



Blind deconvolution on graphs: Exact and stable recovery[☆]

Chang Ye^a, Gonzalo Mateos^{a,b,*}

^a Department of Electrical and Computer Engineering, University of Rochester, Rochester, USA

^b Goergen Institute for Data Science, University of Rochester, Rochester, USA

ARTICLE INFO

Keywords:

Blind deconvolution
Network source localization
Graph signal processing
Convex relaxation
Exact recovery

ABSTRACT

We study a blind deconvolution problem on graphs, which arises in the context of localizing a few sources that diffuse over networks. While the observations are bilinear functions of the unknown graph filter coefficients and sparse input signals, a requirement on invertibility of the diffusion filter enables an efficient convex relaxation leading to a linear programming formulation that can be tackled with off-the-shelf solvers. Under the Bernoulli–Gaussian model for the inputs, we derive sufficient exact recovery conditions in the noise-free setting. A stable recovery result is then established, ensuring the estimation error remains manageable even when the observations are corrupted by a small amount of noise. Numerical tests with synthetic and real-world network data illustrate the merits of the proposed algorithm, its robustness to noise as well as the benefits of leveraging multiple signals to aid the (blind) localization of sources of diffusion. At a fundamental level, the results presented here broaden the scope of classical blind deconvolution of (spatio-) temporal signals to irregular graph domains.

1. Introduction

Network processes such as neural activities at different cortical brain regions [1–3], vehicle flows over transportation networks [4,5], COVID-19 infections across demographic areas connected via a commute flow mobility graph [6], or spatial temperature profiles monitored by distributed sensors [7,8], can be represented as signals supported on the nodes of a graph. In this context, the graph signal processing (GSP) paradigm hinges on recognizing that signal properties are shaped by the underlying graph topology (e.g., in a network diffusion or percolation process), to develop models, signal representations, and information processing algorithms that exploit this relational structure. Accordingly, generalizations of key signal processing tasks have been widely explored in recent work; see [9,10] for recent tutorial accounts. Notably graph filters were conceived as information-processing operators acting on graph-valued signals [11,12], and they are central to graph convolutional neural network models; see e.g. [13]. Mathematically, graph filters are linear transformations that can be expressed as polynomials of the so-termed graph-shift operator (GSO; see Section 2). The GSO offers an algebraic representation of network structure and can be viewed as a local diffusion operator. Its spectral decomposition can be used to represent signals and filters in the graph frequency domain [14]. For the cycle digraph representing e.g., periodic temporal

signals, the GSO boils down to the time-shift operator [9,11,15]. Given a GSO, the polynomial coefficients fully determine the graph filter and are referred to as filter coefficients.

Problem description. In this paper, we revisit the blind deconvolution task for graph signals introduced in [16], with an emphasis on modeling diffusion processes and localization network diffusion sources. Specifically, given P observations of graph signals $\{y_i\}_{i=1}^P$ that we model as outputs of a diffusion filter (i.e., a polynomial in a known GSO), we seek to jointly identify the filter coefficients \mathbf{h} and the input signals $\{x_i\}_{i=1}^P$ that generated the network observations. Since the resulting bilinear inverse problem is ill-posed, we assume that the inputs are sparse – a well-motivated setting when few seeding nodes (the sources) inject a signal that is diffused throughout a network [17]. Localizing sources of network diffusion is a challenging problem with applications in several fields, including sensor-based environmental monitoring, social networks, neural signal processing, or, epidemiology, to name a few. This inverse problem broadens the scope of classical blind deconvolution of temporal or spatial signals to graphs [18–20].

Related work and contributions. A noteworthy approach was put forth in [17], which casts the (bilinear) blind graph-filter identification task as a linear inverse problem in the “lifted” rank-one, row-sparse matrix $\mathbf{x}\mathbf{h}^T$. While the rank and sparsity minimization algorithms

[☆] This work was supported in part by the National Science Foundation, USA under award ECCS-1809356 and by the Center of Excellence in Data Science, an Empire State Development-designated Center of Excellence, USA.

Part of the results in this paper were presented at the 2018 *EUSIPCO* conference (Ye et al., 2018).

* Correspondence to: University of Rochester, 726 Computer Studies Building, Box 270231, Rochester, NY 14627, USA.

E-mail addresses: cye7@ur.rochester.edu (C. Ye), gmateosb@ece.rochester.edu (G. Mateos).

in [17,21] can successfully recover sparse inputs along with low-order graph filters, reliance on matrix lifting can hinder applicability to large graphs. Beyond this computational consideration, the overarching assumption of [17] is that the inputs $\{\mathbf{x}_i\}_{i=1}^P$ share a common support. Here instead we show how a requirement on invertibility of the graph filter facilitates an efficient convex relaxation for the multi-signal case with arbitrary supports (Section 3); see also [20] for a time-domain precursor as well as generalizations of [22] in a supervised learning setting [23] and a study of robustness to graph perturbations [24]. In Section 4, we establish sufficient conditions under which the proposed convex estimator can exactly recover sparse input signals, assumed to adhere to a Bernoulli-Gaussian model. We also derive a stability result for the pragmatic setting where the graph signal observations are corrupted by additive noise. The analysis is challenging, since the favorable (circulant) structure of time-domain filters in [20] is no longer present in the network-centric setting dealt with here. Numerical tests with synthetic and real data corroborate the effectiveness of the proposed approach in recovering the sparse input signals (Section 5). Concluding remarks are given in Section 6. We defer proofs and mathematical details as well as non-essential experimental details to the appendices.

Different from most existing model-based works dealing with source localization on graphs, e.g., [25–27], like [28] we advocate a GSP approach based on an admittedly simple forward (graph filtering) model. In our case, this simplification facilitates a thorough theoretical analysis of recovery performance, offering valuable insights. Often the models of diffusion are probabilistic in nature, and resulting maximum-likelihood source estimators can only be optimal for particular (e.g., tree) graphs [26], or rendered scalable under restrictive dependency assumptions [29]. Relative to [2,28], the proposed framework can accommodate signals defined on general undirected graphs and relies on a *convex* estimator of the sparse sources of diffusion. Furthermore, the setup where multiple output signals are observed (each one corresponding to a different sparse input), has not been thoroughly explored in convex-relaxation approaches to blind deconvolution of (non-graph) signals, e.g., [18,30]; see [20] for a recent and inspiring alternative that we leverage here.

Relative to the conference precursor [22] that discussed inherent scaling and (non-cyclic) permutation ambiguities arising with some particular graphs as well as identifiability issues [31], the exact and stable recovery guarantees in Section 4 are significant contributions of this journal paper. Here we offer a comprehensive presentation along with full-blown technical details, we study robustness to noise, and provide a markedly expanded experimental evaluation anchored in our theoretical analyses.

Notation: Entries of a matrix \mathbf{X} and a (column) vector \mathbf{x} are denoted as X_{ij} and x_i . Calligraphic uppercase letters are used for sets, e.g., \mathcal{S} . Operators $(\cdot)^T$, $\mathbb{E}[\cdot]$, $\text{vec}[\cdot]$, $\sigma_{\max}[\cdot]$, \circ , \otimes and \odot stand for matrix transpose, expectation, matrix vectorization, largest singular value, Hadamard (entry-wise), Kronecker, and Khatri–Rao (column-wise Kronecker) products, respectively. The diagonal matrix $\text{diag}(\mathbf{x})$ has (i, i) th entry x_i . The $n \times n$ identity matrix is represented by \mathbf{I}_n , while $\mathbf{0}_n$ stands for the $n \times 1$ vector of all zeros, and $\mathbf{0}_{n \times p} = \mathbf{0}_n \mathbf{0}_p^T$. A similar convention is adopted for vectors and matrices of all ones. For matrix $\mathbf{M} \in \mathbb{R}^{N \times k}$, we use $\text{span}\{\mathbf{M}\} := \{\mathbf{z} \in \mathbb{R}^N | \mathbf{z} = \mathbf{M}\mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^k\}$ to represent the linear subspace spanned by its columns. Moreover, $[\mathbf{M}]_{\mathcal{S}}$ has entries M_{ij} if $(i, j) \in \mathcal{S}$ and 0 otherwise, where $\mathcal{S} \subseteq \{1, \dots, N\} \times \{1, \dots, k\}$.

The notation $\|\mathbf{X}\|_{p,q} = \left(\sum_j (\sum_i |X_{ij}|^p)^{q/p} \right)^{1/q}$ stands for the elementwise matrix norm, $\|\mathbf{X}\|_F = \|\mathbf{X}\|_{2,2}$ denotes Frobenius norm, and $\|\mathbf{X}\|_{l \rightarrow 2}, l = 1, 2$ are operator norms, i.e., the maximum ℓ_2 norm of a column of \mathbf{X} for $l = 1$ and $\sigma_{\max}[\mathbf{X}]$ for $l = 2$, respectively.

2. Preliminaries and problem statement

We briefly review the necessary GSP background to introduce the observation model and formally state the problem.

2.1. Graph signal processing background

Consider a weighted and undirected graph denoted as $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V} := \{1, \dots, N\}$ comprises the vertex set. The symmetric graph adjacency matrix $\mathbf{A} \in \mathbb{R}_+^{N \times N}$ has entries $A_{ij} = A_{ji} \geq 0$, that represent the weight of the edge (i, j) between nodes i and j . Naturally, if such edge (i, j) is not present in \mathcal{G} then $A_{ij} = 0$. We do not allow for self-loops, hence $A_{ii} = 0$ for all $i \in \mathcal{V}$. Directed graphs are important [32], but beyond the scope of this work; see also Remark 2.

As a general algebraic descriptor of graph connectivity, we henceforth rely on a symmetric GSO $\mathbf{S} \in \mathbb{R}^{N \times N}$, inheriting and encoding the sparsity pattern of \mathcal{G} [11,33]. That is, the only requirement for an admissible GSO \mathbf{S} is that $S_{ij} = 0$ when there is no edge connecting i and j . Typical choices are \mathbf{A} itself, the combinatorial graph Laplacian $\mathbf{L} = \mathbf{A} - \text{diag}(\mathbf{A} \cdot \mathbf{1}_N)$, or, their various degree-normalized forms [9,34] that we will adopt for our experiments in Section 5. Given that \mathbf{S} is real and symmetric, it is always diagonalizable and we can decompose it as $\mathbf{S} = \mathbf{V}\mathbf{A}\mathbf{V}^T$, where $\mathbf{A} = \text{diag}(\lambda_1, \dots, \lambda_N)$ collects the eigenvalues and \mathbf{V} 's columns comprise an orthonormal basis of GSO eigenvectors.

Lastly, we define a graph signal $\mathbf{x} : \mathcal{V} \mapsto \mathbb{R}^N$ as an N -dimensional vector, where x_i denotes the signal value at node $i \in \mathcal{V}$. This value could for instance represent the rating user i assigns to a particular item (e.g., an article as in the experiments in Section 5.3), the measurement collected by sensor i , or, the neural activity recorded in the i th cortical region-of-interest as defined by some brain parcellation. In GSP, the prevalent view is that graph signal \mathbf{x} should not be viewed in isolation, but rather as a tuple along with the GSO \mathbf{S} – the later provides useful contextual (or prior) information about pairwise relationships among individual signal elements in \mathbf{x} . Next, we elaborate on the operator viewpoint of \mathbf{S} as a map between graph signals.

2.2. Graph-filter models of network diffusion

Let \mathbf{y} represent a graph signal supported on \mathcal{G} , which is assumed to be generated from an initial state \mathbf{x} through linear network dynamics of the form

$$\mathbf{y} = a_0 \prod_{l=1}^{\infty} (\mathbf{I}_N - a_l \mathbf{S}) \mathbf{x} = \sum_{l=0}^{\infty} b_l \mathbf{S}^l \mathbf{x}. \quad (1)$$

The linear transformation \mathbf{S} represents one-hop network aggregation (or averaging), so repeated $(l = 1, 2, \dots)$ applications of the GSO in (1) diffuse \mathbf{x} across \mathcal{G} . Accordingly, $\{a_l\}_{l=0}^{\infty}$ and $\{b_l\}_{l=0}^{\infty}$ can be viewed as diffusion coefficients for the multiplicative and additive signal model parametrizations in (1), respectively. This generative model for \mathbf{y} is admittedly simple, but it nonetheless encompasses a broad class of linear network processes such as heat diffusion, average consensus, PageRank, and the DeGroot model of opinion dynamics [35,36].

The Cayley–Hamilton theorem [37, p. 109] ensures that the infinite series in the right-hand-side of (1) can always be equivalently reparametrized using *bounded-degree* polynomials of \mathbf{S} . Introducing the coefficient vector $\mathbf{h} := [h_0, \dots, h_{L-1}]^T$ and the shift-invariant graph filter [9]

$$\mathbf{H} := h_0 \mathbf{I}_N + h_1 \mathbf{S} + h_2 \mathbf{S}^2 + \dots + h_{L-1} \mathbf{S}^{L-1} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l, \quad (2)$$

the signal model (1) becomes

$$\mathbf{y} = \left(\sum_{l=0}^{L-1} h_l \mathbf{S}^l \right) \mathbf{x} = \mathbf{H}\mathbf{x}, \quad (3)$$

for some \mathbf{h} and $L \leq N$. While graph filters are leveraged here as simple generative mechanisms to describe diffusion processes on networks, these convolutional operators play a central role in GSP and machine learning on graphs; see e.g., [12] for a recent tutorial treatment.

Frequency representation. Graph filters and signals admit representations in the frequency (i.e., graph spectral) domain [9,14]. To this end, recall the GSO eigenvalues $\lambda_1, \dots, \lambda_N$ and introduce the $N \times L$

Vandermonde matrix Ψ_L , where $\Psi_{ij} := \lambda_i^{j-1}$ and L is the order of the graph filter (2). The graph Fourier transform (GFT) of a signal \mathbf{x} and the frequency response of filter \mathbf{h} are $\tilde{\mathbf{x}} := \mathbf{V}^T \mathbf{x}$ and $\tilde{\mathbf{h}} := \Psi_L \mathbf{h}$, respectively. This follows by evaluating the GFT of the filter's output $\mathbf{y} = \mathbf{H}\mathbf{x}$ and using the spectral decomposition $\mathbf{S} = \mathbf{V}\mathbf{A}\mathbf{V}^T$, to yield

$$\tilde{\mathbf{y}} = \text{diag}(\Psi_L \mathbf{h}) \mathbf{V}^T \mathbf{x} = \text{diag}(\tilde{\mathbf{h}}) \tilde{\mathbf{x}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}. \quad (4)$$

Apparently, the graph filter \mathbf{H} is diagonalized by the graph's spectral basis \mathbf{V} . As a result $\tilde{\mathbf{y}}$ is given by the element-wise product (\circ) of $\tilde{\mathbf{x}}$ and the filter's frequency response $\tilde{\mathbf{h}} := \Psi_L \mathbf{h}$, analogous to the convolution theorem for temporal signals.

2.3. Problem formulation

Suppose we observe P graph signals that we collect in a matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_P] \in \mathbb{R}^{N \times P}$. For given shift operator \mathbf{S} and filter order $L \leq N$, observations adhere to the (graph convolutional) diffusion model $\mathbf{Y} = \mathbf{H}\mathbf{X}$ in (3), where $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_P] \in \mathbb{R}^{N \times P}$ is sparse having at most $S \ll N$ non-zero entries per column. We do not require that the columns of \mathbf{X} share a common support. The pragmatic setting whereby observations are corrupted by additive noise will be considered in Section 4.2.

The goal is to perform blind deconvolution on the graph \mathcal{G} , which amounts to estimating sparse \mathbf{X} and the filter coefficients \mathbf{h} up to scaling and (possibly) permutation ambiguities [22, Sec. IV-A]. All we are given is \mathbf{Y} , a forward model in terms of the parameterized filter family in (2), and a structural assumption on \mathbf{X} . Sparsity is well motivated when the signals in \mathbf{Y} represent diffused versions of a few localized sources in \mathcal{G} , here indexed by $S := \text{supp}(\mathbf{X}) = \{(i, j) \mid X_{ij} \neq 0\}$. From this vantage point, the blind deconvolution task can be viewed as a source localization one, where recovering \mathbf{X} is of primary interest and \mathbf{h} becomes a nuisance parameter. In any case, the non-sparse formulation is ill-posed, since the number of unknowns $NP + L$ in $\{\mathbf{X}, \mathbf{h}\}$ exceeds the NP observations in \mathbf{Y} .

