BLIND IDENTIFICATION OF GRAPH FILTERS WITH MULTIPLE SPARSE INPUTS

Santiago Segarra[†]*, Antonio G. Marques*[*]*, Gonzalo Mateos*[‡] *and Alejandro Ribeiro*[†]

[†]Dept. of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA
[*]Dept. of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain
[‡]Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA

ABSTRACT

Network processes are often represented as signals defined on the vertices of a graph. To untangle the latent structure of such signals, one can view them as outputs of *linear graph filters* modeling underlying network dynamics. This paper deals with the problem of joint identification of a graph filter and its input signal, thus broadening the scope of classical blind deconvolution of temporal and spatial signals to the less-structured graph domain. Given a graph signal $\mathbf{y}$ modeled as the output of a graph filter, the goal is to recover the vector of filter coefficients $\mathbf{h}$, and the input signal $\mathbf{x}$ which is assumed to be sparse. While $\mathbf{y}$ is a bilinear function of $\mathbf{x}$ and $\mathbf{h}$, the filtered graph signal is also a linear combination of the entries of the "lifted" rank-one, row-sparse matrix $\mathbf{x}\mathbf{h}^T$. The blind graph filter identification problem can be thus tackled via rank and sparsity minimization subject to linear constraints, an approach amenable to convex relaxation. An algorithm for jointly processing multiple output signals corresponding to different sparse inputs is also developed. Numerical tests with synthetic and real-world networks illustrate the merits of the proposed algorithm, as well as the benefits of leveraging multiple signals to aid the blind identification task.

***Index Terms*—** Graph signal processing, blind system identification, graph filter, graph process, multiple sparse signals.

1. INTRODUCTION

Coping with the challenges posed by fields such as network science and big data necessitates broadening the scope beyond classical time-varying signal analysis and processing, to also accommodate *signals defined on graphs* [1–3]. Under the assumption that the signal properties are related to the topology of the graph where they are supported, the goal of graph signal processing is to develop algorithms that fruitfully leverage this relational structure. A suitable way to achieve this is to rely on the so-called graph-shift operator, which is a matrix that reflects the local connectivity of the graph [2].

We consider here that each node has a certain value, and these values are collected across nodes to form a graph signal. With this definition, graph filters – which are a generalization of classical time-invariant systems – are a specific class of operators whose input and output are graph signals (cf. Section 2). Mathematically, graph filters are linear transformations that can be expressed as polynomials of the graph-shift operator [4]. The polynomial coefficients determine completely the transformation and are referred to as *filter coefficients*. Such linear transformations can be implemented via local interactions among nodes, and may be used to model e.g., diffusion or percolation dynamics in the network [5, 6].

Contributions. This paper investigates the problem of blind identification of graph filters. Specifically, we are given a graph signal $\mathbf{y}$ which is assumed to be modeled as the output of a graph filter, and seek to *jointly identify* the *filter* coefficients $\mathbf{h}$ and the *input signal* $\mathbf{x}$ that gave rise to $\mathbf{y}$. Such a challenging problem broadens to graphs the scope of classical blind system identification or blind deconvolution of signals in the time or spatial domains [7]. Since the inverse problem is ill-posed, we assume that the length of $\mathbf{h}$ is small and that $\mathbf{x}$ is sparse. This is the case when, e.g., a few seeding nodes inject a signal that is diffused throughout a network [8,9]. While $\mathbf{y}$ is a bilinear function of $\mathbf{x}$ and $\mathbf{h}$, in Section 3 we show the filtered graph signal is also a linear combination of the entries of the "lifted" rank-one, row-sparse matrix $\mathbf{x}\mathbf{h}^T$ [7, 10]. The blind graph filter identification problem can be thus tackled via joint rank and sparsity minimization subject to linear constraints, an approach amenable to convex relaxation [11, 12]. An algorithm is developed in Section 4 to perform blind identification when multiple outputs are available; see also [13] for identifiability claims in a setting unrelated to graphs. Simulations in Section 5 showcase the effectiveness of the proposed algorithm on synthetic and real-world graphs, and the benefits of leveraging multiple signals to aid the blind identification task.

