, any entry X_{ij} is bounded by its operator norm $\|\mathbf{X}\|$, i.e.,

$$\max_{\substack{1\le i\le M\\1\le j\le N}}|X_{ij}|\le\|\mathbf{X}\|.$$

Lemma 4: Under Assumptions 1, 2 and 3, condition (b) holds with probability at least $1-\alpha_b$ where

$$\alpha_b=4L^2S^2\exp\left(-\frac{5}{4\rho_{\mathbf{P}}(2)(\rho_{\mathbf{U}}(2)-1)}h\left(\frac{4}{5\sqrt{LS}}\right)\right)\quad(3)$$

if $\alpha_b\le 1$.

Proof: Let $\widetilde{\mathbf{m}}_{ij}$ $(1\le i\le L,\ 1\le j\le N)$ denotes the column of $\mathbf{M}$. Due to the fact that for $1\le i\le L$ and $j\ge S$,

$$\begin{aligned}\|\mathbf{M}_\Omega^H\widetilde{\mathbf{m}}_{ij}\|_2&\le\sqrt{\sum_{i'=1}^{L}\sum_{j'=1}^{S}\|\widetilde{\mathbf{m}}_{i'j'}^H\widetilde{\mathbf{m}}_{ij}\|_2^2} \\ &\le\sqrt{LS}\mu\end{aligned}$$

where $\mu=\max_{i_1\ne i_2\text{ or }j_1\ne j_2}|\langle\widetilde{\mathbf{m}}_{i_1,j_1},\widetilde{\mathbf{m}}_{i_2,j_2}\rangle|$ is the coherence, i.e., the largest correlation between two different columns of $\mathbf{M}$, we just need to find a suitable bound of μ.

For any two different columns of $\mathbf{M}$, $\widetilde{\mathbf{m}}_{i_1,j_1}$ and $\widetilde{\mathbf{m}}_{i_2,j_2}$ where $1 \le i_1, i_2 \le L$, $1 \le j_1, j_2 \le N$ and $(i_1, j_1) \neq (i_2, j_2)$. Let $\widetilde{\Omega} = \{(i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\}$, then $|\widetilde{\Omega}| \le 4$. Let $\mathbf{M}_{\widetilde{\Omega}}$ denote the orthogonal projections of $\mathbf{M}$'s columns onto $\widetilde{\Omega}$ for convenience, then $\mathbf{M}_{\widetilde{\Omega}}$ contains $\widetilde{\mathbf{m}}_{i_1,j_1}$ and $\widetilde{\mathbf{m}}_{i_2,j_2}$. From the proof of Lemma 1, by setting $Q = N$ we have $\|\mathbf{M}^*_{\widetilde{\Omega}}\mathbf{M}_{\widetilde{\Omega}} - \mathbf{I}_{\widetilde{\Omega}}\| \le \delta$ with probability at least $1-\alpha$ where

$$\alpha = 8\exp\left(-\frac{5}{4\rho_{\mathbf{P}}(2)(\rho_{\mathbf{U}}(2)-1)}h\left(\frac{4}{5}\delta\right)\right)$$

if $\alpha \le 1$. Therefore, by Proposition 5 we have $|\langle \widetilde{\mathbf{m}}_{i_1,j_1}, \widetilde{\mathbf{m}}_{i_2,j_2}\rangle| \le \delta$ with probability at least $1-\alpha$ if $\alpha \le 1$.

Then we take the union bound over $\frac{(LS)^2}{2}$ pairs of different indices (i_1, j_1) and (i_2, j_2). Thus μ satisfies $\mu \le \frac{1}{\sqrt{LS}}$ with probability at least $1 - \frac{(LS)^2}{2}\alpha$ and $\delta = \frac{1}{\sqrt{LS}}$ if $\alpha \le 1$.

Therefore,

$$\max_{\substack{1\le i\le L\\ j>s}} \|\mathbf{M}^H_{\Omega}\widetilde{\mathbf{m}}_{ij}\|_2 \le 1$$

with probability at least $1-\alpha_b$ with

$$\alpha_b = 4L^2S^2\exp\left(-\frac{5}{4\rho_{\mathbf{P}}(2)(\rho_{\mathbf{U}}(2)-1)}h\left(\frac{4}{5\sqrt{LS}}\right)\right)$$

if $\alpha_b \le 1$. ■

C. Construction of Inexact Dual Certificate

For conditions (c), (d) and (e) in Proposition 2, we construct inexact dual certificate by golfing scheme proposed in [21]. By this method, we explicitly construct a sequence of random matrices, use independent samples for each iteration to decouple statistical dependency and show the convergence to $\mathrm{sign}(\mathbf{Z})$ where $\mathrm{sign}(\cdot)$ is a elementwise sign function.

Lemma 5: Under Assumptions 1, 2 and 3, condition (c) and (d) hold with probability at least $1-\alpha_c$ where

$$\alpha_c = \alpha_d = 2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{1}{5}\right)\right)$$

if $\alpha_c \le 1$ and $P \ge \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}$.