All in all, using (4) the diffused source localization task can be stated as a non-convex feasibility problem of the form

$$\text{find } \{\mathbf{X}, \mathbf{h}\} \text{ s. to } \mathbf{Y} = \mathbf{V}\text{diag}(\Psi_L \mathbf{h}) \mathbf{V}^T \mathbf{X}, \|\mathbf{X}\|_0 \leq PS, \quad (5)$$

where the ℓ_0 -(pseudo) norm $\|\mathbf{X}\|_0 := |\text{supp}(\mathbf{X})|$ counts the non-zero entries in \mathbf{X} . In other words, the goal is to find the solution to a system of bilinear equations subject to a sparsity constraint in \mathbf{X} ; a hard problem due to the cardinality function as well as the bilinear equality constraints. To deal with the latter, building on [20] we will henceforth assume that the filter \mathbf{H} is invertible.

3. Convex relaxation for invertible graph filters

In this section we propose a convex relaxation of (5), which is feasible under the additional assumption of diffusion filter invertibility. We wrap up with a brief discussion about algorithms to solve the resulting linear program.

3.1. A linear programming problem reformulation

From the filter's input-output relationship in (4), it follows that \mathbf{H} will be invertible if $\tilde{\mathbf{h}}$ does not vanish at any of the graph frequencies $\{\lambda_i\}_{i=1}^N$. That is, no frequency component of the input should be completely annihilated by the filter. We make the following assumption on the signal model.

Assumption 1 (Invertible Graph Filter). Recall the observation model $\mathbf{Y} = \mathbf{H}\mathbf{X}$ in Section 2.2, where $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ is a graph filter. We assume \mathbf{H} is invertible, meaning $\tilde{h}_i = \sum_{l=0}^{L-1} h_l \lambda_i^l \neq 0$ holds for all $i = 1, \dots, N$.

The inverse operator $\mathbf{G} := \mathbf{H}^{-1}$ is also a graph filter on \mathcal{G} , which can be uniquely represented as a polynomial in the shift \mathbf{S} of degree at most $N - 1$ [11, Theorem 4]. Specifically, let $\mathbf{g} \in \mathbb{R}^N$ be the vector of inverse-filter coefficients, i.e., $\mathbf{G} = \sum_{l=0}^{N-1} g_l \mathbf{S}^l$. Then one can equivalently rewrite the observation model $\mathbf{Y} = \mathbf{H}\mathbf{X}$ as

$$\mathbf{X} = \mathbf{G}\mathbf{Y} = \mathbf{V}\text{diag}(\tilde{\mathbf{g}}) \mathbf{V}^T \mathbf{Y}, \quad (6)$$

where $\tilde{\mathbf{g}} := \Psi_N \mathbf{g} \in \mathbb{R}^N$ is the inverse filter's frequency response and $\Psi_N \in \mathbb{R}^{N \times N}$ is a square Vandermonde matrix. Naturally, $\mathbf{G} = \mathbf{H}^{-1}$ implies the condition $\tilde{\mathbf{g}} \circ \tilde{\mathbf{h}} = \mathbf{1}_N$ on the frequency responses, where $\mathbf{1}_N$ denotes the $N \times 1$ vector of all ones.

The fundamental implication of Assumption 1 is that, leveraging (6), one can recast (5) as a linear inverse problem

$$\min_{\{\mathbf{X}, \tilde{\mathbf{g}}\}} \|\mathbf{X}\|_0, \text{ s. to } \mathbf{X} = \mathbf{V}\text{diag}(\tilde{\mathbf{g}}) \mathbf{V}^T \mathbf{Y}, \mathbf{X} \neq \mathbf{0}_{N \times P}. \quad (7)$$

This approach to handle the bilinear Eqs. (5) is markedly different from the matrix lifting technique in [17]. Relative to [17], the additional invertibility requirement we impose offers major algorithmic simplifications and stronger theoretical guarantees; see also the numerical evaluation in Section 5.2. We were inspired by the blind deconvolution method in [20], but in the GSP setting dealt with here the convolution kernel \mathbf{H} is not circulant, and so \mathbf{V} is no longer the discrete Fourier transform (DFT) matrix. This has major technical implications when it comes to establishing exact and stable recovery guarantees; see Section 4.

The ℓ_0 pseudo-norm in (7) renders the problem NP-hard to optimize. Convex-relaxation approaches to tackle sparse recovery problems have enjoyed remarkable success, since they often entail no loss of optimality. An important contribution of this work is to establish this holds for (7) as well; see Section 4. Accordingly, we instead: (i) seek to minimize the ℓ_1 -norm convex surrogate of the cardinality function, that is $\|\mathbf{X}\|_{1,1} = \sum_{i,j} |X_{ij}|$; and (ii) express the filter in the graph spectral domain as in (6). This way, we arrive at the convex cost function

$$\|\mathbf{X}\|_{1,1} = \|\mathbf{G}\mathbf{Y}\|_{1,1} = \|\mathbf{V}\text{diag}(\tilde{\mathbf{g}}) \mathbf{V}^T \mathbf{Y}\|_{1,1} = \|(\mathbf{Y}^T \mathbf{V} \circ \mathbf{V}) \tilde{\mathbf{g}}\|_1,$$

where the last equality is obtained after vectorizing the norm's matrix argument, and \circ denotes the Khatri-Rao (i.e., columnwise Kronecker) product. Our idea is to solve the convex ℓ_1 -synthesis problem (in this case a linear program), e.g., [38], namely

$$\hat{\tilde{\mathbf{g}}} = \underset{\tilde{\mathbf{g}} \in \mathbb{R}^N}{\text{argmin}} \|(\mathbf{Y}^T \mathbf{V} \circ \mathbf{V}) \tilde{\mathbf{g}}\|_1, \text{ s. to } \mathbf{1}_N^T \tilde{\mathbf{g}} = N. \quad (8)$$

While the linear constraint in (8) avoids $\hat{\tilde{\mathbf{g}}} = \mathbf{0}_N$, it also serves to fix the scale of the solution since the bilinear problem (5) is only identifiable up to scaling. Without loss of generality we henceforth let $\mathbf{1}_N^T \hat{\tilde{\mathbf{g}}} = N$. An insightful discussion on the role of the constraint will emerge as an upshot of the exact recovery guarantees derived in Section 4.1. A reason for choosing the (arbitrary) value N in the constraint is that it simplifies said recovery conditions; see also (12).

Remark 1 (Badly-Conditioned Graph Filters). For badly conditioned filters \mathbf{H} that are close to being singular, some entries of the frequency response $\tilde{h}_i \approx 0$ and the corresponding \tilde{g}_i can be quite large. Our theoretical analysis in Section 4 will reveal that, as expected, in such cases exact recovery via the convex relaxation (8) can be challenging. Adoption of regularization techniques can help mitigate these adverse effects, say by augmenting the cost function in (8) with a (weighted) ridge or Tikhonov penalty $\|\tilde{\mathbf{g}}\|_2^2$ to encourage shrinkage. We leave these extensions to our model as part of our future work.

3.2. Algorithms

From an algorithmic point of view, under the assumption that the diffusion filter is invertible (cf. Assumption 1), one can readily use e.g., an off-the-shelf interior-point method or a specialized sparsity-minimization algorithm to solve the linear programming formulation

Algorithm 1 Iteratively-reweighted ℓ_1 minimization for (8).

-
- 1: **Input:** Matrix $\mathbf{Y}^\top \mathbf{V} \odot \mathbf{V}$, $\mu > 0$ and $\epsilon > 0$.
 - 2: **Initialize** $t = 0$, $\mathbf{w}^{(0)} = \mathbf{1}_{NP}$, $\mathbf{X}^{(0)} = \mathbf{0}_{N \times P}$.
 - 3: **repeat**
 - 4: Solve

$$\hat{\mathbf{g}}^{(t+1)} = \underset{\mathbf{g}}{\operatorname{argmin}} \left\| \mathbf{w}^{(t)} \circ [(\mathbf{Y}^\top \mathbf{V} \odot \mathbf{V}) \hat{\mathbf{g}}] \right\|_1, \quad \text{s. to } \mathbf{1}_N^\top \hat{\mathbf{g}} = N.$$
 - 5: Form $\operatorname{vec}(\mathbf{X}^{(t+1)}) = (\mathbf{Y}^\top \mathbf{V} \odot \mathbf{V}) \hat{\mathbf{g}}^{(t+1)}$.
 - 6: Update $w_i^{(t+1)} = \frac{1}{[\operatorname{vec}(\mathbf{X}^{(t+1)})]_i + \mu}$, $i = 1, 2, \dots, NP$.
 - 7: $t \leftarrow t + 1$.
 - 8: **until** $\|\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}\|_2 \leq \epsilon$.
 - 9: **return** $\hat{\mathbf{g}} := \hat{\mathbf{g}}^{(t)}$ and $\hat{\mathbf{X}} := \mathbf{X}^{(t)}$.
-

(8) efficiently. Different from the solvers in [17,39], the aforementioned algorithmic alternatives are free of expensive singular-value decompositions per iteration. Regardless of the particular algorithm chosen, we have found that the overall performance can be improved via the iteratively-reweighted ℓ_1 -norm minimization procedure tabulated under Algorithm 1; see also [40] for a justification of such refinement that effectively zeros out small residual entries in intermediate estimates of \mathbf{X} . In practice, a couple refinement iterations suffice so the additional overhead is minimal.

Once the frequency response $\hat{\mathbf{g}}$ of the inverse filter is recovered, one can readily reconstruct the sparse sources via

$$\hat{\mathbf{X}} = \operatorname{unvec}((\mathbf{Y}^\top \mathbf{V} \odot \mathbf{V}) \hat{\mathbf{g}}), \quad (9)$$

where here the $\operatorname{unvec}(\cdot)$ operator reshapes its vector argument to an $N \times P$ matrix. If so desired, one can likewise form the diffusion filter estimate $\hat{\mathbf{H}} = \mathbf{V} \operatorname{diag}(\mathbf{1}_N / \hat{\mathbf{g}}) \mathbf{V}^\top$, where the division is to be conducted entrywise.

In the next section, we will discuss the exact recovery condition for (8) and its robustness to noise-corrupted observations. Before moving on, a remark is in order.

Remark 2 (Extensions to Directed Graphs). In this work we consider undirected graphs \mathcal{G} with symmetric GSO \mathbf{S} . Yet most of the methods developed here are still applicable in the directed case, provided that the asymmetric GSO is diagonalizable as $\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1}$. Eigenvectors and eigenvalues can be complex valued, so the decision variable $\hat{\mathbf{g}} \in \mathbb{C}^N$ will also be complex, challenging optimization using Algorithm 1. A simple workaround is to go to the vertex domain and search for the inverse filter coefficients $\mathbf{g} \in \mathbb{R}^N$, after substituting $\hat{\mathbf{g}} = \boldsymbol{\Psi}_N \mathbf{g}$ in (8). Beyond algorithms, our recovery guarantees in the ensuing section will likely carry over after minor modifications resulting from the aforementioned change of variable $\hat{\mathbf{g}} = \boldsymbol{\Psi}_N \mathbf{g}$. We defer these and other intriguing extensions (see Section 6) to our future work.

4. Recovery guarantees

Here we conduct a theoretical analysis of the proposed convex estimator in (8), which is a relaxation of the blind deconvolution problem on the graph \mathcal{G} . We first derive exact recovery guarantees, which hold with high probability under a Bernoulli-Gaussian model for the sparse inputs. A stable recovery result is then established, ensuring the estimation error on the inverse filter's frequency response can be kept in check when the observations are corrupted by a small amount of additive noise.

Because of its analytical tractability, the Bernoulli-Gaussian model is widely adopted to describe and generate random sparse matrices such as the unknown sources $\mathbf{X} \in \mathbb{R}^{N \times P}$ in Section 2.3. We henceforth adopt the following model specification that is consistent with the definition in [31].

Definition 1 (Bernoulli-Gaussian Model). We say a random matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ adheres to a Bernoulli-Gaussian model with parameter $\theta \in (0, 1)$, if its entries are $X_{ip} = \Omega_{ip} \gamma_{ip} / \sqrt{\theta}$, where $\Omega_{ip} \sim \operatorname{Bernoulli}(\theta)$ and $\gamma_{ip} \sim \operatorname{Normal}(0, 1)$ are i.i.d. for all i, p .

In the context of Definition 1 we say that the matrix entries X_{ip} are Bernoulli-Gaussian random variables, with $\mathbb{E}[X_{ip}] = 0$ and $\operatorname{var}[X_{ip}] = 1$. Apparently, the model parameter θ offers a handle on the sparsity level of \mathbf{X} , while entries in $S = \operatorname{supp}(\mathbf{X}) \equiv \operatorname{supp}(\boldsymbol{\Omega})$ are drawn from a standard Normal distribution.

4.1. Exact recovery conditions

Suppose that (7) is identifiable (see e.g., [22, Remark 1] for sufficient conditions under the Bernoulli-Gaussian model), and let $\{\mathbf{X}_0, \tilde{\mathbf{g}}_0\}$ be the solution. For the ensuing discussion and to state our exact recovery result, some preliminary notation is in order. We compactly denote polynomials of the given graph-shift operator $\mathbf{S} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\top$ as $\mathcal{P}(\mathbf{h}) := \mathbf{V} \operatorname{diag}(\mathbf{h}) \mathbf{V}^\top$, where \mathbf{h} is the corresponding filter's frequency response. So given observations $\mathbf{Y} = \mathcal{P}(\tilde{\mathbf{h}}_0) \mathbf{X}_0 \in \mathbb{R}^{N \times P}$ (implying $\mathbf{X}_0 = \mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{Y}$ under Assumption 1, for $\tilde{\mathbf{g}}_0 \circ \tilde{\mathbf{h}}_0 = \mathbf{1}_N$), we will study the following problem [cf. (8)]

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathcal{P}(\mathbf{g}) \mathbf{Y}\|_{1,1}, \quad \text{s. to } \mathbf{r}^\top \hat{\mathbf{g}} = c, \quad (10)$$

where $\mathbf{r} \in \mathbb{R}^N$ is a generic constraint vector, and $c = \mathbf{r}^\top \tilde{\mathbf{g}}_0 \neq 0$ is a constant to control the (inherently ambiguous) scale of the solution $\hat{\mathbf{g}}$. In general we do not have any prior knowledge about the ground-truth filter, so we can just select $\mathbf{r} = \mathbf{1}_N$ as in (8). Our main result will offer some insights on how the recovery performance is affected by the choice of \mathbf{r} .

We let $\mathbf{P}_1^\perp := \mathbf{I}_N - \frac{\mathbf{1}_N \mathbf{1}_N^\top}{N}$ denote the projection operator onto the orthogonal complement of the subspace $\operatorname{span}(\mathbf{1}_N)$. Finally, consider the matrix $\tilde{\mathbf{U}} := (\mathbf{V} \odot \mathbf{V}) \mathbf{P}_1^\perp \in \mathbb{R}^{N \times N}$, which one can show has a maximum singular value $\sigma_{\max}(\tilde{\mathbf{U}}) \leq 1$. With these definitions in place, we have all the elements to state the main theorem in this section. The result provides sufficient conditions to guarantee exact recovery of the inverse filter $\tilde{\mathbf{g}}_0$, and hence the sparse sources \mathbf{X}_0 via (9), with high probability.

Theorem 1 (Exact Recovery). Consider graph signal observations $\mathbf{Y} = \mathcal{P}(\tilde{\mathbf{h}}_0) \mathbf{X}_0 \in \mathbb{R}^{N \times P}$, where \mathbf{X}_0 adheres to the Bernoulli-Gaussian model with $\theta \in (0, 0.324]$. Recall that under Assumption 1, we can write $\mathbf{X}_0 = \mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{Y}$. Let $P \geq C' \sigma_m^{-2} \log \frac{4}{\delta}$, where $\sigma_m = \min(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ and $\sigma_1 \in \left(0, \frac{\sqrt{\pi} \theta^{3/2}}{2}\right]$, $\sigma_2 \in \left(0, \frac{\sqrt{\pi} \theta}{2}\right]$, $\sigma_3 > 0$, $\sigma_4 \in (0, 1)$, $\delta \in (0, 1)$ are parameters, while C' is a constant that does not depend on P , σ_m , or δ . Then $\hat{\mathbf{g}} = \tilde{\mathbf{g}}_0$ is the unique solution to (10) with probability at least $1 - \delta$, if

$$\left\| \mathbf{P}_1^\perp \operatorname{diag}(\mathbf{r}) \tilde{\mathbf{g}}_0 \right\|_2 \leq (c/N) d_0, \quad (11)$$

$$\text{where } d_0 := \frac{\sqrt{1 - \sigma_{\max}^2(\tilde{\mathbf{U}}) [(1 - \sigma_1) - 2\theta(1 + \sigma_2)] (1 - \sigma_4)}}{(1 + \sigma_3) \sqrt{\theta}}.$$

Proof. See Appendix A. \square

Notice that when $\mathbf{r} = \mathbf{1}_N$, we have $c = \mathbf{1}_N^\top \tilde{\mathbf{g}}_0 = N$ as per our arbitrary choice of scale in (8). Thus, the sufficient recovery condition (11) simplifies to

$$\left\| \mathbf{P}_1^\perp \tilde{\mathbf{g}}_0 \right\|_2 \leq d_0. \quad (12)$$

Condition (12) essentially states that when the frequency response of the inverse filter $\tilde{\mathbf{g}}_0$ is closer to the all-ones vector $\mathbf{1}_N$, meaning that $\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0$ is smaller in magnitude, the inverse filter is easier to recover, as the right-hand-side of the bound (12) does not directly depend on $\tilde{\mathbf{g}}_0$ (except through the ratio of its arbitrarily fixed scale and dimensionality, and here $c = N$). In any case, it is pleasing to see the conditions (11) and (12) only depend on the scale-normalized

filter frequency responses, e.g., $\tilde{\mathbf{g}}_0/c$ in (11). Going back to Remark 1, Eq. (12) and the prior discussion corroborate that a badly-conditioned filter \mathbf{H} will have an inverse \mathbf{G} that is more challenging to recover. Indeed, in such case some of the entries of $\tilde{\mathbf{g}}_0$ will be quite large — hence markedly deviating from $\mathbf{1}_N$. All in all, one can interpret $\|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0\|_2$ as a measure of the problem's ill conditioning.