Relation to prior work. Our ideas were inspired by the pioneering work in [7], where matrix lifting is used for blind deconvolution of temporal and spatial signals. In the current paper, the linear operator mapping $\mathbf{x}\mathbf{h}^T$ to the output signal $\mathbf{y}$ depends on the spectral properties of the graph-shift operator [14], a departure from the random (Gaussian or partial Fourier) operators arising with the biconvex compressed sensing approach of [10]. Unlike our previous work in [14], we consider that the support of the input signal $\mathbf{x}$ is unknown – a pragmatic setting useful to trace those influential nodes that led to the observed network state. Accordingly, envisioned applications domains include opinion formation in social networks (who started the rumor?), inverse problems of biological signals supported on graphs (which brain regions were activated?), and modeling and estimation of diffusion processes (who is "patient zero" for the disease outbreak?) [3]. Furthermore, the setup where multiple output signals are observed (each one corresponding to a different sparse input), has not been explored in recent convex relaxation approaches to blind deconvolution [7].

2. GRAPH SIGNALS AND GRAPH FILTERS

Graph signals. Let $\mathcal{G}$ denote a directed graph with a set of nodes $\mathcal{N}$ (with cardinality N) and a set of links $\mathcal{E}$, such that if node i is connected to j, then $(i,j) \in \mathcal{E}$. Since $\mathcal{G}$ is directed, the set $\mathcal{N}_i : \{j \ \ |(j,i) \in \mathcal{E}\}$ stands for the (incoming) neighborhood of i. For any given $\mathcal{G}$ we define the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ as a sparse matrix with non-zero elements A_{ji} if and only if $(i,j) \in \mathcal{E}$. The value of A_{ji} captures the strength of the connection from i to j. The focus of the paper is on analyzing (graph) signals defined on $\mathcal{N}$. These signals can be represented as vectors $\mathbf{x} = [x_1, ..., x_N]^T \in \mathbb{R}^N$, where x_i represents the value of the signal at node i.

Work in this paper is supported by USA NSF CCF-1217963 and Spanish MINECO TEC2013-41604-R.

Graph shift operator. The graph $\mathcal{G}$ can be endowed with the so-called *graph-shift operator* $\mathbf{S}$ [2,4]. The shift $\mathbf{S} \in \mathbb{R}^{N\times N}$ is a matrix whose entry S_{ji} can be non-zero only if $i = j$ or if $(i, j) \in \mathcal{E}$. The sparsity pattern of the matrix $\mathbf{S}$ captures the local structure of $\mathcal{G}$, but we make no specific assumptions on the values of its non-zero entries The intuition behind $\mathbf{S}$ is to represent a linear transformation that can be computed locally at the nodes of the graph. More rigorously, if $\mathbf{y}$ is defined as $\mathbf{y} = \mathbf{S}\mathbf{x}$, then node i can compute y_i provided that it has access to the value of x_j at $j \in \mathcal{N}_i$. Typical choices for $\mathbf{S}$ are the adjacency matrix $\mathbf{A}$ [2, 4], and the graph Laplacian [1]. We assume henceforth that $\mathbf{S}$ is diagonalizable, so that $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ with $\mathbf{\Lambda} \in \mathbb{R}^{N\times N}$ being diagonal.

Graph filters. The shift $\mathbf{S}$ can be used to define linear graph-signal *operators* of the form

$$\mathbf{H} := \textstyle\sum_{l=0}^{L-1} h_l \mathbf{S}^l \tag{1}$$

which are called *graph filters* [2]. For a given input $\mathbf{x}$, the output of the filter is simply $\mathbf{y} = \mathbf{H}\mathbf{x}$. The coefficients of the filter are collected into $\mathbf{h} := [h_0, \ldots, h_{L-1}]^T$, with $L - 1$ denoting the filter degree. Graph filters are of particular interest because they represent linear transformations that can be implemented locally [5, 9].