Proof: We use golfing scheme to construct the dual certificate $\mathbf{z}$ as follows. The process aims to produce a sequence of random matrices $\{\mathbf{z}_p \in \mathrm{range}(\mathcal{M}^*) | p = 0, 1, \cdots, P\}$ such that $\mathbf{z}_p$ converges to $\mathrm{sign}(\mathbf{Z})$ exponentially fast and each entry of $\mathbf{z}_p$ in $\Omega^\perp$ remains small. We set $\mathbf{z}_0 = \mathbf{0}$ and

$$\mathbf{z}_p = \mathbf{z}_{p-1} + \frac{N}{Q}\mathcal{M}^*_p\mathcal{M}_p(\mathrm{sign}(\mathbf{Z}) - \mathbf{z}_{p-1,\Omega})$$

Thus, the residuals between $\mathrm{sign}(\mathbf{Z})$ and $\mathbf{z}_p$ on Ω is given by $\mathbf{W}_p = \mathbf{z}_{p-1,\Omega} - \mathrm{sign}(\mathbf{Z})$ with $\mathbf{W}_0 = -\mathrm{sign}(\mathbf{Z})$ and $\|\mathbf{W}_0\|_F \le \sqrt{LS}$. Then for $p \ge 1$,

$$\begin{aligned}\mathbf{W}_p &= (\mathbf{z}_{p-1,\Omega} - \mathbf{z}_{p-2,\Omega}) + (\mathbf{z}_{p-2,\Omega} - \mathrm{sign}(\mathbf{Z}))\\ &= \mathbf{W}_{p-1} - \frac{N}{Q}\mathcal{M}^*_{p,\Omega}\mathcal{M}_{p,\Omega}(\mathbf{W}_{p-1})\\ &= \frac{N}{Q}\left(\frac{Q}{N} - \mathcal{M}^*_{p,\Omega}\mathcal{M}_{p,\Omega}\right)(\mathbf{W}_{p-1})\end{aligned}$$

By Lemma 2, with probability at least $1-\alpha$ where

$$\alpha = 2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{1}{5}\right)\right)$$

if $\alpha \le 1$, we have

$$\begin{aligned}\|\mathbf{W}_p\|_F &= \frac{N}{Q}\left\|\left(\frac{Q}{N} - \mathcal{M}^*_{p,\Omega}\mathcal{M}_{p,\Omega}\right)(\mathbf{W}_{p-1})\right\|_F\\ &\le \frac{N}{Q}\left\|\frac{Q}{N} - \mathcal{M}^*_{p,\Omega}\mathcal{M}_{p,\Omega}\right\|\|\mathbf{W}_{p-1}\|_F\\ &\le \frac{1}{2}\|\mathbf{W}_{p-1}\|_F\end{aligned}$$

where the first inequality comes from the compatible property of the operator norm with the l_2 vector norm since $\|\mathbf{W}_{p-1}\|_F = \|\mathrm{vec}(\mathbf{W}_{p-1})\|_2$. Hence,

$$\begin{aligned}\|\mathbf{W}_p\|_F &\le 2^{-p}\|\mathbf{W}_0\|_F\\ &\le 2^{-p}\sqrt{LS}\end{aligned}$$

Set $\mathbf{z} = \mathbf{z}_P$. To achieve $\|\mathbf{W}_P\|_F = \|\mathbf{z}_\Omega - \mathrm{sign}(\mathbf{Z})\| \le \frac{1}{4\sqrt{2}\gamma}$, we just need $2^{-P}\sqrt{LS} \le \dfrac{1}{4\sqrt{2}\gamma}$, i.e.,

$$P \ge \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}.$$

Since $\mathbf{z} = -\dfrac{N}{Q}\sum_{p=1}^{P}\mathcal{M}^*_p\mathcal{M}_p(\mathbf{W}_{p-1})$, we have

$$\mathbf{y} = -\frac{N}{Q}\begin{pmatrix}\mathcal{M}_1(\mathbf{W}_0) & \mathcal{M}_2(\mathbf{W}_1) & \cdots & \mathcal{M}_P(\mathbf{W}_{P-1})\end{pmatrix}^\top$$