Furthermore, the influence of prior knowledge of the true filter $\tilde{\mathbf{h}}_0$ on the recovery performance can be illustrated through a specific and straightforward case where $\mathbf{r} = \tilde{\mathbf{h}}_0$ in (11). In this idealized scenario, since $\|\mathbf{P}_1^\perp \text{diag}(\tilde{\mathbf{h}}_0) \tilde{\mathbf{g}}_0\|_2 = \|\mathbf{P}_1^\perp \mathbf{1}_N\|_2 = 0$, the left-hand-side of (11) will certainly be less than $(c/N)d_0$. Although this is a trivial case, it suggests that when \mathbf{r} is closer to $\tilde{\mathbf{h}}_0$ (meaning that $\text{diag}(\mathbf{r})\tilde{\mathbf{g}}_0 = \mathbf{r} \circ \tilde{\mathbf{g}}_0$ is closer to $\mathbf{1}_N$), the recovery performance of the proposed convex relaxation (10) will improve.

Additionally, increasing θ within the feasible range $\theta \in (0, 0.324]$ will decrease d_0 , making the recovery of the inverse filter more challenging. Conversely, when $\|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0\|_2$ is smaller, a lower d_0 can be tolerated, allowing for a denser input signal (higher θ). Moreover, in this last case the feasible parameters $\{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$ (and hence σ_m) can be larger, resulting in a lower required number of observations P . Naturally, increasing the recovery probability $1 - \delta$ (provided that the sufficient condition is satisfied) necessitates a larger sample size; see the $\log(4/\delta)$ scaling.

4.2. Stable recovery from noisy data

Suppose we are now given P noise-corrupted graph signal observations $\mathbf{Y} = \mathcal{P}(\tilde{\mathbf{h}}_0)\mathbf{X}_0 + \mathbf{N}$, where $\mathbf{N} \in \mathbb{R}^{N \times P}$ is an additive noise matrix. Since the filter is invertible, we can write $\mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{Y} = \mathbf{X}_0 + \mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{N}$. It will be convenient to split the effective noise $\mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{N}$ into matrix components respectively corrupting entries in $S := \text{supp}(\mathbf{X}_0)$ and its complement S^C . The latter will be denoted by $\mathbf{N}^{(C)} := [\mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{N}]_{S^C} = \mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{N} - [\mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{N}]_S$, where matrix $[\mathbf{M}]_{\mathcal{A}}$ has entries M_{ij} if $(i, j) \in \mathcal{A}$ and 0 otherwise.

We can establish the following error bound for the solution of (10), which holds under the same conditions of Theorem 1 and asserts that the inverse filter recovery error will be stable.

Theorem 2 (Stable Recovery). Consider graph signal observations $\mathbf{Y} = \mathcal{P}(\tilde{\mathbf{h}}_0)\mathbf{X}_0 + \mathbf{N}$, where $\mathbf{N} \in \mathbb{R}^{N \times P}$ is an additive noise matrix. Assume the conditions in Theorem 1 are satisfied. Then the estimation error associated to the solution $\hat{\mathbf{g}}$ of problem (10) is bounded by

$$\|\hat{\mathbf{g}} - \tilde{\mathbf{g}}_0\|_l \leq \frac{2 \left\| \text{diag}(\tilde{\mathbf{g}}_0) \left(\mathbf{I}_N - \mathbf{1}_N \frac{(\mathbf{r} \circ \tilde{\mathbf{g}}_0)^T}{c} \right) \right\|_{l \rightarrow 2} \|\mathbf{N}^{(C)}\|_{1,1}}{\sqrt{\frac{2}{\pi}} PQ - d_0 \|\mathbf{N}^{(C)}\|_{1,1} - \|\mathbf{N}^{(C)}\|^T \mathbf{V} \circ \mathbf{V}\|_{1 \rightarrow 2}}, \quad (13)$$

where

$$Q := \frac{(1 + \sigma_3)\sqrt{\theta}N}{c} \left[\sqrt{\left(\frac{cd_0}{N} \right)^2 - (1 - \sigma_3)^2 \|\mathbf{P}_1^\perp \text{diag}(\mathbf{r})\tilde{\mathbf{g}}_0\|_2^2} - \sigma_3 \|\mathbf{P}_1^\perp \text{diag}(\mathbf{r})\tilde{\mathbf{g}}_0\|_2 \right],$$

for some $\sigma_5 \in [0, 1]$, and $\|\cdot\|_l$ stands for the ℓ_1 and ℓ_2 norms when $l = 1, 2$, respectively.

Proof. See Appendix B. \square

Notice first that $Q \geq 0$ when (11) holds. Importantly, the denominator in the right-hand-side of (13) should be non-negative to obtain feasible upper bound. This effectively imposes a constraint on the magnitude of the noise component $\mathbf{N}^{(C)}$ in S^C , which should satisfy

$$\|\mathbf{N}^{(C)}\|_F \leq \frac{\sqrt{\frac{2}{\pi}} PQ}{d_0 \|\tilde{\mathbf{N}}^{(C)}\|_{1,1} + \|[\tilde{\mathbf{N}}^{(C)}]^T \mathbf{V} \circ \mathbf{V}\|_{1 \rightarrow 2}}, \quad (14)$$

where $\tilde{\mathbf{N}}^{(C)} := \mathbf{N}^{(C)}/\|\mathbf{N}^{(C)}\|_F$. The right-hand-side of (14) provides an upper bound to the strength of the noise that is tolerable. For further intuition, suppose $\mathbf{r} = \mathbf{1}_N$ as in (8). Once more, in favorable settings

where $\|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0\|_2$ is small, e.g., if $\tilde{\mathbf{g}}_0$ is closer to the all-ones vector $\mathbf{1}_N$, we will have a larger upper bound in (14) because Q will be larger. Similarly Q will increase when d_0 is large, for instance in sparse settings where θ is small. Either way, we have $Q \approx (1 + \sigma_3)\sqrt{\theta}d_0$ and the noise condition simplifies to $\|\mathbf{N}^{(C)}\|_{1,1} \leq \sqrt{\frac{2\theta}{\pi}}(1 + \sigma_3)P$.

5. Numerical results

We carry out numerical experiments to assess the performance of the proposed convex relaxation in a variety of settings. To this end, we run Algorithm 1 and solve the per-iteration sparse recovery problems using CVX [41]. We first rely on synthetic data to simulate various controlled settings, which allow us to verify some of the conclusions drawn from our theoretical analysis in Section 4. We then compare against the matrix lifting approach in [17] with a focus on the sparsity levels and filter orders that lead to successful recovery. Finally, we conduct a network source localization experiment using real social network data, and include the non-convex ℓ_1 recovery algorithm in [28] as an additional baseline.

5.1. Key parameters affecting recovery performance

Here we consider Erdős-Renyi random graphs with $N = 20$ and edge connection probability $p = 0.4$. For the GSO we adopt the degree-normalized adjacency matrix $\mathbf{S} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, where $\mathbf{D} := \text{diag}(\mathbf{A} \cdot \mathbf{1}_N)$. Notice that the eigenvalues of \mathbf{S} are bounded, i.e., $\lambda_i \in (-1, 1]$. We generate graph filters \mathbf{H} such that their inverse filters have frequency responses $\tilde{\mathbf{g}} = \mathbf{1}_N + \alpha \mathbf{P}_1^\perp \tilde{\mathbf{g}}_\perp$, with perturbation $\tilde{\mathbf{g}}_\perp \in \mathbb{R}^N$ drawn from $\text{Normal}(\mathbf{0}_N, \mathbf{I}_N)$ (multivariate standard Gaussian) and subsequently scaled such that $\|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_\perp\|_2 = 1$. Hence, the left-hand-side of (12) becomes $\|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0\|_2 = \alpha \|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_\perp\|_2 = \alpha$ and (12) will be satisfied whenever $\alpha \leq d_0$. The matrix \mathbf{X} of sparse sources is drawn according to the Bernoulli-Gaussian model in Definition 1, and we will control the sparsity parameter θ . Besides, in order to verify the stability properties conveyed by the error bound (13), we also consider noise corrupted observations $\mathbf{Y} = \mathbf{H}\mathbf{X} + \eta\mathbf{N}$, where \mathbf{N} has i.i.d. standard Gaussian entries.

Given this simulation setup, we conducted four experiments to further examine the exact and stable recovery properties of (8). The results are shown in Fig. 1, which depicts different figures of merit as a function of: (i) α vs P , (ii) α vs θ , (iii) θ vs P , (iv) η vs α . As figures of merit we consider the root mean square error (RMSE) of the estimated sources $\hat{\mathbf{X}}$ as well as the support recovery accuracy, i.e.,

$$\text{RE}_x := \frac{\|\hat{\mathbf{X}} - \mathbf{X}_0\|_F}{\|\mathbf{X}_0\|_F}, \quad \text{ACC}_x := \frac{|\text{supp}_\kappa(\hat{\mathbf{X}}) \cap \text{supp}_\kappa(\mathbf{X}_0)|}{|\text{supp}_\kappa(\mathbf{X}_0)|},$$

where $\text{supp}_\kappa(\cdot)$ is the support function with threshold κ . For these tests, we use $\kappa = 0.1$.

In Fig. 1, the first and third columns depict $1 - \text{RE}_x$ and the second and fourth columns show ACC_x . In all cases, lighter-colored pixels indicate better performance. In the top row, we examine the relation of $\alpha = \|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0\|_2$ v.s. P in (a)–(b) and α v.s. θ in (c)–(d). Our previous discussion for Theorem 1 and the recovery condition (12) is corroborated by these results which show a smaller α would both reduce the required sample size P and tolerate denser sources, i.e., a larger θ . From Figs. 1(e)–(f), it can be seen that while $\alpha = 2$ is fixed, which implies the lower bound of d_0 is unchanged, source signals generated with a larger θ require a larger sample size P for successful recovery. And this observation is also consistent with Theorem 1. Notice that while our sufficient conditions hold for $\theta \in (0, 0.324]$, in practice successful recovery is possible for larger θ provided that e.g., P is large enough or α is sufficiently small.

Moving on to the noisy setting, from Figs. 1(g)–(h) it can be seen that a smaller α would result in higher robustness to noise. This finding is also consistent with Theorem 2, because $\|\mathbf{P}_1^\perp \tilde{\mathbf{g}}_0\|_2 = \alpha$ when we use the all-ones vector in the constraint, i.e., $\mathbf{r} = \mathbf{1}_N$. All in all, as expected the numerical results in Fig. 1 are in line with the theoretical guarantees in Theorems 1 and 2.

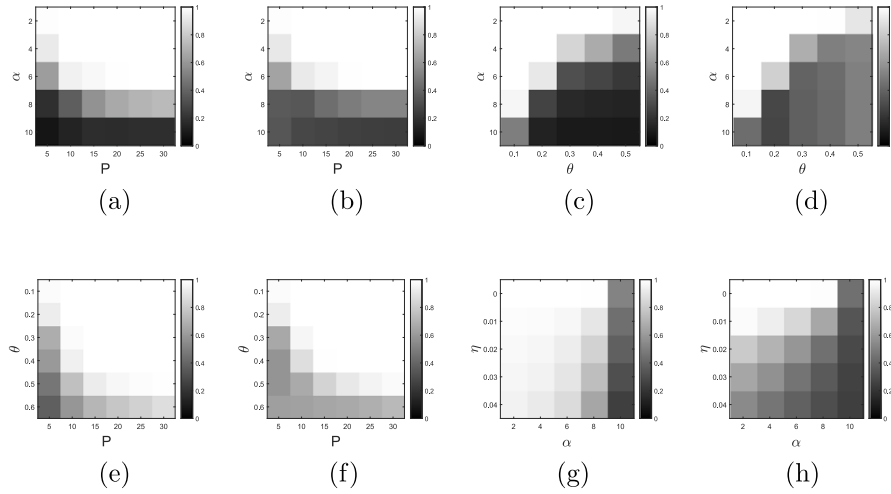


Fig. 1. Recovery performance of Algorithm 1 for problems involving Erdős-Renyi random graphs with $N = 20$, $p = 0.4$. The first and third columns show $1 - \text{RE}_x$ for different settings (the whiter the better, meaning lower error), the second and fourth columns show the accuracy of support estimate for different settings (the whiter the better, meaning higher fraction of true sources recovered). For (a) and (b) the recovery performance of α v.s. P , with $\theta = 0.15$, is shown. For (c) and (d) α v.s. θ with $P = N = 20$. For (e) and (f) θ v.s. P , with $\alpha = 0.1N = 2$. For (g) and (h) we consider the noisy case, i.e. η v.s. α , with $\theta = 0.1$, $P = N = 20$. All of the results represent averages over 100 independent realizations. Apparently, the proposed approach exhibits satisfactory performance over a wide range of settings.

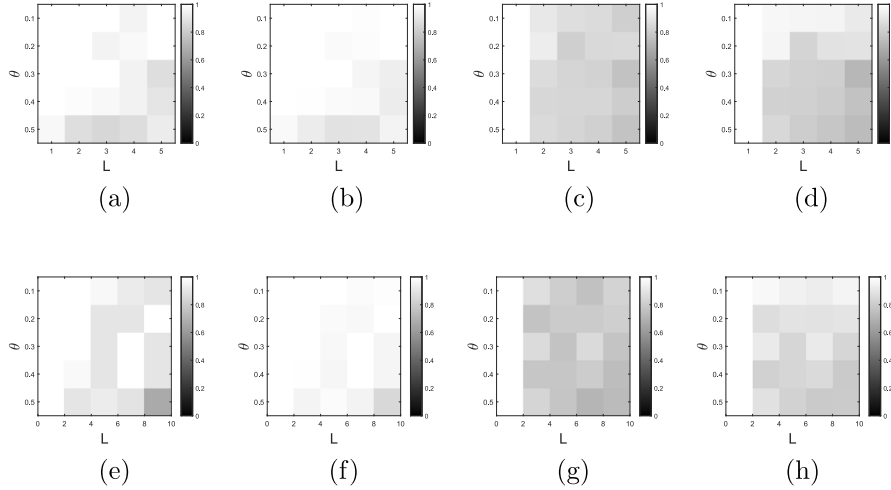


Fig. 2. Recovery performance comparing Algorithm 1 with the matrix-lifting approach in [17], for an ER graph with $N = P = 20$, $p = 0.4$ and a filter perturbation magnitude of $\alpha' = 0.5$. Plots (a) and (b) show $1 - \text{RE}_x$ and accuracy of support estimate for the proposed estimator, while (c) and (d) show $1 - \text{RE}_x$ and accuracy of support estimate for the lifting approach. The same is shown in plots (e)–(h), but for the 66-node structural brain network studied in [42], with $P = 30$. All the results represent averages over 20 independent realizations. Algorithm 1 uniformly outperforms the matrix-lifting baseline, and is less sensitive to the filter order L .