Frequency domain representation. Leveraging the spectral decomposition of $\mathbf{S}$, graph filters and signals can be represented in the frequency domain. To be precise, let us use the eigenvectors of $\mathbf{S}$ to define the $N \times N$ matrix $\mathbf{U} := \mathbf{V}^{-1}$, and the eigenvalues of $\mathbf{S}$ to define the $N \times L$ Vandermonde matrix $\mathbf{\Psi}$, where $\Psi_{ij} := (\Lambda_{ii})^{j-1}$. Using these conventions, the frequency representations of a *signal* $\mathbf{x}$ and of a *filter* $\mathbf{h}$ are defined as $\widehat{\mathbf{x}} := \mathbf{U}\mathbf{x}$ and $\widehat{\mathbf{h}} := \mathbf{\Psi}\mathbf{h}$, respectively [4]. Exploiting such representations, the output $\mathbf{y} = \mathbf{H}\mathbf{x}$ of a graph filter in the frequency domain is given by

$$\widehat{\mathbf{y}} = \mathrm{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x} = \mathrm{diag}\big(\widehat{\mathbf{h}}\big)\widehat{\mathbf{x}}. \tag{2}$$

Identity (2) is the counterpart of the celebrated convolution theorem for temporal signals, and follows from $\mathbf{H} = \mathbf{V}\big(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l\big)\mathbf{U}$ [cf. (1)] and $\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l = \mathrm{diag}(\mathbf{\Psi}\mathbf{h})$; see e.g., [14] for a detailed derivation. To establish further connections with the time domain, let us consider the directed cycle graph whose adjacency matrix $\mathbf{A}_{dc}$ is zero, except for entries $A_{ij} = 1$ whenever $i = \mathrm{mod}_N(j)+1$, where $\mathrm{mod}_N(x)$ denotes the modulus (remainder) obtained after dividing x by N. If $\mathbf{S} = \mathbf{A}_{dc}$, one can verify that: i) $\mathbf{y} = \mathbf{H}\mathbf{x}$ can be found as the circular convolution of $\mathbf{h}$ and $\mathbf{x}$, and ii) both $\mathbf{U}$ and $\mathbf{\Psi}$ correspond to the Discrete Fourier Transform (DFT) matrix. Interestingly, while in the time domain $\mathbf{U} = \mathbf{\Psi}$, this is not true for general graphs.

3. BLIND IDENTIFICATION OF GRAPH FILTERS

The concepts introduced in the previous section can be used to formally state the problem. For given shift operator $\mathbf{S}$ and filter degree $L - 1$, suppose that we observe the output signal $\mathbf{y} = \mathbf{H}\mathbf{x}$ [cf. (1)], where $\mathbf{x}$ is *sparse* having at most $S \ll N$ non-zero entries. For future reference introduce the ℓ_0 (pseudo) norm $\|\mathbf{x}\|_0 := |\mathrm{supp}(\mathbf{x})|$, where the support of $\mathbf{x}$ is $\mathrm{supp}(\mathbf{x}) := \{i \,|\, x_i \neq 0\}$ and hence $\|\mathbf{x}\|_0 \leq S$. The present paper deals with *blind identification* of the graph filter (and its input signal), which amounts to estimating sparse $\mathbf{x}$ and the filter coefficients $\mathbf{h}$ from the observed output signal $\mathbf{y}$. Such a challenging problem is a natural extension to graphs of classical blind system identification, or blind deconvolution of signals in the temporal or spatial domains.

Remark 1 (Sparse input) Sparsity in $\mathbf{x}$ is well-motivated due to its practical relevance and modeling value – network signals such as $\mathbf{y}$ are oftentimes the diffused version of few localized sources, hereby indexed by $\mathrm{supp}(\mathbf{x})$. In addition, the non-sparse formulation with $S = N$ is ill-posed, since the number of unknowns $N + L$ in $\{\mathbf{x}, \mathbf{h}\}$ exceeds the number of observations N in $\mathbf{y}$. Alternatively, a low-dimensional subspace model for $\mathbf{x}$ could be also adopted to effectively reduce the degrees of freedom in the problem [14].