Therefore,

$$\begin{aligned}\|\mathbf{y}\|_2 &= \sqrt{\frac{N^2}{Q^2}\sum_{p=1}^{P}\|\mathcal{M}_p(\mathbf{W}_{p-1})\|_F^2}\\ &= \sqrt{\frac{N^2}{Q^2}\sum_{p=1}^{P}\langle\mathcal{M}_p(\mathbf{W}_{p-1}), \mathcal{M}_p(\mathbf{W}_{p-1})\rangle}\\ &= \sqrt{\frac{N^2}{Q^2}\sum_{p=1}^{P}\langle\mathcal{M}^*_p\mathcal{M}_p(\mathbf{W}_{p-1}), \mathbf{W}_{p-1}\rangle}\\ &\le \sqrt{\frac{N^2}{Q^2}\cdot\frac{3Q}{2N}\sum_{p=1}^{P}\|\mathbf{W}_{p-1}\|_F^2}\\ &\le \sqrt{\frac{3N}{2Q}\sum_{p=1}^{P}4^{-p+1}LS}\\ &\le \sqrt{\frac{3N}{2Q}}\sum_{p=1}^{P}2^{-p+1}\sqrt{LS}\\ &\le \sqrt{\frac{3N}{2Q}}(2-2^{1-P})\sqrt{LS}\\ &\le 2\sqrt{\frac{3}{2}P}\sqrt{LS}\end{aligned}$$

The choices of $\tau = 2\sqrt{\frac{3\ln(4\sqrt{2LS}\gamma)}{2\ln 2}}$ and $P = \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}$ complete the proof. ■

Lemma 6: Under Assumptions 1, 2 and 3, condition (e) holds with probability at least $1-\alpha_e$ where

$$\alpha_e = \alpha_c + \alpha - \alpha_c\alpha \tag{4}$$

and

$$\alpha = 2P\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}h\left(\frac{\sqrt{\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}}{5\sqrt{\rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)LS}}\right)\right)$$

if $\alpha_e \geq 1$.

Proof: Notice that

$$\mathbf{z} = -\frac{N}{Q}\sum_{p=1}^{P}\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})$$

we just need to find the bound of

$$\mathbb{P}\left(\|[\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})]_{\Omega^\perp}\|_\infty > \frac{Q}{2^{p+1}N}\right)$$

under the condition of $\|\mathbf{W}_p\|_F \leq 2^{-p}\sqrt{LS}$.

$$\|[\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})]_{\Omega^\perp}\|_\infty = \max_{s\in\Omega^\perp}|\langle\mathbf{e}_s, \mathbf{M}_p^H\mathbf{M}_p\,\mathrm{vec}(\mathbf{W}_{p-1})\rangle|$$

where $\mathbf{e}_s \in \mathbb{R}^{LN}$ is a unit vector with its s-th entry being 1 ($s > NS$). For any $s\in\Omega^\perp$,

$$\begin{aligned}\langle\mathbf{e}_s, \mathbf{M}_p^H\mathbf{M}_p\,\mathrm{vec}(\mathbf{W}_{p-1})\rangle &= \langle\mathbf{e}_s, \sum_{i\in\Gamma_p}(\overline{\psi}_i\psi_i^\top\otimes\bar{\mathbf{u}}_i\mathbf{u}_i^\top)\,\mathrm{vec}(\mathbf{W}_{p-1})\rangle\\ &= \sum_{i\in\Gamma_p}\langle\mathbf{e}_s, (\overline{\psi}_i\psi_i^\top\otimes\bar{\mathbf{u}}_i\mathbf{u}_i^\top)\,\mathrm{vec}(\mathbf{W}_{p-1})\rangle\end{aligned}$$

Let $a_i = \langle\mathbf{e}_s, (\overline{\psi}_i\psi_i^\top\otimes\bar{\mathbf{u}}_i\mathbf{u}_i^\top)\,\mathrm{vec}(\mathbf{W}_{p-1})\rangle$. Notice that

$$\begin{aligned}\mathbb{E}(a_i) &= \mathbf{e}_s^H(\overline{\psi}_i\psi_i^\top\otimes\mathbf{I}_N)\,\mathrm{vec}(\mathbf{W}_{p-1})\\ &= \mathbf{e}_s^H\,\mathrm{vec}(\mathbf{W}_{p-1}\overline{\psi}_i\psi_i^\top)\\ &= 0\end{aligned}$$

since the support of $\mathrm{vec}(\mathbf{W}_{p-1}\overline{\psi}_i\psi_i^\top)$ is the same as the support of $\mathbf{W}_{p-1} = \mathbf{z}_{p-1,\Omega} - \mathrm{sign}(\mathbf{Z})$ defined in the previous content. Next, we need to find the bounds of $|a_i|$ and $Var(a_i) = \mathbb{E}a_i^2$ so that Lemma 1 can be leveraged to show the final result.