5.2. Comparing with the matrix lifting approach in [17]

Here we start by examining the sensitivity of the estimator (8) to the graph filter order L , which may impact $\|\mathbf{P}_1^+ \tilde{\mathbf{g}}_0\|_2$ and hence the satisfiability of (11). To this end, let $\mathbf{S} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ and consider graph filter coefficients of the form $\mathbf{h} = [1, h_1, \dots, h_{L-1}]^T$. The frequency response $\tilde{\mathbf{h}} = \Psi_L \mathbf{h} = \mathbf{1}_N + (\sum_{l=1}^{L-1} h_l) \mathbf{e}_1 + [\Psi_L]_{-1} \mathbf{h}_{-1}$, where $[\cdot]_{-1}$ zeros out the first row of its matrix argument, i.e., $[\Psi_L]_{-1} = \text{diag}(\mathbf{1}_N - \mathbf{e}_1) \Psi_L$ and likewise $\mathbf{h}_{-1} = \text{diag}(\mathbf{1}_N - \mathbf{e}_1) \mathbf{h}$, where $\mathbf{e}_1 = [1, 0, \dots, 0]^T \in \mathbb{R}^N$. As for the normalized adjacency matrix used as GSO, all its eigenvalues are within $(-1, 1)$ except for the first (biggest) eigenvalue $\lambda_1 = 1$, so the term $[\Psi_L]_{-1} \mathbf{h}_{-1}$ can be viewed as a small perturbation relative to $\mathbf{1}_N$. This suggests that a higher L together with filter coefficients \mathbf{h}_{-1} of larger magnitude would lead to greater deviations of $\tilde{\mathbf{h}}$ from $\mathbf{1}_N$. This in turn implies a bigger $\|\mathbf{P}_1^+ \tilde{\mathbf{g}}_0\|_2$ that would make filter recovery harder. The preceding discussion suggests evaluating the recovery performance of Algorithm 1 for graph filters with coefficients of the form $\mathbf{h} = \mathbf{e}_1 + \alpha' \mathbf{h}_{-1}$, with $\mathbf{h}_{-1} = [0, \mathbf{h}_{L-1}^T]^T$ and $\mathbf{h}_{L-1} \in \mathbb{R}^{L-1}$ is drawn from $\text{Normal}(\mathbf{0}_{L-1}, \mathbf{I}_{L-1})$. Parameter α' controls the filter perturbation,

similar to α in Section 5.1. However, working with (and perturbing) \mathbf{h} instead of the inverse filter $\tilde{\mathbf{g}}$ as in Section 5.1, directly allows us to examine the effect of L .

In this context, we compare the recovery performance of the proposed approach (8) against the matrix-lifting baseline in [17] – the latter is known to be more sensitive to L , especially for higher-order filters [22]. We evaluate and report the same figures of merit RE_x and ACC_x used in the previous experiment, but now as a function of sparsity (θ) and filter order (L). We fix the filter perturbation level to $\alpha' = 0.5$. The numerical results are shown in Figs. 2(a)–(d) for a $N = 20$ -node Erdős-Renyi random graph with $p = 0.4$, and in Figs. 2(e)–(h) for a $N = 66$ -node structural brain connectome from the study in [42]. The first two columns in Fig. 2 depict $1 - \text{RE}_x$ and ACC_x for Algorithm 1, respectively; while the last two columns show the counterparts for the estimator in [17]. Again, lighter-colored pixels are indicative of better performance. Results in Fig. 2 clearly show that Algorithm 1 uniformly outperforms the matrix-lifting baseline, and is markedly less (adversely) affected by large values of L – especially when the input sources are sufficiently sparse. The results are fairly consistent across the graphs

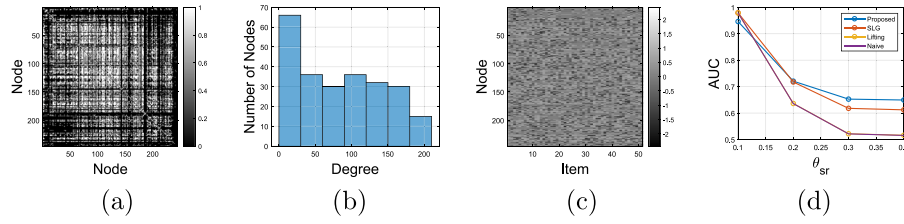


Fig. 3. Recovery performance on the Epinions dataset. (a) The adjacency matrix W of the sampled directed social network and the (b) the degree distribution of the symmetrized undirected graph with adjacency $A = (W + W^T)/2$. (c) The centered rating data Y_{obs} . (d) Source localization performance is quantified in terms of the AUC as a function source density level $\theta_{\text{sr}} = \{0.1, 0.2, 0.3, 0.4\}$. We compare four different approaches: (i) the proposed estimator (8); (ii) the source localization on graphs (SLG) algorithm in [28], (iii) the matrix lifting-based approach from [17], and (iv) the naive predictor whereby $\hat{X} = Y_{\text{obs}}$. We find that Algorithm 1 offers more robust predictions across a broader range of input sparsity levels.

tested here (compare the top and bottom rows in Fig. 2), but we note recovery appears to be more challenging when signals are diffused over the real brain connectome (even with 50% more observations). In part, this could be explained by the fact that $\sigma_{\text{max}}(\tilde{U}_{\text{ER}}) = 0.5054$ (averaged over 20 realizations of ER graphs with $N = 20$ and $p = 0.4$), while $\sigma_{\text{max}}(\tilde{U}_{\text{brain}}) = 0.8769$. Hence, d_0 in (12) will be smaller for the brain network — everything else being equal.

5.3. Experiments on real social network data

In this section we test the proposed approach on the Epinions dataset [43], a who-trusts-whom online social network that includes 132k users, 755k items (articles written by some of the users), 13M user-to-item ratings (1–5 scale, with timestamp), and the signed trust/distrust pairwise relations (717k for trust and 123k for distrust) between users. Leveraging these data, we want to tackle the following network source localization problem: given a connected social graph and timestamped item ratings generated by users in this social network, can we locate the subset of users who generate the earliest ratings, assuming that other users' rating might be impacted by those early ratings?

More precisely, for a connected social network with N user nodes and a set of P items, we want to identify the θ_{sr} earliest ratings (θ_{sr} is a prescribed proportion of earliest ratings assigned to these items, say $\theta_{\text{sr}} = 10\%, 20\%, 30\%, 40\%$) from the observed ratings matrix $Y_{\text{obs}} \in \mathbb{R}^{N \times P}$. Note that the rating density of the whole dataset is fairly low, i.e., 0.0015% [43]. As a result, for a sparse observation matrix Y_{obs} the optimal solution of (8) would likely yield the trivial result $\hat{g} = \mathbf{1}_N$, since this solution both satisfies the recovery condition (11) and is compatible with the sparsity requirement on the sources $\hat{X} = V \text{diag}(\hat{g}) V^T Y_{\text{obs}} = Y_{\text{obs}}$. To obtain an interesting problem instance that is compatible with our setting, we generate a sub-dataset with higher rating density. To this end, we sample and pre-process the original data (details can be found in Appendix C), and obtain a reduced dataset of $N = 245$ users and $P = 51$ items with rating density 0.64. All of the N users were connected via a directed trust network \mathcal{G} with binary adjacency matrix W , i.e., the link W_{ij} from user i to user j indicates i was trusted by j , and hence user i 's opinion would impact that of user j . The resulting adjacency matrix W is shown in Fig. 3(a) and the degree distribution of the symmetrized undirected graph with $A = (W + W^T)/2$ can be found in Fig. 3(b). Like in the previous experiments, we adopt $S = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ as GSO. Following a data pre-processing step, the resulting centered ratings matrix Y_{obs} with values in $[-2, 2]$ is shown in Fig. 3(c). Given that the ground-truth sparsity of the source θ_{sr} is unknown, we examine different sparsity level assumptions on the input, namely, $\theta_{\text{sr}} \in \{0.1, 0.2, 0.3, 0.4\}$. Specifically, to populate the source signal X_{sr} we retain different proportions θ_{sr} of the earliest ratings per item $p = 1, \dots, P$. This way θ_{sr} is closely related to S as defined in Section 2.3.

To assess the source recovery performance, we take S and Y_{obs} as inputs and compare three approaches: two methods for graph-aware blind deconvolution including the proposed estimator (8) and the

lifting approach in [17], plus the non-convex ℓ_1 recovery algorithm for source localization on graphs (SLG) [28]. Since the support of sources X_{sr} is a strict subset of the support of Y_{obs} , we consider the area under the curve (AUC) of the predicted sources in \hat{X} as figure of merit in this numerical test case. In addition to the aforementioned three methods, we also consider a naive baseline whereby $\hat{X} = Y_{\text{obs}}$. The resulting AUCs for different source signal sparsity levels θ_{sr} are shown in Fig. 3(d). It can be seen that Algorithm 1 achieves the highest AUC for the denser settings $\theta_{\text{sr}} = 0.3, 0.4$; for $\theta_{\text{sr}} = 0.2$, the proposed approach and SLG attain a similar AUC (which is higher than the naive baseline); for $\theta_{\text{sr}} = 0.1$, the proposed method performs marginally worse than all of the other three predictors. The matrix lifting algorithm is only competitive when the sources are the sparsest.

Overall, we find that the proposed estimator (8) offers more robust predictions across a broader range of input sparsity levels, a promising finding to support the prospect of solving real world network deconvolution problems. In all fairness though, the performance of none of the methods is stellar. But we note these real data are complex and we lack a ground truth for validation, since the chronological order alone does not imply causality. Secondly, even through there are a few earliest ratings that impact other ratings and hence should be reasonably viewed as sources, the observations could still be highly noisy.

6. Concluding summary and future work

We studied the problem of blind graph filter identification from multiple sparse inputs, which extends blind deconvolution of time (or spatial) domain signals to graphs. By introducing an assumption on invertibility of the graph filter, we obtained a computationally simpler convex relaxation for (diffused) source localization in the multi-signal case. In terms of theoretical analyses, we first derived sufficient conditions for exact recovery, which hold with high probability under a Bernoulli-Gaussian model for the sparse inputs. A stable recovery result is then established, ensuring the estimation error on the inverse filter's frequency response is manageable when the observations are corrupted by a small amount of noise.

Ongoing work includes additional analyses on the *robustness* of the proposed approach to imperfections in the observed graph, as well as when measurements are collected only in a fraction of nodes. On the algorithmic side, developing an *online* network source localization method capable of processing streaming graph signal observations is also of interest.

CRedit authorship contribution statement

Chang Ye: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Conceptualization. **Gonzalo Mateos:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Appendix A. Proof of Theorem 1

Recall the notation introduced in Section 4.1. The proof of Theorem 1 begins by considering an equivalent problem to (10), obtained via an invertible change of variable $\mathbf{w} = \tilde{\mathbf{g}}\mathbf{h}_0$, namely

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathcal{P}(\mathbf{w})\mathbf{X}\|_{1,1}, \text{ s. to } \tilde{\mathbf{r}}^\top \mathbf{w} = c, \quad (\text{A.1})$$

where $\tilde{\mathbf{r}}^\top = \mathbf{r}^\top \operatorname{diag}(\tilde{\mathbf{g}}_0)$. Note that the solution candidate $\hat{\mathbf{w}} = \mathbf{1}_N$ implies $\tilde{\mathbf{g}} = \tilde{\mathbf{g}}_0$ in (10), so we let $c = \tilde{\mathbf{r}}^\top \mathbf{1}_N$. Then we have the following proposition, which simply restates Theorem 1 in terms of the equivalent problem (A.1).

Proposition 1 (Exact Recovery for the Equivalent Problem). *Consider graph signal observations $\mathbf{Y} = \mathcal{P}(\tilde{\mathbf{h}}_0)\mathbf{X}_0 \in \mathbb{R}^{N \times P}$, where \mathbf{X}_0 adheres to the Bernoulli-Gaussian model with $\theta \in (0, 0.324]$. Recall that under Assumption 1, we can write $\mathbf{X}_0 = \mathcal{P}(\tilde{\mathbf{g}}_0)\mathbf{Y}$. Let $P \geq C' \sigma_m^{-2} \log \frac{4}{\delta}$, where $\sigma_m = \min(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ and $\sigma_1 \in \left(0, \frac{\sqrt{\pi}\theta^{3/2}}{2}\right]$, $\sigma_2 \in \left(0, \frac{\sqrt{\pi}\theta}{2}\right]$, $\sigma_3 > 0$, $\sigma_4 \in (0, 1)$, $\delta \in (0, 1)$ are parameters, while C' is a constant that does not depend on P , σ_m , or δ . Then $\hat{\mathbf{w}} = \mathbf{1}_N$ is the unique solution to (A.1) with probability at least $1 - \delta$, if*

$$\left\| \mathbf{P}_1^\perp \tilde{\mathbf{r}} \right\|_2 \leq (c/N)d_0, \quad (\text{A.2})$$

$$\text{where } d_0 := \frac{\sqrt{1 - \sigma_{\max}^2(\mathbf{0})[(1 - \sigma_1) - 2\theta(1 + \sigma_2)](1 - \sigma_4)}}{(1 + \sigma_3)\sqrt{\theta}}.$$

To establish Proposition 1 (and thus Theorem 1), we will argue that for all feasible perturbations $\delta \in \mathbb{R}^N$ such that $\tilde{\mathbf{r}}^\top \delta = 0$, then $\|\mathcal{P}(\mathbf{1}_N + \delta)\mathbf{X}\|_{1,1} \geq \|\mathbf{X}\|_{1,1}$ holds for any Bernoulli-Gaussian $\mathbf{X} \in \mathbb{R}^{N \times P}$ with high probability. Our approach in the proof of Proposition 1 at the end of this appendix is to compute and bound the sub-gradient of $\|\mathcal{P}(\mathbf{1}_N + \delta)\mathbf{X}\|_{1,1}$ at $\delta = \mathbf{0}_N$. To leverage the independence of some random variables, it will be convenient to decompose the sub-gradient expression using a hollow matrix $\mathcal{M}(\delta)$ as $\mathcal{P}(\delta)$ with 0 diagonal elements (i.e., $\mathcal{M}(\delta)_{ij} = \mathcal{P}(\delta)_{ij}, \forall i \neq j$; $\mathcal{M}(\delta)_{ii} = 0, \forall i$). The key to our bounding strategy in the proof of Proposition 1 is to derive a concentration property for hollow matrices [cf. (A.10)], that follows from Proposition 2.

Proposition 2. *Consider vectors $\mathbf{m}_i \in \mathbb{R}^N$, $i \in \{1, \dots, N\}$, such that $[\mathbf{m}_i]_i = 0$. Suppose $\mathbf{X} \in \mathbb{R}^{N \times P}$ is drawn from the Bernoulli-Gaussian model in Definition 1, with $\theta \in (0, e^{-1}]$. Given parameters $\delta \in (0, 1)$, $\sigma_1 \in \left(0, \frac{\sqrt{\pi}\theta^{3/2}}{2}\right]$, and $\sigma_2 \in \left(0, \frac{\sqrt{\pi}\theta}{2}\right]$, let $P \geq \frac{C}{\min(\sigma_1^2, \sigma_2^2)} \log \frac{4}{\delta}$ for some constant C .*

Then, for each $i \in \{1, \dots, N\}$ we have

$$\begin{aligned} \text{(a) Pr} \left[\left| \frac{1}{\beta_i P} \|\mathbf{m}_i^\top \mathbf{X}\|_1 - \|\mathbf{m}_i\|_2 \right| \leq \sigma_1 \|\mathbf{m}_i\|_2 \right] &\geq 1 - \delta \\ \text{(b) Pr} \left[\left| \frac{1}{\beta_i \theta P} \|(\omega_i^\top) \circ (\mathbf{m}_i^\top \mathbf{X})\|_1 - \|\mathbf{m}_i\|_2 \right| \leq \sigma_2 \|\mathbf{m}_i\|_2 \right] &\geq 1 - \delta, \end{aligned} \quad (\text{A.3})$$

where $\beta_i := \mathbb{E}[\|\mathbf{m}_i^\top \mathbf{x}_p\|] / \|\mathbf{m}_i\|_2$, with $\mathbf{x}_p = [x_{1p}, \dots, x_{Np}]^\top \in \mathbb{R}^N$ denoting the p th column of \mathbf{X} and $\omega_i = [\Omega_{i1}, \dots, \Omega_{ip}]^\top \in \mathbb{R}^P$ is the i th row of Ω .