Since the observed filter output is $\mathbf{y} = \mathbf{V}\widehat{\mathbf{y}} = \mathbf{V}\mathrm{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x}$ [cf. (2)], the blind graph filter identification problem can be stated as the following feasibility problem

$$\begin{aligned} &\text{find} \;\; \{\mathbf{h}, \mathbf{x}\} \\ &\text{s. to} \;\; \mathbf{y} = \mathbf{V}\mathrm{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x}, \;\; \|\mathbf{x}\|_0 \leq S. \end{aligned} \tag{3}$$

In other words, the goal is to find the solution to a set of bilinear equations subject to a sparsity constraint in $\mathbf{x}$. While very natural, (3) is in fact a difficult problem due to the non-convex ℓ_0-norm as well as the bilinear constraints. To deal with the latter, it is convenient to rewrite the first constraint in (3) as

$$\mathbf{y} = \mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \mathrm{vec}\big(\mathbf{x}\mathbf{h}^T\big), \tag{4}$$

where $\odot$ denotes the Khatri-Rao (i.e., columnwise Kronecker) product, and $\mathrm{vec}(\cdot)$ is the matrix vectorization operator. To establish (4), let $\mathbf{u}_i^T$ and $\boldsymbol{\psi}_i^T$ denote the i-th rows of $\mathbf{U}$ and $\mathbf{\Psi}$, respectively. It follows from (2) that $\widehat{y}_i = (\boldsymbol{\psi}_i^T\mathbf{h})(\mathbf{u}_i^T\mathbf{x}) = \big(\boldsymbol{\psi}_i^T \otimes \mathbf{u}_i^T\big)\mathrm{vec}\big(\mathbf{x}\mathbf{h}^T\big)$, where $\otimes$ denotes the Kronecker product. Upon stacking the entries $\widehat{y}_i$ to form $\widehat{\mathbf{y}}$ and then using $\mathbf{y} = \mathbf{V}\widehat{\mathbf{y}}$, the result follows by identifying $\boldsymbol{\psi}_i^T \otimes \mathbf{u}_i^T$ with the i-th row of $\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T$.

While (4) confirms that the filtered graph signal $\mathbf{y}$ is a *bilinear* function of $\mathbf{x}$ and $\mathbf{h}$, it also shows that $\mathbf{y}$ is a *linear* combination of the entries of the "lifted" rank-one, outer-product matrix $\mathbf{Z} := \mathbf{x}\mathbf{h}^T \in \mathbb{R}^{N\times L}$. In other words, there exists a linear mapping $\mathcal{M} : \mathbb{R}^{N\times L} \mapsto \mathbb{R}^N$ such that $\mathbf{y} = \mathcal{M}(\mathbf{Z})$. Note that $\mathcal{M}$ can be expressed in terms of a matrix multiplication with $\mathbf{M} := \mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \in \mathbb{R}^{N\times LN}$, since $\mathbf{y} = \mathbf{M}\mathrm{vec}(\mathbf{Z})$ as per (4). In addition to being of rank one, note that the sparsity in $\mathbf{x}$ renders $\mathbf{Z}$ row-wise sparse, i.e., rows $\mathbf{z}_i^T$ indexed by $\{1, \ldots, N\} \setminus \mathrm{supp}(\mathbf{x})$ are identically zero. Building on the ideas in [7, 10], one can thus pose the blind graph filter identification problem as a *linear* inverse problem, where the goal is to recover a row-sparse, rank-one $N \times L$ matrix $\mathbf{Z}$ from observations $\mathbf{y} = \mathcal{M}(\mathbf{Z})$. To this end, a natural formulation to tackle such inverse problem is

$$\begin{aligned} &\min_{\mathbf{Z}} \;\; \mathrm{rank}(\mathbf{Z}) \\ &\text{s. to} \;\; \mathbf{y} = \mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \mathrm{vec}\big(\mathbf{Z}\big), \;\; \|\mathbf{Z}\|_{2,0} \leq S \end{aligned} \tag{5}$$

where $\|\mathbf{Z}\|_{2,0}$ is equal to the number of non-zero rows of $\mathbf{Z}$.