The bound of $|a_i|$ is given by

$$\begin{aligned}|a_i| &= |\mathbf{e}_s^H(\overline{\psi}_i\psi_i^\top\otimes\bar{\mathbf{u}}_i\mathbf{u}_i^\top)\,\mathrm{vec}(\mathbf{W}_{p-1})|\\ &= |\mathbf{e}_s^H(\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)(\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)^H\,\mathrm{vec}(\mathbf{W}_{p-1})|\\ &\leq |\mathbf{e}_s^H(\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)|\cdot|(\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)^H\,\mathrm{vec}(\mathbf{W}_{p-1})|\\ &= |\mathbf{e}_s^H(\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)|\cdot|(\overline{\psi}_i\otimes\bar{\mathbf{u}}_{i,\Omega})^H\,\mathrm{vec}(\mathbf{W}_{p-1})| \qquad (6)\\ &\leq \|\overline{\psi}_i\otimes\bar{\mathbf{u}}_i\|_\infty\cdot\|\overline{\psi}_i\otimes\bar{\mathbf{u}}_{i,\Omega}\|_2\cdot\|\,\mathrm{vec}(\mathbf{W}_{p-1})\|_2\\ &\leq \|\overline{\psi}_i\|_\infty\cdot\|\bar{\mathbf{u}}_i\|_\infty\cdot\|\overline{\psi}_i\|_2\cdot\|\bar{\mathbf{u}}_{i,\Omega}\|_2\cdot\|\mathbf{W}_{p-1}\|_F\\ &= 2^{-p+1}\sqrt{\rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)LS}\end{aligned}$$

where (6) holds since $\mathbf{W}_{p-1}$ has support on Ω.

The bound of $\mathbb{E}a_i^2$ is given by

$$\begin{aligned}\sum_{i\in\Gamma_p}\mathbb{E}a_i^2 &= \sum_{i\in\Gamma_p}\mathbb{E}(\mathbf{e}_s^H(\overline{\psi}_i\psi_i^\top\otimes\bar{\mathbf{u}}_i\mathbf{u}_i^\top)\,\mathrm{vec}(\mathbf{W}_{p-1}))^2\\ &= \sum_{i\in\Gamma_p}\mathbb{E}[(\mathbf{e}_s^H(\overline{\psi}_i\otimes\bar{\mathbf{u}}_i))^2((\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)^H\,\mathrm{vec}(\mathbf{W}_{p-1}))^2]\\ &\leq \rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\sum_{i\in\Gamma_p}\mathbb{E}[((\overline{\psi}_i\otimes\bar{\mathbf{u}}_i)^H\,\mathrm{vec}(\mathbf{W}_{p-1}))^2]\\ &= \rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\\ &\quad\sum_{i\in\Gamma_p}\mathbb{E}[\mathrm{vec}(\mathbf{W}_{p-1})^H(\overline{\psi}_i\psi_i^\top\otimes\bar{\mathbf{u}}_i\mathbf{u}_i^\top)\,\mathrm{vec}(\mathbf{W}_{p-1})]\\ &= \rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\\ &\quad\sum_{i\in\Gamma_p}\mathrm{vec}(\mathbf{W}_{p-1})^H(\overline{\psi}_i\psi_i^\top\otimes\mathbf{I}_N)\,\mathrm{vec}(\mathbf{W}_{p-1})\\ &= \rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\,\mathrm{vec}(\mathbf{W}_{p-1})^H(\mathbf{T}_p\otimes\mathbf{I}_N)\,\mathrm{vec}(\mathbf{W}_{p-1})\\ &\leq \rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\|\mathbf{T}_p\|\|\mathbf{W}_{p-1}\|_F^2\\ &\leq \rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)\cdot\frac{5Q}{4N}\cdot 2^{-2(p-1)}LS\\ &= \frac{2^{-2p}\cdot 5QLS\rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)}{N}\end{aligned}$$

we apply the Bernstein inequality and get

$$\begin{aligned}&\mathbb{P}\left(\|[\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})]_{\Omega^\perp}\|_\infty > \frac{Q}{2^{p+1}N}\right)\\ \leq &2\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}h\left(\frac{\sqrt{\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}}{5\sqrt{\rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)LS}}\right)\right)\end{aligned}$$

under the condition of $\|\mathbf{W}_p\|_F \leq 2^{-p}\sqrt{LS}$, for all $1\leq p\leq P$. Therefore,

$$\mathbb{P}\left(\max_{1\leq p\leq P}\|[\mathcal{M}_p^*\mathcal{M}_p(\mathbf{W}_{p-1})]_{\Omega^\perp}\|_\infty \leq \frac{Q}{2^{p+1}N}\right) \geq 1-\alpha_e$$

where $\alpha_e = \alpha_c + \alpha - \alpha_c\alpha$ and

$$\alpha = 2P\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}h\left(\frac{\sqrt{\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}}{5\sqrt{\rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)LS}}\right)\right)$$

if $\alpha_e \leq 1$. ■

D. Proof of Theorem 2

Proof: From Lemma 3, 4, 5 and 6, we have conditions in Proposition 2 hold with probability at least $1-\alpha$ where