The idea behind (a) in (A.3) comes from [44, Theorem 5.1], namely that the absolute value of linear combinations of i.i.d. Bernoulli-Gaussian $\{X_{ij}\}$ will concentrate to its expectation. For (b), note that Ω_{ip} and $|\mathbf{m}_i^\top \mathbf{x}_p|$ are independent $\forall i, p$, when \mathbf{m}_i is a hollow vector. This implies

$$\begin{aligned} \mathbb{E}[\|(\omega_i^\top) \circ (\mathbf{m}_i^\top \mathbf{X})\|_1] &= \mathbb{E} \left[\sum_{p=1}^P \Omega_{ip} |\mathbf{m}_i^\top \mathbf{x}_p| \right] \\ &= \mathbb{E}_{\mathbf{X}} \left[\sum_{p=1}^P \mathbb{E}[\Omega_{ip}] |\mathbf{m}_i^\top \mathbf{x}_p| \right] \\ &= \theta \mathbb{E}_{\mathbf{X}} \left[\sum_{p=1}^P |\mathbf{m}_i^\top \mathbf{x}_p| \right] = \theta P \beta_i \|\mathbf{m}_i\|_2. \end{aligned}$$

To prove Proposition 2, we first establish preliminary Lemmata 1 and 2 that are similar to [44, Lemma 5.3], and then prove Lemmata 3 and 4 that follow ideas from [44, Proposition 5.2]. Specifically, for a hollow matrix \mathbf{M} and Bernoulli-Gaussian distributed \mathbf{X} where $S := \operatorname{supp}(\mathbf{X}) = \operatorname{supp}(\Omega)$, let $\tilde{\mathbf{X}} := \mathbf{M}\mathbf{X}$ and $\tilde{\mathbf{X}}^{(S)} := \frac{1}{\theta} \Omega \circ (\mathbf{M}\mathbf{X})$, with $\tilde{\mathbf{M}} := [\mathbf{m}_1 / \|\mathbf{m}_1\|_2, \dots, \mathbf{m}_N / \|\mathbf{m}_N\|_2]^\top \in \mathbb{R}^{N \times N}$. In the first step we show that the absolute value of the entries of both of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{X}}^{(S)}$ have bounded expectation and variance (cf. Lemmata 1 and 2). Then in the second step, Lemma 3 shows that $|\tilde{X}_{ip}|$ (and $|\tilde{X}_{ip}^{(S)}|$) have uniform exponential tails, meaning that for some constant $b' > 0$ and all $t \geq 0$ we have $\Pr[|\tilde{X}_{ip}| \geq t] \leq e^{-b't}$. Finally and using Lemmata 1–3, Lemma 4 establishes sub-Gaussian tail bounds for the centered and scaled ℓ_1 norms of the rows of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{X}}^{(S)}$. Lemma 4 is then instrumental to show that the ℓ_1 norms of the rows of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{X}}^{(S)}$ concentrate around their means, as asserted in Proposition 2.

Lemma 1. *Consider vectors $\tilde{\mathbf{m}}_i \in \mathbb{R}^N$ $i \in \{1, \dots, N\}$, such that $[\tilde{\mathbf{m}}_i]_i = 0$, $\|\tilde{\mathbf{m}}_i\|_\infty = \tilde{\alpha}_i$, and $\|\tilde{\mathbf{m}}_i\|_2 = 1$. Let $\tilde{\mathbf{X}}^{(S)} = \frac{1}{\theta} \Omega \circ (\mathbf{M}\mathbf{X}) \in \mathbb{R}^{N \times P}$, where $\mathbf{X} \in \mathbb{R}^{N \times P}$ is drawn from the Bernoulli-Gaussian model in Definition 1, with $\theta \in (0, e^{-1})$, and $\tilde{\mathbf{M}} = [\tilde{\mathbf{m}}_1, \dots, \tilde{\mathbf{m}}_N]^\top \in \mathbb{R}^{N \times N}$. Then $\mathbb{E}[|\tilde{X}_{ip}^{(S)}|^2] = \frac{1}{\theta}$, and $\beta_0(1 - \sigma'_i) \leq \mathbb{E}[|\tilde{X}_{ip}^{(S)}|] \leq \beta_0, \forall (i, p)$, where $\beta_0 := \mathbb{E}[|\gamma_{ip}|] = \sqrt{2/\pi}$ is the expectation of a folded standard Normal distributed random variable $|\gamma_{ip}|$ and*

$$\sigma'_i = \begin{cases} \frac{\tilde{\alpha}_i^2}{\theta}, & \tilde{\alpha}_i \in (0, \sqrt{\theta}] \cap [N^{-1/2}, 1], \\ 1 - \sqrt{\theta} \tilde{\alpha}_i \left[1 + \frac{(1 - \tilde{\alpha}_i^2)\theta^2}{2\tilde{\alpha}_i^2(1 - \theta)} \right], & \tilde{\alpha}_i \in (\sqrt{\theta}, 1] \cap [N^{-1/2}, 1]. \end{cases} \quad (\text{A.4})$$

Proof of Lemma 1. Notice that for any (i, p) , $\tilde{M}_{ii} = 0$, so $\tilde{X}_{ip}^{(S)} = \frac{1}{\theta} \Omega_{ip} \sum_{j=1}^N \tilde{M}_{ij} X_{jp} = \frac{1}{\theta} \Omega_{ip} \sum_{j \neq i} \tilde{M}_{ij} \Omega_{jp} \gamma_{jp} \theta^{-1/2}$. By conditioning on $Z_{ip} = \sum_{j \neq i} \Omega_{jp} \tilde{M}_{ij}^2 = z$, then $\tilde{X}_{ip} = \theta^{-1/2} \sum_{j \neq i} \tilde{M}_{ij} \Omega_{jp} \gamma_{jp}$ is distributed as $\text{Normal}(0, \frac{z}{\theta})$. Also notice that $\theta \tilde{X}_{ip}^{(S)} = \Omega_{ip} \tilde{X}_{ip}$, where Ω_{ip} and \tilde{X}_{ip} are independent. Then $\mathbb{E}[|\theta \tilde{X}_{ip}^{(S)}| | Z_{ip} = z] = \theta \beta_0 \sqrt{\frac{z}{\theta}}$, and thus $\mathbb{E}[|\tilde{X}_{ip}^{(S)}|] = \theta^{-1/2} \beta_0 \mathbb{E}[\sqrt{Z_{ip}}]$. Following ideas in the proof of [44, Lemma 5.3], we can derive upper and lower bounds for $\mathbb{E}[\sqrt{Z_{ip}}]$.

For the upper bound, we have $\mathbb{E}[\sqrt{Z_{ip}}] \leq \sqrt{\mathbb{E}[Z_{ip}]} \leq \sqrt{\theta}$. For the lower bound, we consider two cases, $\tilde{\alpha}_i < \sqrt{\theta}$ and $\tilde{\alpha}_i \geq \sqrt{\theta}$. When $\tilde{\alpha}_i < \sqrt{\theta}$, we use the strategy in [44, Lemma 5.3]. Let $t = \frac{Z_{ip}}{\theta} - 1 \geq -1$, and recall that $\sqrt{1+t} \geq 1 + \frac{t}{2} - t^2$, hence

$$\begin{aligned} \mathbb{E}[\sqrt{Z_{ip}}] &= \sqrt{\theta} \mathbb{E}[\sqrt{1+t}] \\ &\geq \sqrt{\theta} \left(1 + \mathbb{E}\left[\frac{Z_{ip}}{\theta} - 1\right] - \mathbb{E}\left[\left(\frac{Z_{ip}}{\theta} - 1\right)^2\right] \right) \\ &= \sqrt{\theta} (1 - \theta^{-2} \operatorname{var}[Z_{ip}]). \end{aligned}$$

Now $\operatorname{var}[Z_{ip}] = \sum_{j \neq i} \tilde{M}_{ij}^4 \operatorname{var}[\Omega_{jp}] \leq \sum_{j \neq i} \tilde{M}_{ij}^4 \theta \leq \theta \tilde{\alpha}_i^2 \sum_{j \neq i} \tilde{M}_{ij}^2 = \theta \tilde{\alpha}_i^2$ (as $\tilde{\mathbf{m}}_i$ has unit ℓ_2 norm). Hence, $\mathbb{E}[\sqrt{Z_{ip}}] \geq \sqrt{\theta}(1 - \frac{\tilde{\alpha}_i^2}{\theta}) \geq \sqrt{\theta}(1 - \sigma'_i)$, where $\sigma'_i = \frac{\tilde{\alpha}_i^2}{\theta}$. As a result, $\beta_0(1 - \sigma'_i) \leq \mathbb{E}[|\tilde{X}_{ip}^{(S)}|] \leq \beta_0, \forall (i, p)$, with $\sigma'_i = \tilde{\alpha}_i^2/\theta < 1$.

When $\tilde{\alpha}_i \geq \sqrt{\theta}$, we let k bet the index associated with the maximum entry of $\tilde{\mathbf{m}}_i$ (i.e., $\tilde{\alpha}_i = \|\tilde{\mathbf{m}}_i\|_\infty = |\tilde{M}_{ik}|$). Then, we have

$$\begin{aligned} \mathbb{E}[\sqrt{Z_{ip}}] &= \mathbb{E} \left[\sqrt{\Omega_{kp} \tilde{M}_{ik}^2 + \sum_{j \neq i, k} \Omega_{jp} \tilde{M}_{ij}^2} \right] \\ &\geq \mathbb{E} \left[\sqrt{\Omega_{kp} \tilde{M}_{ik}^2 + \sum_{j \neq i, k} \Omega_{jp} \tilde{M}_{ij}^2} \mid \Omega_{kp} = 1 \right] \Pr[\Omega_{kp} = 1] \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{=} \theta \bar{\alpha}_i \mathbb{E} \left[\sqrt{1 + \frac{1-\theta}{\theta} t} \right] \geq \theta \bar{\alpha}_i \mathbb{E} \left[\sqrt{e^t} \right] \geq \theta \bar{\alpha}_i \mathbb{E} \left[1 + \frac{t}{2} \right] \\
&= \theta \bar{\alpha}_i \mathbb{E} \left[1 + \frac{\theta}{2(1-\theta)} \sum_{j \neq i, k} \frac{\Omega_{jp} \bar{M}_{ij}^2}{\bar{\alpha}_i^2} \right] \\
&\stackrel{(b)}{=} \sqrt{\theta}(1 - \bar{\sigma}_i)
\end{aligned}$$

In (a) we let $t = \frac{\theta}{1-\theta} \sum_{j \neq i, k} \frac{\Omega_{jp} \bar{M}_{ij}^2}{\bar{\alpha}_i^2} \leq 1$, as $\bar{\alpha}_i^2 \geq \theta$. And in (b), we let $\bar{\sigma}_i = 1 - \sqrt{\theta} \bar{\alpha}_i \left[1 + \frac{(1-\bar{\alpha}_i^2)\theta^2}{2\bar{\alpha}_i^2(1-\theta)} \right]$. It can be checked that $f(\bar{\alpha}_i, \theta) := 1 - \bar{\sigma}_i = \sqrt{\theta} \bar{\alpha}_i \left[1 + \frac{(1-\bar{\alpha}_i^2)\theta^2}{2\bar{\alpha}_i^2(1-\theta)} \right] \in [\theta + \frac{\theta^2}{2}, \sqrt{\theta}]$, for $\bar{\alpha}_i \geq \sqrt{\theta}$ and $\theta \leq e^{-1}$. As a result, for any $\bar{\alpha}_i \in [\sqrt{\theta}, 1]$, we get $\bar{\sigma}_i = 1 - f(\bar{\alpha}_i, \theta) \in [1 - \sqrt{\theta}, 1 - \theta - \frac{\theta^2}{2}]$. So $\bar{\sigma}_i \in (0, 1)$ is a feasible lower bound for $\bar{\alpha}_i \geq \sqrt{\theta}$ with $\theta \leq e^{-1}$.

Note that the feasible range of $\bar{\alpha}_i$ is $[N^{-1/2}, 1]$, and we do not know whether $\sqrt{\theta} \leq e^{-1/2} \approx 0.6$ would be within it. So we apply the following strategy to select the lower bound σ'_i , i.e.,

$$\sigma'_i = \begin{cases} \frac{\bar{\alpha}_i^2}{\theta}, & \bar{\alpha}_i \in (0, \sqrt{\theta}] \cap [N^{-1/2}, 1] \\ 1 - f(\bar{\alpha}_i, \theta), & \bar{\alpha}_i \in (\sqrt{\theta}, 1] \cap [N^{-1/2}, 1]. \end{cases} \quad (\text{A.5})$$

Besides, we have $\mathbb{E} \left[|\theta \bar{X}_{ip}^{(S)}|^2 \right] = \mathbb{E} \left[|\Omega_{ip} \bar{X}_{ip}|^2 \right] = \mathbb{E} \left[\Omega_{ip} \right] \mathbb{E} \left[|\bar{X}_{ip}|^2 \right] = \theta$, so $\mathbb{E} \left[|\bar{X}_{ip}^{(S)}|^2 \right] = \frac{1}{\theta}$ as desired.

Lemma 2. Consider vectors $\bar{\mathbf{m}}_i \in \mathbb{R}^N$ $i \in \{1, \dots, N\}$, such that $[\bar{\mathbf{m}}_i]_i = 0$, $\|\bar{\mathbf{m}}_i\|_\infty = \bar{\alpha}_i$, and $\|\bar{\mathbf{m}}_i\|_2 = 1$. Let $\bar{\mathbf{X}} = \bar{\mathbf{M}}\mathbf{X} \in \mathbb{R}^{N \times P}$, where $\mathbf{X} \in \mathbb{R}^{N \times P}$ is drawn from the Bernoulli-Gaussian model in Definition 1 with $\theta \in (0, e^{-1})$, and $\bar{\mathbf{M}} = [\bar{\mathbf{m}}_1, \dots, \bar{\mathbf{m}}_N]^T \in \mathbb{R}^{N \times N}$. Then $\mathbb{E} \left[|\bar{X}_{ip}|^2 \right] = 1$ and $\beta_0(1 - \sigma'_i) \leq \mathbb{E} \left[|\bar{X}_{ip}| \right] \leq \beta_0, \forall (i, p)$, where σ'_i is given by (A.4) and $\beta_0 := \mathbb{E} \left[|\gamma_{ip}| \right] = \sqrt{2/\pi}$ is the expectation of a folded standard Normal distributed random variable $|\gamma_{ip}|$.

Proof of Lemma 2. Because $\bar{M}_{ii} = 0$, $\bar{X}_{ip} = \sum_{j=1}^N \bar{M}_{ij} X_{jp} = \sum_{j \neq i} \bar{M}_{ij} \Omega_{jp} \gamma_{jp} \theta^{-1/2}$, for all (i, p) . As $\bar{\mathbf{m}}_i$ has unit ℓ_2 norm, upon conditioning on $Z_{ip} = \sum_{j \neq i} \Omega_{jp} \bar{M}_{ij}^2 = z$, \bar{X}_{ip} is distributed as $\text{Normal}(0, \frac{z}{\theta})$. Then $\mathbb{E} \left[|\bar{X}_{ip}| \mid Z_{ip} = z \right] = \beta_0 \sqrt{\frac{z}{\theta}}$, and $\mathbb{E} \left[|\bar{X}_{ip}| \right] = \theta^{-1/2} \beta_0 \mathbb{E} \left[\sqrt{Z_{ip}} \right]$. As Z_{ip} here is defined exactly as in the proof of Lemma 1, we also have $\sqrt{\theta}(1 - \sigma'_i) \leq \mathbb{E} \left[Z_{ip} \right] \leq \sqrt{\theta}$, with σ'_i defined as in (A.5). Notice that we have the same bounds for $\mathbb{E} \left[|\bar{X}_{ip}| \right]$ and $\mathbb{E} \left[|\bar{X}_{ip}^{(S)}| \right]$. However, for $\mathbb{E} \left[|\bar{X}_{ip}|^2 \right]$, the result will be different, as $\mathbb{E} \left[|\bar{X}_{ip}|^2 \right] = 1$. ■

From Lemmata 1 and 2, we know for an arbitrary hollow-matrix $\bar{\mathbf{M}}$ and Bernoulli-Gaussian distributed \mathbf{X} with $\theta \in (0, e^{-1})$, the product $\bar{\mathbf{X}} = \bar{\mathbf{M}}\mathbf{X}$ and its masked version $\bar{\mathbf{X}}^{(S)}$ have entries whose absolute values are bounded in expectation. Next, Lemma 3 establishes their entries have uniform exponential tails.

Lemma 3. The elements of both $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^{(S)}$ have uniform exponential tails. Specifically, $\Pr \left[|\bar{X}_{ip}| > t \right] \leq e^{-ut+O(u^2)}, \forall u \in [0, \sqrt{2\theta}/\bar{\alpha}_i], t \geq 0$. Likewise, $\Pr \left[|\bar{X}_{ip}^{(S)}| > t \right] \leq e^{-ut+O(u^2)}, \forall u \in [0, \sqrt{2\theta^3}/\bar{\alpha}_i], t \geq 0$.