A basic question is whether (5) is equivalent to the original blind identification problem. To give a rigorous answer, some definitions are introduced next. For a given matrix $\mathbf{U}$, $\mathrm{spark}(\mathbf{U})$ is the smallest number n such that there exists a subgroup of n columns from $\mathbf{U}$ that are linearly dependent [15]. Given a set of row indices $\mathcal{I}$, define the complement set of indices $\mathcal{I}^c := \{1, \ldots, N\}\backslash\mathcal{I}$ and the matrix $\mathbf{U}_{\mathcal{I}}$ formed by the rows of $\mathbf{U}$ indexed by $\mathcal{I}$. Moreover, for a given graph-shift operator $\mathbf{S}$ – fixed $\mathbf{V}$, $\mathbf{\Psi}$, and $\mathbf{U}$ – define the set $\mathcal{O}_{\mathbf{y}}$ of matrix minimizers of (5) as a function of $\mathbf{y}$. Then, the following result on the validity of the matrix problem formulation in (5) holds.

Proposition 1 *Let $|\cdot|$ denote the number of non-repeated elements of a set and $\mathcal{I}_S$ be a set of row indices such that* $\mathrm{spark}(\mathbf{U}_{\mathcal{I}_S}) \leq S$. *Then*

$$\mathcal{O}_y = \big\{\mathbf{x}\mathbf{h}^T \,\big|\, \mathbf{y} = \textstyle\sum_{l=0}^{L-1} h_l \mathbf{S}^l \mathbf{x},\ \ \|\mathbf{x}\|_0 \leq S\big\}, \tag{6}$$

for all $\mathbf{y}$ *if and only if*

$$\min_{\mathcal{I}_S} \big|\{\lambda_i\}_{i\in\mathcal{I}_S^c}\big| > L-1. \tag{7}$$

Proof: If we show that (7) is violated if and only if there exists a rank-1 matrix $\mathbf{Z} = \mathbf{x}\mathbf{h}^T$ such that $\mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \text{vec}\big(\mathbf{Z}\big) = \mathbf{0}$ and $\|\mathbf{Z}\|_{2,0} \leq S$, then Corollary 1 in [16] completes the proof. Leveraging the fact that $\mathbf{V}$ is full-rank, the above system of homogenous equations can be written as $(\boldsymbol{\psi}_i^T\mathbf{h})(\mathbf{u}_i^T\mathbf{x}) = 0$ for $i = 1,\ldots,N$, where $\boldsymbol{\psi}_i^T$ denotes the i-th row of $\mathbf{\Psi}$ and similarly for $\mathbf{U}$. Since $\text{spark}(\mathbf{U}_{\mathcal{I}_S^c}) \leq S$, there exists $\mathbf{x} \neq \mathbf{0}$ with $\|\mathbf{x}\|_0 \leq S$ such that $(\boldsymbol{\psi}_i^T\mathbf{h})(\mathbf{u}_i^T\mathbf{x}) = 0$ holds for $i \in \mathcal{I}_S$. Exploiting the Vandermonde structure of $\mathbf{\Psi}$, it follows that $\mathbf{h} \neq \mathbf{0}$ satisfying the equality for $i \in \mathcal{I}_S^c$ can be found if and only if (7) is violated. ■

Ideally, when solving (5) for some output $\mathbf{y}$ one should recover the set of outer products of all possible combinations of sparse inputs $\mathbf{x}$ and filter coefficients $\mathbf{h}$ that can give rise to such output [cf. (6)]. This is not true in general [16, Theorem 1], however, Proposition 1 states conditions on the graph-shift operator [cf. (7)] for the desired equivalence to hold. Notice that condition (7) does not guarantee that the solution of (5) is unique, but rather that the outer product of the desired sparse signal and filter coefficients is contained in $\mathcal{O}_\mathbf{y}$. Conditions that guarantee identifiability of (5), i.e. uniqueness of solution, are subject of ongoing investigation.