$$\alpha \geq \max\{\alpha_a, \alpha_b, \alpha_c, \alpha_d, \alpha_e\}$$

and

$$\begin{aligned}\alpha_a &= 2LS\exp\left(-\frac{5}{4\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{2}{5}\right)\right)\\ \alpha_b &= 4L^2S^2\exp\left(-\frac{5}{4\rho_{\mathbf{P}}(2)(\rho_{\mathbf{U}}(2)-1)}h\left(\frac{4}{5\sqrt{LS}}\right)\right)\\ \alpha_c = \alpha_d &= 2PLS\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)(\rho_{\mathbf{U}}(S)-1)}h\left(\frac{1}{5}\right)\right)\\ \alpha_e &= \alpha' + \alpha_c - \alpha'\alpha_c\\ \alpha' &= 2P\exp\left(-\frac{5Q}{4N\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}h\left(\frac{\sqrt{\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)}}{5\sqrt{\rho_{\mathbf{P}}(1)\rho_{\mathbf{U}}(1)LS}}\right)\right)\end{aligned}$$

where the last three equations hold only for $P \geq \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}$.

One choice of α could be

$$\alpha = 8PL^2S^2 \exp\left(-\frac{5}{4P\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)} h\left(\frac{1}{5\sqrt{LS}}\right)\right)$$

since $h(u)$ is increasing with respect to u. ■

E. Proof of Corollary 2

Proof: To achieve $N \geq P \geq \frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2}$, notice that

$$\begin{aligned}\frac{\ln(4\sqrt{2LS}\gamma)}{\ln 2} &= \frac{1}{\ln 2}\left(\ln 4 + \frac{1}{2}\ln(2LS) + \frac{1}{2}\ln\gamma^2\right)\\ \gamma^2 &= 2N[\ln(2LN)+1]+1\\ &= 2N\left[\ln(2N)+\ln L+1+\frac{1}{2N}\right]\\ &\leq 2N[\ln(2N)+\ln L+2]\end{aligned}$$

we have

$$\gamma^2 \leq 2\cdot 2N\ln(2N) \leq 2\cdot(2N)^2$$

where the first inequality holds with the condition of $\ln L + 2 \leq \ln(2N)$, i.e.,

$$N \geq \frac{e^2}{2}L \tag{5}$$

Therefore, we just need to require

$$N \geq \frac{1}{\ln 2}\left(\ln 4 + \frac{1}{2}\ln(2LS) + \frac{1}{2}\ln[2\cdot(2N)^2]\right)$$

By rearranging the above equation, we have

$$\exp(\ln 2\cdot N - C_1) \geq \ln 2\cdot N$$

where $C_1 = -\ln\left(\frac{\ln 2}{2}\right) + \frac{5\ln 2+\ln(2LS)}{2} = -\ln(\ln 2) + \frac{3}{2}\ln 2 + \frac{1}{2}\ln(2LS) > 1$. Solved for $\ln 2\cdot N$, we have

$$\ln 2\cdot N \geq -W_{-1}(-e^{-C_1}) \tag{6}$$

where $W_{-1}(z)$ is one of the branch of the Lambert function with $z \in \left[-\frac{1}{e}, 0\right]$, $W_{-1}(z) \in (-\infty, -1]$ and $W_{-1}(z)e^{W_{-1}(z)} = z$[22].

The lower bound of $W_{-1}(z)$ in [22] is given by

$$W_{-1}(-e^{-u}) > -u - \sqrt{2(u+1)}$$

for $u > 1$. Thus, inequality (8) holds if

$$\ln 2\cdot N \geq C_1 + \sqrt{2(C_1+1)}$$

i.e.

$$N \geq \frac{1}{\ln 2}\left(C_1 + \sqrt{2(C_1+1)}\right) \tag{7}$$

Next, we hope to find the condition for the establishment of $\alpha \leq 1$, i.e.

$$C_2 P \exp\left(-\frac{C_3}{P}\right) \leq 1$$

where $C_2 = 8L^2S^2 \geq 0$ and $C_3 = \frac{5}{\rho_{4\mathbf{P}}(L)\rho_{\mathbf{U}}(S)} h\left(\frac{1}{5\sqrt{LS}}\right) \geq 0$. Solved for P, we have

$$P \geq \frac{C_3}{W_0(C_2C_3)} \tag{8}$$

where $W_0(z)$ is the primal branch of the Lambert function with $z \in [-\frac{1}{e}, +\infty)$, $W_0(z) \geq -1$ and $W_0(z)e^{W_0(z)} = z$.

The lower bound of $W_0(z)$ in [23] is given by

$$W_0(z) \geq \ln z - \ln(\ln z)$$

for $z \geq e$. Thus, inequality (10) holds if

$$P \geq \frac{C_3}{\ln(C_2C_3) - \ln[\ln(C_2C_3)]} \tag{9}$$

and

$$C_2C_3 \geq e \tag{10}$$

By combining the results in (7)-(12), we have conditions

$$\begin{aligned}N &\geq \max\left\{\frac{1}{\ln 2}\left(C_1+\sqrt{2(C_1+1)}\right), \frac{C_3}{\ln(C_2C_3)-\ln[\ln(C_2C_3)]}\right\}\\ C_2C_3 &\geq e\end{aligned}$$

where

$$\begin{aligned}C_1 &= -\ln(\ln 2) + \frac{3}{2}\ln 2 + \frac{1}{2}\ln(2LS)\\ C_2 &= 8L^2S^2\\ C_3 &= \frac{5}{4\rho_{\mathbf{P}}(L)\rho_{\mathbf{U}}(S)} h\left(\frac{1}{5\sqrt{LS}}\right)\end{aligned}$$

■

V. Numerical Results

To examine the effectiveness of our model, we carry out the following experiments.