Proof of Lemma 3. To show that $\bar{X}_{ip}^{(S)}$ has a uniform exponential tail for all (i, p) , we consider $0 \leq u \leq u_i = \sqrt{2\theta^3}/\bar{\alpha}_i$,

$$\begin{aligned}
\mathbb{E} \left[e^{u \bar{X}_{ip}^{(S)}} \right] &= \mathbb{E} \left[e^{\frac{u \Omega_{ip}}{\theta} \sum_{j \neq i} \bar{M}_{ij} X_{jp}} \right] \\
&= \prod_{j \neq i} \mathbb{E} \left[e^{\frac{u \Omega_{ip} \Omega_{jp}}{\theta^{3/2}} \bar{M}_{ij} \gamma_{jp}} \right] \\
&= \prod_{j \neq i} \mathbb{E}_{\gamma_{jp}} \left[\theta \left[\theta e^{\frac{u}{\theta^{3/2}} \bar{M}_{ij} \gamma_{jp}} + (1-\theta) \right] + (1-\theta) \right]
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \prod_{j \neq i} \left[\theta \left[\theta (1 + u^2 \bar{M}_{ij}^2 / \theta^3) + (1-\theta) \right] + (1-\theta) \right] \\
&= \prod_{j \neq i} (1 + \theta^{-1} u^2 \bar{M}_{ij}^2) \leq \prod_{j \neq i} e^{\frac{1}{\theta} u^2 \bar{M}_{ij}^2} = e^{\frac{u^2}{\theta}}. \quad (\text{A.6})
\end{aligned}$$

Notice that in (a), we have applied: (i) $\mathbb{E} \left[e^{uX} \right] = e^{\frac{u^2}{2}}, \forall u \geq 0$ when $X \sim \text{Normal}(0, 1)$, so $\mathbb{E} \left[e^{u\theta^{-3/2} \bar{M}_{ij} \gamma_{jp}} \right] = e^{u\theta^{-3/2} \bar{M}_{ij}^2 / 2} = e^{\frac{u^2}{2\theta^3} \bar{M}_{ij}^2}$ because $\gamma_{jp} \sim \text{Normal}(0, 1)$; and then (ii) $e^\tau \leq 1 + 2\tau, \forall \tau \in [0, 1]$ as $\tau = \frac{u^2}{2\theta^3} \bar{M}_{ij}^2 \leq \frac{u^2}{2\theta^3} \bar{\alpha}_i^2 \leq 1$. All in all, $\forall u \in (0, u_i)$ and $\forall t \geq 0$ we have

$$\Pr \left[\bar{X}_{ip}^{(S)} \geq t \right] = \Pr \left[e^{u \bar{X}_{ip}^{(S)}} \geq e^{ut} \right] \leq e^{-ut} \mathbb{E} \left[e^{u \bar{X}_{ip}^{(S)}} \right] \leq e^{-ut + \frac{u^2}{\theta}}.$$

By symmetry, it follows $\Pr \left[\bar{X}_{ip}^{(S)} \leq -t \right] \leq e^{-ut + \frac{u^2}{\theta}}$. So $\bar{X}_{ip}^{(S)}, \forall (i, p)$ has a uniform exponential tail, i.e., $\Pr \left[|\bar{X}_{ip}^{(S)}| > t \right] \leq e^{-ut+O(u^2)}, \forall u \in [0, \sqrt{2\theta^3}/\bar{\alpha}_i]$, and all $t \geq 0$.

To show that \bar{X}_{ip} also has a uniform exponential tail for all (i, p) , we follow a similar approach and consider $0 \leq u \leq u_i = \sqrt{2\theta}/\bar{\alpha}_i$ to obtain

$$\begin{aligned}
\mathbb{E} \left[e^{u \bar{X}_{ip}} \right] &= \mathbb{E} \left[e^{u \sum_{j \neq i} \bar{M}_{ij} X_{jp}} \right] \\
&= \prod_{j \neq i} \mathbb{E} \left[e^{\frac{u \Omega_{jp}}{\theta^{1/2}} \bar{M}_{ij} \gamma_{jp}} \right] \\
&= \prod_{j \neq i} \mathbb{E}_{\gamma_{jp}} \left[\theta e^{\frac{u}{\theta^{1/2}} \bar{M}_{ij} \gamma_{jp}} + (1-\theta) \right] \\
&\stackrel{(b)}{\leq} \prod_{j \neq i} \left[\theta (1 + u^2 \bar{M}_{ij}^2 / \theta) + (1-\theta) \right] \\
&= \prod_{j \neq i} (1 + u^2 \bar{M}_{ij}^2) \leq \prod_{j \neq i} e^{u^2 \bar{M}_{ij}^2} = e^{u^2}
\end{aligned}$$

Notice that in (b), we applied the same inequalities as (a) in (A.6). Then $\forall u \in (0, u_i)$ and $\forall t \geq 0$,

$$\Pr \left[\bar{X}_{ip} \geq t \right] = \Pr \left[e^{u \bar{X}_{ip}} \geq e^{ut} \right] \leq e^{-ut} \mathbb{E} \left[e^{u \bar{X}_{ip}} \right] \leq e^{-ut+u^2}$$

By symmetry, $\Pr \left[\bar{X}_{ip} \leq -t \right] \leq e^{-ut+u^2}$ as well. So $\bar{X}_{ip}, \forall (i, p)$ has a uniform exponential tail, i.e., $\Pr \left[|\bar{X}_{ip}| > t \right] \leq e^{-ut+O(u^2)}, \forall u \in [0, \sqrt{2\theta}/\bar{\alpha}_i]$, and all $t \geq 0$. ■

The key ingredient needed to prove Proposition 2 is Lemma 4, which derives sub-Gaussian tail bounds for the centered and scaled ℓ_1 norms of the i th rows of $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^{(S)}$. To this end, we will rely on the established Lemmata 1–3.

Lemma 4. Let $\check{x}_i^{(S)} := \frac{1}{\sqrt{P}} \left(\sum_{p=1}^P |\bar{X}_{ip}^{(S)}| - P\beta_i \right)$ and $\check{x}_i := \frac{1}{\sqrt{P}} \left(\sum_{p=1}^P |\bar{X}_{ip}| - P\beta_i \right)$. Then both $\check{x}_i^{(S)}$ and \check{x}_i have sub-Gaussian tails (as defined in [44, Definition 2.1]), up to $\frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i}} \sqrt{P}$ and $\frac{\theta}{\sqrt{2\bar{\alpha}_i}} \sqrt{P}$, respectively.

Proof of Lemma 4. Firstly, note that $\mathbb{E} \left[\bar{X}_{ip} \right] = 0$ and from Lemma 2 it follows that $\text{var} \left[\bar{X}_{ip} \right] = 1, \mathbb{E} \left[|\bar{X}_{ip}| \right] = \beta_i$. Moreover, according to Lemma 3, \bar{X}_{ip} has an exponential tail, i.e., $\Pr \left[|\bar{X}_{ip}| > t \right] \leq e^{-ut+O(u^2)}, \forall u \in [0, \sqrt{2\theta}/\bar{\alpha}_i]$, and all $t \geq 0$. From [44, Proposition 5.2, Lemma 2.3], it can be shown that $\forall i, \check{x}_i$ has a sub-Gaussian tail up to $\frac{\theta}{\sqrt{2\bar{\alpha}_i}} \sqrt{P}$, i.e., $\mathbb{E} \left[e^{u \check{x}_i} \right] < e^{O(u^2)}$ and $\mathbb{E} \left[e^{-u \check{x}_i} \right] < e^{O(u^2)}, \forall u \in \left(0, \frac{\theta}{\sqrt{2\bar{\alpha}_i}} \sqrt{P} \right)$.

For $\check{x}_i^{(S)}$, we also have $\mathbb{E} \left[\bar{X}_{ip}^{(S)} \right] = 0$ and $\mathbb{E} \left[|\bar{X}_{ip}^{(S)}| \right] = \beta_i$. From Lemma 1 we know the only difference with \bar{X}_{ip} is $\text{var} \left[\bar{X}_{ip}^{(S)} \right] = \frac{1}{\theta}$. So we can still apply [44, Proposition 5.2, Lemma 2.3] and show $\forall i, \check{x}_i^{(S)}$ has a sub-Gaussian tail up to $\frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i}} \sqrt{P}$, i.e., $\mathbb{E} \left[e^{u \check{x}_i^{(S)}} \right] < e^{O(u^2)}$ and $\mathbb{E} \left[e^{-u \check{x}_i^{(S)}} \right] < e^{O(u^2)}, \forall u \in \left(0, \frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i}} \sqrt{P} \right)$. ■

Now, using [Lemma 4](#) we establish that the ℓ_1 norms of the rows of $\bar{\mathbf{X}}$ and $\bar{\mathbf{X}}^{(S)}$ concentrate around their means, as asserted in [Proposition 2](#).

Proof of Proposition 2. To show (b), from [Lemma 4](#) we know $\check{x}_i^{(S)} = \frac{1}{\sqrt{P}} \left(\sum_{p=1}^P |\bar{X}_{ip}^{(S)}| - \beta_i P \right) = \frac{1}{\sqrt{P}} \left(\frac{1}{\theta} \|(\omega_i^\top) \circ (\bar{\mathbf{m}}_i^\top \mathbf{X})\|_1 - P\beta_i \right)$ has a sub-Gaussian tail up to $\frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i}} \sqrt{P}$. As a result, for some $\sigma_2 \in (0, 1)$, we have

$$\Pr \left[\frac{1}{\theta \beta_i P} \|(\omega_i^\top) \circ (\bar{\mathbf{m}}_i^\top \mathbf{X})\|_1 \geq 1 + \sigma_2 \right] = \Pr \left[\check{x}_i^{(S)} \geq \beta_i \sqrt{P} \sigma_2 \right] \leq e^{-C' \beta_i^2 P \sigma_2^2},$$

for some constant $C' > 0$. We require $\beta_i \sqrt{P} \sigma_2 \leq \frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i}} \sqrt{P} \Rightarrow \sigma_2 \leq \min \left\{ \frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i \beta_i}}, 1 \right\}$. Because $\theta \in (0, e^{-1}]$ and $\frac{1}{\sqrt{2\bar{\alpha}_i \beta_i}} \geq \frac{1}{\sqrt{2}\beta_0} = \frac{\sqrt{\pi}}{2}$, we have $\frac{\sqrt{\pi} \theta^{3/2}}{2} \leq \min \left\{ \frac{\theta^{3/2}}{\sqrt{2\bar{\alpha}_i \beta_i}}, 1 \right\}$. So it suffices to select $\sigma_2 \in \left(0, \frac{\sqrt{\pi} \theta^{3/2}}{2} \right]$. Now for $e^{-C' \beta_i^2 P \sigma_2^2} \leq \frac{\delta}{2}$, we have $P \geq \frac{C}{\sigma_2^2} \log \frac{2}{\delta}$, where $C \geq \frac{1}{C' \beta_i^2}$ is some constant. Putting all pieces together we have $\Pr \left[\frac{1}{\theta \beta_i P} \|(\omega_i^\top) \circ (\bar{\mathbf{m}}_i^\top \mathbf{X})\|_1 \leq 1 - \sigma_2 \right] \leq \frac{\delta}{2}$ and [Proposition 2\(b\)](#) follows.

To show (a), notice that from [Lemma 4](#) we know $\check{x}_i = \frac{1}{\sqrt{P}} \left(\sum_{p=1}^P |\bar{X}_{ip}| - \beta_i P \right) = \frac{1}{\sqrt{P}} \left(\| \bar{\mathbf{m}}_i^\top \mathbf{X} \|_1 - P\beta_i \right)$ has a sub-Gaussian tail up to $\frac{\theta}{\sqrt{2\bar{\alpha}_i}} \sqrt{P}$. Following similar steps as above for (b), one can readily arrive at [Proposition 2\(a\)](#). ■

Having established [Proposition 2](#), we have almost all ingredients needed to prove [Proposition 1](#) (and thus [Theorem 1](#)). Before doing so, we introduce a final lemma to show that Bernoulli-Gaussian random variables have bounded-energy.

Lemma 5 (Bounded Energy). Let $\{X_{ip}\}$ be i.i.d. random variables drawn from the Bernoulli-Gaussian model in [Definition 1](#), with $\theta \in (0, e^{-1}]$. For any vector $\phi \in \mathbb{R}^N$ and some $\sigma_3 > 0$, if $P \geq \frac{\check{C}}{\beta_0^2 \sigma_3^2} \log \frac{2}{\delta}$ where \check{C} is some constant, we have

$$\left| \sum_{i,p} \phi_i |X_{ip}| \right| \leq (1 + \sigma_3) \left| \mathbb{E} \left[\sum_{i,p} \phi_i |X_{ip}| \right] \right| \quad (\text{A.7})$$

holds with probability at least $1 - \delta$.

Proof of Lemma 5. Note that for $\theta \in (0, e^{-1}]$, we have $\mathbb{E}[|X_{ip}|] = \sqrt{\theta} \beta_0$, $\mathbb{E}[|X_{ip}|^2] = 1$, and $\text{var}[|X_{ip}|] = 1 - \theta \beta_0^2 > 0$, for $i \in \{1, \dots, N\}$, $p \in \{1, \dots, P\}$. Besides, $|X_{ip}|$ has an upper sub-Gaussian tail since it has the folded Normal distribution and,

$$\Pr[|X_{ip}| \geq t] \leq e^{-\frac{1}{2} t^2},$$

Similar to [\[44, Lemma 2.4\]](#), one can show $\mathbb{E}[e^{u|X_{ip}|}] \leq e^{\sqrt{\theta} \beta_0 + C u^2}$, for some constant C and all $u > 0$. Now, let $\check{X}_i := \frac{1}{\sqrt{P}} \sum_{p=1}^P (|X_{ip}| - \sqrt{\theta} \beta_0)$, so $\mathbb{E}[\check{X}_i] = 0$, $\text{var}[\check{X}_i] = \mathbb{E}[\check{X}_i^2] = 1$. Following similar ideas as in the proof of [Lemma 4](#), one has

$$\mathbb{E}[e^{u \check{X}_i}] = \prod_{p=1}^P \mathbb{E}[e^{u(\sqrt{P}(|X_{ip}| - \sqrt{\theta} \beta_0))}] \leq (e^{C \frac{u^2}{P}})^P = e^{O(u^2)}.$$

An analogous calculation also yields $\mathbb{E}[e^{-u \check{X}_i}] \leq e^{O(u^2)}$.

Let $\check{\phi}_i = \phi_i / \|\phi\|_2^2$, for $i \in \{1, \dots, N\}$. From [\[44, Lemma 2.2\]](#), both $\check{Y}_1 := \sum_i \check{\phi}_i \check{X}_i$ and $\check{Y}_2 := -\sum_i \check{\phi}_i \check{X}_i$ have sub-Gaussian tails, i.e., $\Pr[\check{Y}_1 \geq t] \leq e^{-\check{C}_1 t^2}$ and $\Pr[\check{Y}_2 \geq t] \leq e^{-\check{C}_2 t^2}$, for some constants \check{C}_1, \check{C}_2 . Hence with probability at least $1 - e^{-\check{C}_1 t^2}$, we have

$$\begin{aligned} & \sum_{i=1}^N \frac{\check{\phi}_i}{\sqrt{P}} \sum_{p=1}^P (|X_{ip}| - \sqrt{\theta} \beta_0) < t \\ & \Rightarrow \sum_{i,p} \phi_i (|X_{ip}| - \sqrt{\theta} \beta_0) < \|\phi\|_2^2 \sqrt{P} t \end{aligned}$$

$$\Rightarrow \sum_{i,p} (\phi_i |X_{ip}| - \mathbb{E}[\phi_i |X_{ip}|]) < \sigma_3 \mathbb{E} \left[\sum_{i,p} \phi_i |X_{ip}| \right],$$

where $\sigma_3 := \frac{\sqrt{P} t}{\mathbb{E}[\sum_{i,p} \phi_i |X_{ip}|]} = \frac{t}{\sqrt{\theta} P \beta_0}$. If we let $\Pr[\check{Y}_1 \geq t] \leq e^{-\check{C}_1 t^2} \leq \frac{\delta}{2}$, then we will have $P \geq \frac{\check{C}_1}{\theta \beta_0^2 \sigma_3^2} \log \frac{2}{\delta}$. One can also derive the other bound from $\Pr[\check{Y}_2 \leq -t] \leq e^{-\check{C}_2 t^2}$, i.e., $-\sum_{i,p} \phi_i |X_{ip}| < -(1 - \sigma_3) \mathbb{E}[\sum_{i,p} \phi_i |X_{ip}|]$ with probability at least $1 - e^{-\check{C}_2 t^2}$. To finish the proof, set $\check{C} = \max(\check{C}_1, \check{C}_2) / \theta$. ■

Having established the key preliminary results in [Proposition 2](#) and [Lemma 5](#), we can now proceed to prove our main exact recovery result.

Proof of Proposition 1. To prove $\hat{\mathbf{w}} = \mathbf{1}_N$ is the unique solution to problem [\(A.1\)](#), it is equivalent to say for all feasible perturbations $\delta \in \mathbb{R}^N$ such that $\bar{\mathbf{r}}^\top \delta = 0$, then $\|\mathcal{P}(\mathbf{1}_N + \delta) \mathbf{X}\|_{1,1} \geq \|\mathbf{X}\|_{1,1}$ holds for any Bernoulli-Gaussian random $\mathbf{X} \in \mathbb{R}^{N \times P}$ with high probability. Here we define the hollow matrix $\mathcal{M}(\delta)$ as $\mathcal{P}(\delta)$ with 0 diagonal elements (i.e., $\mathcal{M}(\delta)_{ij} = \mathcal{P}(\delta)_{ij}, \forall i \neq j$; $\mathcal{M}(\delta)_{ii} = 0, \forall i$). We can compute the sub-gradient of $\|\mathcal{P}(\mathbf{1}_N + \delta) \mathbf{X}\|_{1,1}$ at $\delta = \mathbf{0}_N$ as,

$$\begin{aligned} \|\mathcal{P}(\mathbf{1}_N + \delta) \mathbf{X}\|_{1,1} & \geq \|\mathbf{X}\|_{1,1} + \|[\mathcal{M}(\delta) \mathbf{X}]_{S^c}\|_{1,1} + \sum_{p=1}^P \epsilon_p^\top \mathcal{P}(\delta) \mathbf{x}_p \\ & \geq \|\mathbf{X}\|_{1,1} + \|\mathcal{M}(\delta) \mathbf{X}\|_{1,1} - 2\|[\mathcal{M}(\delta) \mathbf{X}]_S\|_{1,1} + \sum_{i,p} |X_{ip}| \langle \mathbf{v}_i \circ \mathbf{v}_i, \delta \rangle. \end{aligned} \quad (\text{A.8})$$

Applying [Lemma 5](#), we can bound the last term in [\(A.8\)](#) as

$$\begin{aligned} \sum_{i,p} |X_{ip}| \langle \mathbf{v}_i \circ \mathbf{v}_i, \delta \rangle & \geq -(1 + \sigma_3) \left| \mathbb{E} \left[\sum_{i,p} \frac{\Omega_{ip} |\gamma_{ip}|}{\sqrt{\theta}} \langle \mathbf{v}_i \circ \mathbf{v}_i, \delta \rangle \right] \right| \\ & = -(1 + \sigma_3) \sqrt{\theta} P \beta_0 |\mathbf{1}_N^\top \delta|, \end{aligned} \quad (\text{A.9})$$

where we used $\beta_0 := \mathbb{E}[|\gamma_{ip}|] = \sqrt{\frac{2}{\pi}}$.

Next, we are going to bound $\|\mathcal{M}(\delta) \mathbf{X}\|_{1,1} - 2\|[\mathcal{M}(\delta) \mathbf{X}]_S\|_{1,1}$ by applying [Proposition 2](#). Let \mathbf{m}_i^\top be the i th row of the hollow matrix $\mathcal{M}(\delta)$, i.e., $\mathcal{M}(\delta) = [\mathbf{m}_1, \dots, \mathbf{m}_N]^\top \in \mathbb{R}^{N \times N}$. Likewise, let ω_i^\top be the i th row of Ω . Then from [Proposition 2](#) claims(a) and (b) in [\(A.3\)](#), we have $\|\mathbf{m}_i^\top \mathbf{X}\|_1 \geq (1 - \sigma_1) \beta_i P \|\mathbf{m}_i\|_2$ and $-\|\omega_i^\top \circ (\mathbf{m}_i^\top \mathbf{X})\|_1 \geq -(1 + \sigma_2) \beta_i \theta P \|\mathbf{m}_i\|_2$. So we obtain

$$\begin{aligned} \|\mathbf{m}_i^\top \mathbf{X}\|_1 - 2\|\omega_i^\top \circ (\mathbf{m}_i^\top \mathbf{X})\|_1 & \geq [(1 - \sigma_1) - 2\theta(1 + \sigma_2)] \beta_i P \|\mathbf{m}_i\|_2 \\ & \geq [(1 - \sigma_1) - 2\theta(1 + \sigma_2)] (1 - \sigma_4) \beta_0 P \|\mathbf{m}_i\|_2, \end{aligned}$$

where the last inequality holds because [Lemma 1](#) asserts that $\beta_i \in [(1 - \sigma'_i) \beta_0, \beta_0]$, for some $\sigma'_i \in (0, 1)$. Specifically, given a hollow-vector \mathbf{m}_i with $\bar{\alpha}_i = \|\mathbf{m}_i\|_\infty / \|\mathbf{m}_i\|_2$, the σ'_i can be computed via [\(A.4\)](#). Hence, we can let $\sigma_4 = \max\{\sigma'_i\}_{i=1}^N$ and by vertically stacking the row vectors \mathbf{m}_i^\top , $i \in \{1, \dots, N\}$, the bound

$$\begin{aligned} \|\mathcal{M}(\delta) \mathbf{X}\|_{1,1} - 2\|[\mathcal{M}(\delta) \mathbf{X}]_S\|_{1,1} & \geq [(1 - \sigma_1) - 2\theta(1 + \sigma_2)] (1 - \sigma_4) \beta_0 P \|\mathcal{M}(\delta)\|_{2,1} \end{aligned} \quad (\text{A.10})$$

holds with probability at least $1 - \delta$. Summing [\(A.9\)](#) and [\(A.10\)](#), we have

$$\begin{aligned} \|\mathcal{M}(\delta) \mathbf{X}\|_{1,1} - 2\|[\mathcal{M}(\delta) \mathbf{X}]_S\|_{1,1} + \sum_{i,p} |X_{ip}| \langle \mathbf{v}_i \circ \mathbf{v}_i, \delta \rangle & \geq [(1 - \sigma_1) - 2\theta(1 + \sigma_2)] (1 - \sigma_4) \left(\|\mathcal{M}(\delta)\|_{2,1} - C_1 |\mathbf{1}_N^\top \delta| \right), \end{aligned} \quad (\text{A.11})$$

where we defined $C_1 = \frac{(1 + \sigma_3) \sqrt{\theta}}{[(1 - \sigma_1) - 2\theta(1 + \sigma_2)] (1 - \sigma_4)}$. When $\theta \leq \frac{1 - \sigma_1}{2(1 + \sigma_2)}$, we have $C_1 > 0$ and the lower bound will be non-negative if $\|\mathcal{M}(\delta)\|_{2,1} \geq C_1 |\mathbf{1}_N^\top \delta|$. To show such a θ is feasible, recall that in [Proposition 2](#) we require that $\theta \in (0, e^{-1}]$, $\sigma_1 \in \left(0, \frac{\sqrt{\pi} \theta^{3/2}}{2} \right]$, and $\sigma_2 \in \left(0, \frac{\sqrt{\pi} \theta}{2} \right]$. Hence we need,

$$\theta \leq \frac{1 - (\sqrt{\pi}/2) \theta^{3/2}}{2(1 + \sqrt{\pi} \theta/2)} \Rightarrow f(\theta) := \sqrt{\pi} \theta^2 + 2\theta + \frac{\sqrt{\pi}}{2} \theta^{3/2} - 1 \leq 0.$$

As $f(\theta)$ is monotonically increasing in $\theta \in (0, e^{-1}]$ and $f(0) = -1 < 0$, $f(e^{-1}) > 0$, there exists only one solution $\theta_m \in (0.324, 0.325)$ such that $f(\theta_m) = 0$. So the feasible range is $\theta \in (0, \theta_m]$, or for convenience we let $\theta \in (0, 0.324]$.

Going back to establishing the lower bound in (A.11) is non-negative, note that $\|\mathcal{M}(\delta)\|_{2,1}^2 \geq \|\mathcal{M}(\delta)\|_F^2$ and so it will be sufficient to show $\|\mathcal{M}(\delta)\|_F^2 \geq C_1^2 |\mathbf{1}_N^\top \delta|^2$. For convenience, let us decompose $\bar{\mathbf{r}} = \frac{c}{N} \mathbf{1}_N + \mathbf{d}$ and $\delta = \delta_0 \mathbf{1}_N + \delta_\perp$, for some $\mathbf{d} := \mathbf{P}_1^\perp \bar{\mathbf{r}}$ and $\delta_\perp := \mathbf{P}_1^\perp \delta$ such that $\mathbf{1}_N^\top \delta_\perp = \mathbf{1}_N^\top \mathbf{d} = 0$. From the constraint $\bar{\mathbf{r}}^\top \mathbf{w} = \bar{\mathbf{r}}^\top (\mathbf{1}_N + \delta) = c$, we know $\bar{\mathbf{r}}^\top \delta = 0$ and $\delta_0 = -\frac{\mathbf{d}^\top \delta_\perp}{c}$. So $\delta = -\frac{\mathbf{d}^\top \delta_\perp}{c} \mathbf{1}_N + \delta_\perp$ and $\mathbf{1}_N^\top \delta = -\frac{N \mathbf{d}^\top \delta_\perp}{c}$. Then we have

$$\begin{aligned} \|\mathcal{M}(\delta)\|_F^2 - C_1^2 |\mathbf{1}_N^\top \delta|^2 &= \|\mathcal{M}(\delta_\perp)\|_F^2 - C_1^2 \left| \frac{N \mathbf{d}^\top \delta_\perp}{c} \right|^2 \\ &= \|\delta_\perp\|_2^2 \left(1 - \left\| \begin{bmatrix} \tilde{\mathbf{U}} \\ (N C_1 / c) \mathbf{d}^\top \end{bmatrix} \delta_\perp \right\|_2^2 \right) \\ (\star) &\geq \|\delta_\perp\|_2^2 \left[1 - \left(\sigma_{\max}^2(\tilde{\mathbf{U}}) + \frac{N^2 C_1^2}{c^2} \|\mathbf{d}\|_2^2 \right) \right], \quad (\text{A.12}) \end{aligned}$$

where we let $\tilde{\mathbf{U}} := (\mathbf{V} \circ \mathbf{V}) \mathbf{P}_1^\perp \in \mathbb{R}^{N \times N}$ and $\tilde{\delta}_\perp := \delta_\perp / \|\delta_\perp\|_2$. The inequality in (\star) considers the ‘worst-case’ \mathbf{d} should be colinear to the dominant right singular vector of $\tilde{\mathbf{U}}$, i.e., given the SVD $\tilde{\mathbf{U}} = \tilde{\mathbf{U}} \Sigma \tilde{\mathbf{V}}^\top$, then $\mathbf{d} / \|\mathbf{d}\|_2 = \hat{\mathbf{v}}_1$ is the first column of $\tilde{\mathbf{V}}$.

To ensure $1 - \left(\sigma_{\max}^2(\tilde{\mathbf{U}}) + \frac{N^2 C_1^2}{c^2} \|\mathbf{d}\|_2^2 \right) \geq 0$, we bound $\|\mathbf{d}\|_2$ as

$$\begin{aligned} \|\mathbf{d}\|_2^2 &\leq \frac{c^2(1 - \sigma_{\max}^2(\tilde{\mathbf{U}}))}{N^2 C_1^2} \\ &= \frac{c^2(1 - \sigma_{\max}^2(\tilde{\mathbf{U}}))[(1 - \sigma_1) - 2\theta(1 + \sigma_2)]^2(1 - \sigma_4)^2}{N^2(1 + \sigma_3)^2 \theta}, \quad (\text{A.13}) \end{aligned}$$

which is (A.2), completing the proof. ■

Appendix B. Proof of Theorem 2

When the observations are corrupted by noise $\mathbf{N} \in \mathbb{R}^{N \times P}$, i.e., $\mathbf{Y} = \mathcal{P}(\tilde{\mathbf{h}}_0) \mathbf{X} + \mathbf{N} \in \mathbb{R}^{N \times P}$, (10) can be rewritten as the following equivalent problem

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\mathcal{P}(\mathbf{w}) \mathbf{X} + \mathcal{P}(\tilde{\mathbf{g}}_0 \circ \mathbf{w}) \mathbf{N}\|_{1,1}, \quad \text{s. to } \bar{\mathbf{r}}^\top \mathbf{w} = c \quad (\text{B.1})$$

with the change of variable $\mathbf{w} = \tilde{\mathbf{g}}_0 \circ \tilde{\mathbf{h}}_0$ we used in Appendix A. Again, $\bar{\mathbf{r}}^\top = \mathbf{r}^\top \text{diag}(\tilde{\mathbf{g}}_0)$ is consistent with (A.1). In this case, the solution to (B.1) $\hat{\mathbf{w}}$ is not expected to be $\mathbf{1}_N$, so we let $\hat{\mathbf{w}} = \mathbf{1}_N + \hat{\delta} = \mathbf{1}_N + \left(-\frac{\mathbf{d}^\top \delta_\perp}{c} \mathbf{1}_N + \hat{\delta}_\perp\right) = \left(1 - \frac{\mathbf{d}^\top \delta_\perp}{c}\right) \mathbf{1}_N + \hat{\delta}_\perp$ as before, where $\mathbf{d} = \bar{\mathbf{r}} - \frac{c}{N} \mathbf{1}_N$. Then our goal will be to bound the error of problem (B.1), i.e., $\hat{\mathbf{w}} - \mathbf{1}_N$, which is expected to vanish if the observations were not corrupted by noise. Once this goal is achieved, the recovery error of (10), i.e., $\mathbf{e}_g := \hat{\mathbf{g}} - \tilde{\mathbf{g}}_0$ can also be bounded since $\mathbf{e}_g = \tilde{\mathbf{g}}_0 \circ (\hat{\mathbf{w}} - \mathbf{1}_N)$.

By jointly considering (A.8)(A.11)(A.12), and from (A.12) we can conclude that, if (A.13) is satisfied, e.g. $\|\mathbf{d}\|_2 \leq \frac{c \sqrt{(1 - \sigma_{\max}^2(\tilde{\mathbf{U}}))}}{N C_1} = (c/N) d_0$, the following holds for any feasible $\delta = -\frac{\mathbf{d}^\top \delta_\perp}{c} \mathbf{1}_N + \delta_\perp$

$$\begin{aligned} \|\mathcal{M}(\delta)\|_{2,1} - C_1 |\mathbf{1}_N^\top \delta| &\geq \sqrt{\|\mathcal{M}(\delta)\|_F^2} - C_1 |\mathbf{1}_N^\top \delta| \\ &\geq \sqrt{C_1^2 |\mathbf{1}_N^\top \delta|^2 + \|\delta_\perp\|_2^2 \left(1 - \sigma_{\max}^2(\tilde{\mathbf{U}}) - \frac{N^2 C_1^2}{c^2} \|\mathbf{d}\|_2^2 \right)} - C_1 |\mathbf{1}_N^\top \delta| \\ &= \|\delta_\perp\|_2 \left[\sqrt{1 - \sigma_{\max}^2(\tilde{\mathbf{U}}) - \frac{N^2 C_1^2}{c^2} \|\mathbf{d}\|_2^2 (1 - \sigma_5^2)} - \frac{N C_1}{c} \sigma_5 \|\mathbf{d}\|_2 \right], \quad (\text{B.2}) \end{aligned}$$

where $\sigma_5 := \frac{|\mathbf{d}^\top \delta_\perp|}{\|\delta_\perp\|_2 \|\mathbf{d}\|_2} \in [0, 1]$.