A. Recovery Performance on Different Graphs

First, we examine the recovery performance of the proposed convex-relaxed algorithms on two different kinds of graphs. The first one is a kind of random graphs called Erd-sRnyi model[24] denoted by $G(n,p)$, which is constructed by connecting edges among n nodes with probability p independently. In the experiments, we choose $n = 50$ and $p = 0.1$. In each experiment, we randomly generate a graph, choose different values of S and L, simulate input and output signals, solve the optimization problem and repeat this process for 50 times. The observed output signal contains noise $\mathbf{y}_{\text{obs}} = \mathbf{y} + \mathbf{w}$ where $\|\mathbf{w}\|_2 = \eta = 0.01$. Then we calculate the average recovery accuracy for each different setting, where one experiment is considered as a success if $RMSE = \sqrt{\frac{1}{NL}\|\mathbf{Z} - \hat{\mathbf{Z}}\|_F^2} \leq 0.01$.

The other graph is a real-world social network called Zachary's karate club[25] containing 34 nodes and 78 edges. We use a similar process as the one for the random graphs while here we fix the graph structure.

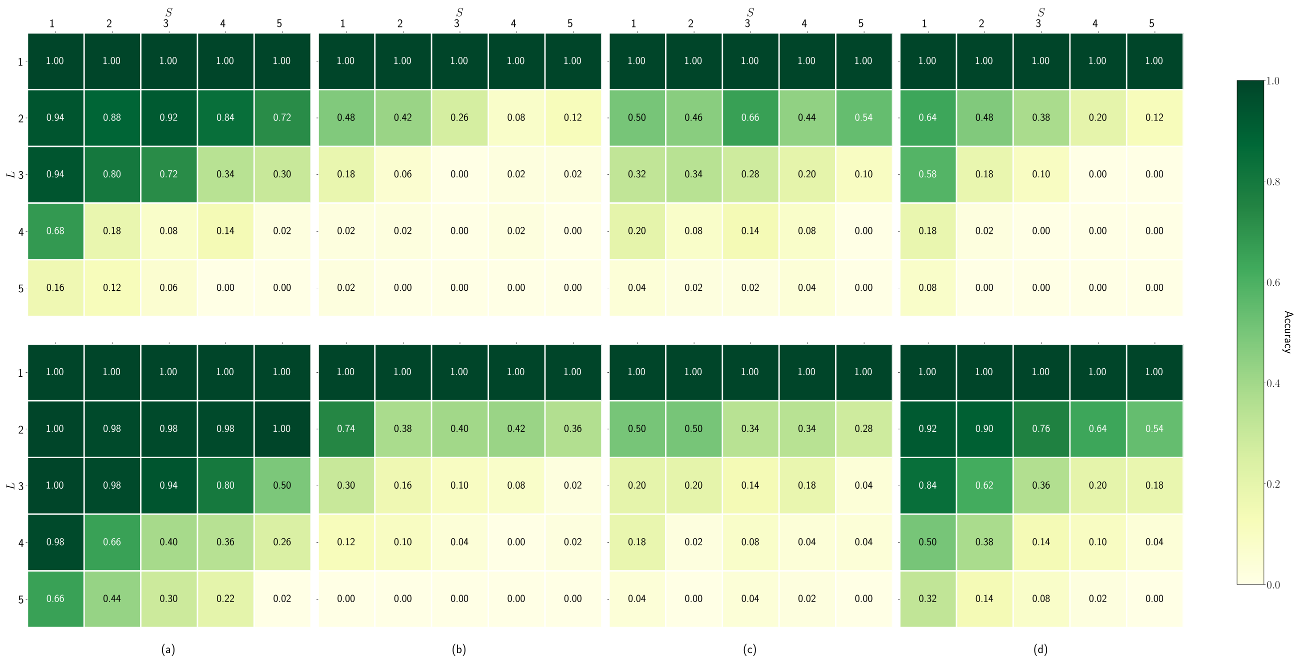

Fig. 2: Average rate of successful recovery as a function of S and L in ErdsRnyi graphs $G(50,0.1)$ (the first three columns) and the social network graph (the fourth column) Zachary's karate club by l_1 (the first row) or $l_* + \lambda l_{2,1}$ (the second row) minimization. (a)-(d) use $\mathbf{A}$, $\mathbf{L}$, $\mathbf{L}^{sym}$, and $\mathbf{A}$ to be the graph-shift operator, respectively.

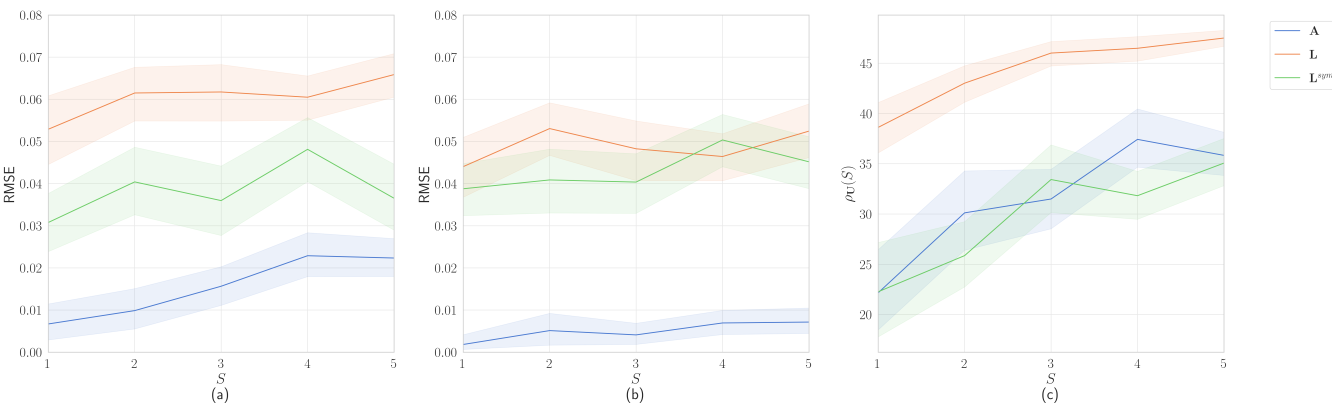

Fig. 3: (a) and (b) show the average RMSE of the l_1 and the $l_* + \lambda l_{2,1}$ minimization algorithms, respectively. (c) demonstrates the values of $\rho_{\mathbf{U}}(S)$ as a function of S. The dash areas denote the 95% confidence intervals around the mean by bootstrapping.

For the random graphs, we carry out experiments for three different choices of $\mathbf{S}$ via the convex-relaxed algorithms: (1) the adjacency matrix A; (2) the unnormalized Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is the diagonal matrix with its i-th diagonal element equals to the degree of node i; (3) the normalized Laplacian matrix $\mathbf{L}^{sym} = \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$ (for any isolated node i, $(\mathbf{D}^{-\frac{1}{2}})_{ii}$ is defined as 0).

For the $l_* + \lambda l_{2,1}$ minimization algorithm, the optimal value of λ is found to be 10 by tuning in the search domain $\{0, 0.2, 0.5, 1, 2, 5, 10\}$. The results are shown in Fig. 2. As we can see, the $l_* + \lambda l_{2,1}$ minimization algorithm performs better than the l_1 minimization algorithm in general, while when $\mathbf{S} = \mathbf{L}^{sym}$, the l_1 minimization algorithm achieves slightly better results.

B. Recovery dependency on Graph Structure

As $\rho_{\mathbf{U}}(S)$ can partially describe the property of a graph, we would like to see how it affects the performance of convex-relaxed algorithms. Next, we fix $L = 3$ and experiment on different choices of sparsity parameter S and graph-shift operator $\mathbf{S}$, which also induces different values of $\rho_{\mathbf{U}}(S)$. The results are shown in Fig. 3. The Laplacian matrix has a somewhat higher value of $\rho_{\mathbf{U}}(S)$, which may be the reason that the l_1 minimization algorithm perform worst when the graph-shift operator is the Laplacian matrix.

C. Recovery Performance of Different Methods

We compare the recovery performance of four kinds of methods on the ErdsRnyi graphs $G(N, 0.1)$ with $L = S =$

3, $\mathbf{S} = \mathbf{A}$ and different values of N: (1) l_1 minimization of Problem (P.7); (2) $l_* + \lambda l_{2,1}$ minimization of Problem (P.6); (3) naive least-square of the bilinear constraint equations via a pseudoinverse; (4) alternative minimization (**AM**). The **AM** algorithm contains two steps. First, given $\mathbf{x}$, we find $\mathbf{h}'$ as a least-square solution of linear equations

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{A}_1\mathbf{x} \\ \mathbf{A}_2\mathbf{x} \\ \vdots \\ \mathbf{A}_L\mathbf{x} \end{bmatrix} \mathbf{h}'$$

where $\mathbf{A}_i \in \mathbb{C}^{N\times N}$ is a matrix formed by concatenating the $[(i-1)\times N+1]$-th to the $(i\times N)$-th columns of $(\mathbf{P}^\top * \mathbf{U}^\top)^\top$. Second, given $\mathbf{h}'$, we find $\mathbf{x}$ which minimizes $\|\hat{\mathbf{y}} - (\mathbf{P}^\top * \mathbf{U}^\top)^\top \operatorname{vec}(\mathbf{x}\mathbf{h}'^\top)\|_2$ and retain the largest S entries in absolute value. We initialize $\mathbf{h}'$ to the first right singular vectors of the simple least-square solution. The **AM** algorithm will stop if the relative error in F-norm of the two consecutive estimators of $\mathbf{Z}$ is smaller than 0.01 or the number of iteration exceeds 1000.