Let $Q := \frac{(1 + \sigma_3) \sqrt{\theta} N}{c} \left[\sqrt{(c/N)^2 d_0^2 - (1 - \sigma_5)^2 \|\mathbf{d}\|_2^2} - \sigma_5 \|\mathbf{d}\|_2 \right]$. Note that $Q \geq 0$ and then we have

$$\begin{aligned} \|\mathcal{P}(\hat{\mathbf{w}}) \mathbf{X} + \mathcal{P}(\tilde{\mathbf{g}}_0 \circ \hat{\mathbf{w}}) \mathbf{N}\|_{1,1} &= \|\mathcal{P}(\hat{\mathbf{w}}) [\mathbf{X} + \mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{N}]\|_{1,1} \\ &\geq \|\mathcal{P}(\hat{\mathbf{w}}) (\mathbf{X} + \mathbf{N}^{(S)})\|_{1,1} - \|\mathcal{P}(\hat{\mathbf{w}}) \mathbf{N}^{(C)}\|_{1,1} \\ &\geq \|\mathbf{X} + \mathbf{N}^{(S)}\|_{1,1} + \beta_0 P Q \|\hat{\delta}_\perp\|_2 - \|\mathcal{P}(\hat{\mathbf{w}}) \mathbf{N}^{(C)}\|_{1,1} \end{aligned} \quad (\text{B.3})$$

where $\mathbf{N}^{(S)} := [\mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{N}]_S$ denotes the sub-matrix of $\mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{N}$ that has the same support S as \mathbf{X} , and its complement as $\mathbf{N}^{(C)} := [\mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{N}]_{S^c} = \mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{N} - [\mathcal{P}(\tilde{\mathbf{g}}_0) \mathbf{N}]_S$. The last inequality holds as \mathbf{X} and $\mathbf{X} + \mathbf{N}^{(S)}$ have the same support. Next, let us find the upper bound for $\|\mathcal{P}(\hat{\mathbf{w}}) \mathbf{N}^{(C)}\|_{1,1}$

$$\begin{aligned} \|\mathcal{P}(\hat{\mathbf{w}}) \mathbf{N}^{(C)}\|_{1,1} &= \left\| \left(1 - \frac{\mathbf{d}^\top \hat{\delta}_\perp}{c} \right) \mathbf{N}^{(C)} + \mathbf{V} \text{diag}(\hat{\delta}_\perp) \mathbf{V}^\top \mathbf{N}^{(C)} \right\|_{1,1} \\ &\leq \|\mathbf{N}^{(C)}\|_{1,1} + (d_0 \|\mathbf{N}^{(C)}\|_{1,1} + \|\mathbf{N}^{(C)}\|^\top \mathbf{V} \circ \mathbf{V} \|\mathbf{N}^{(C)}\|) \|\hat{\delta}_\perp\|_2. \end{aligned} \quad (\text{B.4})$$

Note that we expect $\mathbf{w}_0 = \mathbf{1}_N$ as the ‘ideal’ noise-free solution of (B.1), so $c = \bar{\mathbf{r}}^\top \mathbf{1}_N$. For optimality, we should have

$$\begin{aligned} \|\mathcal{P}(\hat{\mathbf{w}}) \mathbf{X} + \mathcal{P}(\tilde{\mathbf{g}}_0 \circ \hat{\mathbf{w}}) \mathbf{N}\|_{1,1} &\leq \|\mathcal{P}(\mathbf{w}_0) \mathbf{X} + \mathcal{P}(\tilde{\mathbf{g}}_0 \circ \mathbf{w}_0) \mathbf{N}\|_{1,1} \\ &= \|\mathbf{X} + \mathbf{N}^{(S)}\|_{1,1} + \|\mathbf{N}^{(C)}\|_{1,1}. \end{aligned} \quad (\text{B.5})$$

From (B.3), (B.4) and (B.5) we find

$$\begin{aligned} \|\mathbf{X} + \mathbf{N}^{(S)}\|_{1,1} + \|\mathbf{N}^{(C)}\|_{1,1} &\geq \|\mathbf{X} + \mathbf{N}^{(S)}\|_{1,1} - \|\mathbf{N}^{(C)}\|_{1,1} \\ &\quad + (\beta_0 P Q - d_0 \|\mathbf{N}^{(C)}\|_{1,1} - \|\mathbf{N}^{(C)}\|^\top \mathbf{V} \circ \mathbf{V} \|\mathbf{N}^{(C)}\|) \|\hat{\delta}_\perp\|_2 \end{aligned}$$

and while $\beta_0 P Q - d_0 \|\mathbf{N}^{(C)}\|_{1,1} - \|\mathbf{N}^{(C)}\|^\top \mathbf{V} \circ \mathbf{V} \|\mathbf{N}^{(C)}\| > 0$, we have

$$\|\hat{\delta}_\perp\|_2 \leq \frac{2 \|\mathbf{N}^{(C)}\|_{1,1}}{\beta_0 P Q - d_0 \|\mathbf{N}^{(C)}\|_{1,1} - \|\mathbf{N}^{(C)}\|^\top \mathbf{V} \circ \mathbf{V} \|\mathbf{N}^{(C)}\|}.$$

Because $\hat{\mathbf{g}} = \tilde{\mathbf{g}}_0 \circ \hat{\mathbf{w}}$, the error vector between $\hat{\mathbf{g}}$ and the ‘ideal’ ground-truth $\tilde{\mathbf{g}}_0$ is $\mathbf{e}_g = \hat{\mathbf{g}} - \tilde{\mathbf{g}}_0 = \tilde{\mathbf{g}}_0 \circ (\hat{\mathbf{w}} - \mathbf{1}_N) = \tilde{\mathbf{g}}_0 \circ \hat{\delta}$. Recalling that $\hat{\delta} = -\frac{\mathbf{r}^\top \hat{\delta}_\perp}{c} \mathbf{1}_N + \hat{\delta}_\perp = \left(\mathbf{1}_N - \mathbf{1}_N \frac{\bar{\mathbf{r}}^\top}{c} \right) \hat{\delta}_\perp$ and $\beta_0 = \sqrt{\frac{2}{\pi}}$, we can bound the ℓ_1 or ℓ_2 norms of \mathbf{e}_g as

$$\begin{aligned} \|\mathbf{e}_g\|_l &= \|\tilde{\mathbf{g}}_0 \circ \hat{\delta}\|_l \\ &= \left\| \text{diag}(\tilde{\mathbf{g}}_0) \left(\mathbf{1}_N - \mathbf{1}_N \frac{\bar{\mathbf{r}}^\top}{c} \right) \hat{\delta}_\perp \right\|_l \\ &\leq \left\| \text{diag}(\tilde{\mathbf{g}}_0) \left(\mathbf{1}_N - \mathbf{1}_N \frac{\bar{\mathbf{r}}^\top}{c} \right) \right\|_{l \rightarrow 2} \|\hat{\delta}_\perp\|_2 \\ &\leq \frac{2 \left\| \text{diag}(\tilde{\mathbf{g}}_0) \left(\mathbf{1}_N - \mathbf{1}_N \frac{(\mathbf{r} \circ \tilde{\mathbf{g}}_0)^\top}{c} \right) \right\|_{l \rightarrow 2} \|\mathbf{N}^{(C)}\|_{1,1}}{\sqrt{\frac{2}{\pi}} P Q - d_0 \|\mathbf{N}^{(C)}\|_{1,1} - \|\mathbf{N}^{(C)}\|^\top \mathbf{V} \circ \mathbf{V} \|\mathbf{N}^{(C)}\|}, \end{aligned}$$

where $\|\cdot\|_l$ stands for the ℓ_1 and ℓ_2 norms when $l = 1, 2$, respectively. ■

Appendix C. Epinions data sampling and pre-processing

C.1. Sampling

Here we describe the implemented sampling design applied to \mathbf{Y}_{obs} . The goals of data sampling are: (i) the resulting observation density should be as high as possible; and (ii) all of the users in the sampled dataset should be connected. To achieve those goals, for $k = 1, 2, \dots$ we repeat the following three steps. *Step 1:* From the previous item set \mathcal{I}_k we pick all items that have been rated by at least $N_{\min} = 150$ users from \mathcal{U}_k , and let the new item set be \mathcal{I}_{k+1} . *Step 2:* From the previous user set \mathcal{U}_k we pick all users who have rated at least N_{\min} items in \mathcal{I}_{k+1} and let the new user set be the user set candidate \mathcal{U}'_{k+1} . *Step 3:* To maintain the connectivity of the user network, we randomly select a user from \mathcal{U}'_{k+1} and collect all of the users (within \mathcal{U}'_{k+1}) that are accessible from this user (including the selected user) to generate the new user set \mathcal{U}_{k+1} . The above three steps are repeated until there is no feasible update for $\{\mathcal{I}_k, \mathcal{U}_k\}$.

C.2. Pre-processing

As hinted by the results in [Theorem 1](#), the unbiasedness of the graph signals is crucial to satisfactory performance of the proposed approach. Hence we centered the rating matrix by adjusting the range from $[1, 5]$ to $[-2, 2]$.

Data availability

Data will be made available on request.

References

- [1] W. Huang, L. Goldsberry, N.F. Wymbs, S.T. Grafton, D.S. Bassett, A. Ribeiro, Graph frequency analysis of brain signals, *IEEE J. Sel. Top. Signal Process.* 10 (7) (2016) 1189–1203.
- [2] C. Hu, X. Hua, J. Ying, P.M. Thompson, G.E. Fakhri, Q. Li, Localizing sources of brain disease progression with network diffusion model, *IEEE J. Sel. Top. Signal Process.* 10 (7) (2016) 1214–1225.
- [3] Y. Li, G. Mateos, Z. Zhang, Learning to model the relationship between brain structural and functional connectomes, *IEEE Trans. Signal Inf. Process. Netw.* 8 (2022) 830–843.
- [4] J.A. Deri, J.M.F. Moura, New York city taxi analysis with graph signal processing, in: *Proc. IEEE Global Conf. on Signal and Information Process.*, 2016, pp. 1275–1279.
- [5] A. Hasanzadeh, X. Liu, N. Duffield, K.R. Narayanan, Piecewise stationary modeling of random processes over graphs with an application to traffic prediction, in: *Proc. IEEE Int. Conf. on Big Data*, 2019, pp. 3779–3788.
- [6] Y. Li, G. Mateos, Graph frequency analysis of COVID-19 incidence to identify county-level contagion patterns in the United States, in: *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, 2021, pp. 3230–3234.
- [7] I.D. Schizas, G. Mateos, G.B. Giannakis, Distributed LMS for consensus-based in-network adaptive processing, *IEEE Trans. Signal Process.* 57 (6) (2009) 2365–2381.
- [8] L. Stankovic, D.P. Mandic, M. Dakovic, I. Kisil, E. Sejdic, A.G. Constantinides, Understanding the basis of graph signal processing via an intuitive example-driven approach, *IEEE Signal Process. Mag.* 36 (6) (2019) 133–145.
- [9] A. Ortega, P. Frossard, J. Kovačević, J.M.F. Moura, P. Vandergheynst, Graph signal processing: Overview, challenges, and applications, *Proc. IEEE* 106 (5) (2018) 808–828.
- [10] G. Leus, A.G. Marques, J.M. Moura, A. Ortega, D.I. Shuman, Graph signal processing: History, development, impact, and outlook, 40 (4) (2023) 49–60.
- [11] A. Sandryhaila, J.M. Moura, Discrete signal processing on graphs, *IEEE Trans. Signal Process.* 61 (7) (2013) 1644–1656.
- [12] E. Isufi, F. Gama, D.I. Shuman, S. Segarra, Graph filters for signal processing and machine learning on graphs, *IEEE Trans. Signal Process.* (2024) 1–32.
- [13] F. Gama, E. Isufi, G. Leus, A. Ribeiro, Graphs, convolutions, and neural networks: From graph filters to graph neural networks, *IEEE Signal Process. Mag.* 37 (6) (2020) 128–138.
- [14] A. Sandryhaila, J.M. Moura, Discrete signal processing on graphs: Frequency analysis., *IEEE Trans. Signal Process.* 62 (12) (2014) 3042–3054.
- [15] G. Mateos, S. Segarra, A.G. Marques, A. Ribeiro, Connecting the dots: Identifying network structure via graph signal processing, *IEEE Signal Process. Mag.* 36 (3) (2019) 16–43.
- [16] S. Segarra, G. Mateos, A.G. Marques, A. Ribeiro, Blind identification of graph filters with sparse inputs, in: *CAMSAP*, 2015, pp. 449–452.
- [17] S. Segarra, G. Mateos, A.G. Marques, A. Ribeiro, Blind identification of graph filters, *IEEE Trans. Signal Process.* 65 (5) (2017) 1146–1159.
- [18] A. Ahmed, B. Recht, J. Romberg, Blind deconvolution using convex programming, *IEEE Trans. Inform. Theory* 60 (3) (2014) 1711–1732.
- [19] A. Levin, Y. Weiss, F. Durand, W.T. Freeman, Understanding blind deconvolution algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (12) (2011) 2354–2367.
- [20] L. Wang, Y. Chi, Blind deconvolution from multiple sparse inputs, *IEEE Signal Process. Lett.* 23 (10) (2016) 1384–1388.
- [21] D. Ramírez, A.G. Marques, S. Segarra, Graph-signal reconstruction and blind deconvolution for structured inputs, *Signal Process.* 188 (2021) 108180.
- [22] C. Ye, R. Shafipour, G. Mateos, Blind identification of invertible graph filters with sparse inputs, in: *Proc. of European Signal Process. Conf.*, 2018, pp. 121–125.
- [23] C. Ye, G. Mateos, Learning to identify sources of network diffusion, in: *Proc. of European Signal Process. Conf.*, 2022, pp. 727–731.
- [24] V.M. Tenorio, S. Rey, A.G. Marques, Blind deconvolution of sparse graph signals in the presence of perturbations, in: *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, 2024, pp. 9406–9410.
- [25] P. Zhang, J. He, G. Long, G. Huang, C. Zhang, Towards anomalous diffusion sources detection in a large network, *ACM T. Internet Techn.* 16 (1) (2016) 2.
- [26] P.C. Pinto, P. Thiran, M. Vetterli, Locating the source of diffusion in large-scale networks, *Phys. Rev. Lett.* 109 (6) (2012) 068702.
- [27] E. Sefer, C. Kingsford, Diffusion archeology for diffusion progression history reconstruction, *Knowl. Inf. Syst.* 49 (2) (2016) 403–427.
- [28] R. Pena, X. Bresson, P. Vandergheynst, Source localization on graphs via ℓ_1 recovery and spectral graph theory, in: *Proc. IEEE Image, Video, and Multidimensional Signal Process. Workshop*, 2016, pp. 1–5.
- [29] S. Feizi, M. Médard, G. Quon, M. Kellis, K. Duffy, Network infusion to infer information sources in networks, *IEEE Trans. Netw. Sci. Eng.* 6 (3) (2019) 402–417.
- [30] S. Ling, T. Strohmer, Self-calibration and biconvex compressive sensing, *Inverse Problems* 31 (11) (2015) 115002.
- [31] Y. Li, K. Lee, Y. Bresler, Identifiability in bilinear inverse problems with applications to subspace or sparsity-constrained blind gain and phase calibration, *IEEE Trans. Inform. Theory* 63 (2) (2017) 822–842.
- [32] A.G. Marques, S. Segarra, G. Mateos, Signal processing on directed graphs: The role of edge directionality when processing and learning from network data, *IEEE Signal Process. Mag.* 37 (6) (2020) 99–116.
- [33] M. Puschel, J.M.F. Moura, Algebraic signal processing theory: Foundation and 1-D time, *IEEE Trans. Signal Process.* 56 (8) (2008) 3572–3585.
- [34] A. Gavili, X.-P. Zhang, On the shift operator, graph frequency, and optimal filtering in graph signal processing, *IEEE Trans. Signal Process.* 65 (23) (2017) 6303–6318.
- [35] M.H. DeGroot, Reaching a consensus, *J. Amer. Statist. Assoc.* 69 (1974) 118–121.
- [36] H.-T. Wai, S. Segarra, A.E. Ozdaglar, A. Scaglione, A. Jadbabaie, Blind community detection from low-rank excitations of a graph filter, *IEEE Trans. Signal Process.* 68 (2020) 436–451.
- [37] R.A. Horn, C.R. Johnson, *Matrix Analysis*, second ed., Cambridge University Press, 2012.
- [38] H. Zhang, M. Yan, W. Yin, One condition for solution uniqueness and robustness of both ℓ_1 -synthesis and ℓ_1 -analysis minimizations, *Adv. Comput. Math.* 42 (6) (2016) 1381–1399.
- [39] D. Ramírez, A.G. Marques, S. Segarra, Graph-signal reconstruction and blind deconvolution for diffused sparse inputs, in: *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, 2017, pp. 4104–4108.
- [40] E.J. Candes, M.B. Wakin, S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, *J. Fourier Anal. Appl.* 14 (5–6) (2008) 877–905.
- [41] M. Grant, S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, 2014, <http://cvxr.com/cvx>.
- [42] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C.J. Honey, V.J. Wedeen, O. Sporns, Mapping the structural core of human cerebral cortex, *PLoS Biol.* 6 (7) (2008) e159.
- [43] R.M. Hamedani, I. Ali, J. Hong, S.-W. Kim, TrustRec: An effective approach to exploit implicit trust and distrust relationships along with explicitities for accurate recommendations, *Comput. Sci. Inf. Syst.* 18 (1) (2021) 93–114.
- [44] J. Matoušek, On variants of the Johnson–Lindenstrauss lemma, *Random Structures Algorithms* 33 (2) (2008) 142–156.