We plot the median RMSE across 50 simulations for each N as a function of the number of nodes in Fig. 4. The simple least-square method and the **AM** algorithm perform worse than the convex-relaxed algorithms. The **AM** algorithm fluctuates greatly and is hard to converge. Besides, when the input signal is sparse enough, i.e., $\frac{S}{N}$ is small enough, the l_1 minimization algorithm is comparable with the $l_* + \lambda_{2,1}$ minimization algorithm.

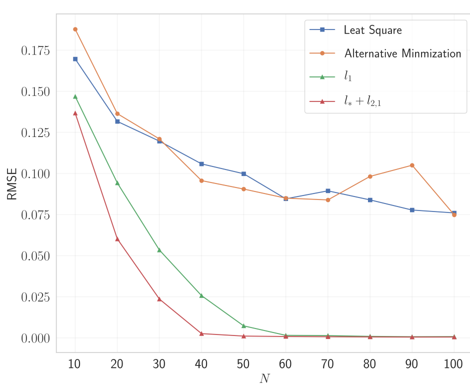

Fig. 4: Median RMSE as a functin of the number of observations N for different methods.

VI. Conclusions

We formulated and studied the deconvolution problem of graphs, which is an extension of signal deconvolution in the time domain or spatial domain. First, we use the lifting technique to overcome the difficulty of bilinear constraints. Thus, we change the vector-valued problem into a matrix-valued problem that aims to recover a row-sparse and low-rank matrix. Also, we prove that in some cases that are very easy to reach, the two issues are equivalent. Second, we utilize different convex-relaxed techniques to get rid of the objective of minimizing the rank. Next, we prove the probabilistic guarantee for the l_1 minimization algorithm to exact recover the input signal and the filter coefficient. Besides, the lower bound of the number of observations required for the theorem to hold is provided. Finally, the numerical experiments show the effectiveness of our convex-relaxed algorithms and the comparable performance of l_1 and $l_* + \lambda l_{2,1}$ minimization algorithms.

References

[1] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

[2] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs," *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[3] A. Sandryhaila and J. M. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, 2014.

[4] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 40–52, 2014.

[5] A. Sandryhaila, S. Kar, and J. M. Moura, "Finite-time distributed consensus through graph filters," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1080–1084, IEEE, 2014.

[6] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters with sparse inputs," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 449–452, IEEE, 2015.

[7] L. Wang and Y. Chi, "Blind deconvolution from multiple sparse inputs," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1384–1388, 2016.

[8] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1146–1159, 2017.

[9] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.

[10] 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.

[11] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," *arXiv preprint arXiv:1506.05163*, 2015.

[12] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in neural information processing systems*, pp. 3844–3852, 2016.

[13] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[14] E. J. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE TRANSACTIONS ON INFORMATION THEORY*, vol. 56, no. 5, p. 2053, 2010.

[15] M. Fazel, H. Hindi, S. P. Boyd, *et al.*, "A rank minimization heuristic with application to minimum order system approximation," in *Proceedings of the American control conference*, vol. 6, pp. 4734–4739, Citeseer, 2001.

[16] C. Steffens, M. Pesavento, and M. E. Pfetsch, "A compact formulation for the l21 mixed-norm minimization problem," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4730–4734, IEEE, 2017.

[17] R. Gribonval, G. Chardon, and L. Daudet, "Blind calibration for compressed sensing by convex optimization," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2713–2716, IEEE, 2012.

[18] Ç. Bilen, G. Puy, R. Gribonval, and L. Daudet, "Convex optimization approaches for blind sensor calibration using sparsity," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4847–4856, 2014.

[19] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," *Bull. Amer. Math. Soc*, vol. 54, pp. 151–165, 2017.

[20] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of computational mathematics*, vol. 12, no. 4, pp. 389–434, 2012.

[21] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[22] I. Chatzigeorgiou, "Bounds on the lambert function and their application to the outage analysis of user cooperation," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1505–1508, 2013.

[23] A. Hoorfar and M. Hassani, "Inequalities on the lambert w function and hyperpower function," *J. Inequal. Pure and Appl. Math*, vol. 9, no. 2, pp. 5–9, 2008.

[24] B. Bollobás and B. Béla, *Random graphs*. No. 73, Cambridge university press, 2001.

[25] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.