3.1. Algorithmic approach via convex relaxation

Albeit natural, problem (5) is challenging since both the rank and the ℓ_0-norm are in general NP-hard to optimize; see e.g., [17]. Over the last decade or so, convex relaxation approaches to tackle rank and/or sparsity minimization problems have enjoyed remarkable success, since they oftentimes entail no loss in optimality. The nuclear norm $\|\mathbf{Z}\|_* = \sum_k \sigma_k(\mathbf{Z})$, where $\sigma_k(\mathbf{Z})$ denotes the k-th singular value of $\mathbf{Z}$, is typically adopted as a convex surrogate to $\text{rank}(\mathbf{Z})$ [11, 17]. Likewise, the ℓ_2/ℓ_1 mixed norm $\|\mathbf{Z}\|_{2,1} := \sum_{i=1}^N \|\mathbf{z}_i^T\|_2$ is the closest convex approximation of $\|\mathbf{Z}\|_{2,0}$ [18]. With τ denoting a tuning parameter to control the rank versus row-sparsity tradeoff, a convex heuristic is to solve

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* + \tau\|\mathbf{Z}\|_{2,1}, \quad \text{s. to } \ \mathbf{y} = \mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \text{vec}\big(\mathbf{Z}\big) \tag{8}$$

hoping that the optimal solution is of rank one and has S non-zero rows, so that we can recover $\mathbf{x}$ and $\mathbf{h}$ up to scaling.

Recovery of simultaneously low-rank and row-sparse matrices from noisy compressive measurements was also considered in [19] for hyperspectral image reconstruction. Recent theoretical results on recovery of simultaneously structured matrix models suggest that minimizing only $\|\mathbf{Z}\|_{2,1}$ could as well suffice [12]; see also [10]. Being convex, (8) is computationally appealing, in fact off-the-shelf interior point solvers are available. Customized scalable algorithms for large-scale problems can be developed to minimize the composite, non-differentiable cost in (8). For instance the solver implemented to run the numerical tests in Section 5 leverages the alternating-direction method of multipliers (ADMM) [20]; see also [19] for a related proximal-splitting algorithm.

Refinement via iteratively-reweighted optimization. Instead of substituting $\|\mathbf{Z}\|_{2,0}$ in (5) by its closest convex approximation, namely $\|\mathbf{Z}\|_{2,1}$, letting the surrogate function to be non-convex can yield tighter approximations, and potentially improve the statistical properties of the estimator. In the context of sparse signal recovery for instance, the ℓ_0 norm of a vector was surrogated in [21] by the logarithm of the geometric mean of its elements.

Building on this last idea, consider replacing $\|\mathbf{Z}\|_{2,1}$ in (8) with $\sum_{i=1}^N \log(\|\mathbf{z}_i^T\|_2 + \delta)$, where δ is a small positive constant. Since the new surrogate term is concave, the overall minimization problem is non-convex and admittedly more complex to solve than (8). With k denoting iterations, local methods based on iterative linearization of $\log(\|\mathbf{z}_i^T\|_2 + \delta)$ around the current iterate $\mathbf{z}_i^T(k)$, can be adopted to minimize the resulting non-convex cost. Skipping details that can be found in [21], application of the majorization-minimization technique leads to an *iteratively-reweighted* version of (8), namely solve for $k = 0, 1, \ldots$

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_* + \sum_{i=1}^N w_i(k)\|\mathbf{z}_i^T\|_2, \quad \text{s. to } \mathbf{y} = \mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T \text{vec}\big(\mathbf{Z}\big)$$

with weights $w_i(k) := \tau/\big(\|\mathbf{z}_i^T(k-1)\|_2 + \delta\big)$. If the value of $\|\mathbf{z}_i^T(k-1)\|_2$ is small, then in the next iteration the regularization term $w_i(k)\|\mathbf{z}_i^T\|_2$ has a large weight, promoting shrinkage of that entire row vector to zero. Numerical tests in Section 5 suggest that few iterations of the iteratively-reweighted procedure suffice to yield improved recovery of $\{\mathbf{x}, \mathbf{h}\}$, when compared to (8). At a more fundamental level, processing multiple output signals can aid the blind identification task as well, and this is the subject dealt with next.

4. MULTIPLE OUTPUT SIGNALS

Suppose now that we have access to a collection of P (possibly time-indexed) output signals $\{\mathbf{y}_p\}_{p=1}^P$, each one corresponding to a different sparse input $\mathbf{x}_p$ fed to the *common* graph filter $\mathbf{H}$ we wish to identify. Although each of the P identification problems could be solved separately (and naively) as per Section 3, the recovery performance can be improved by tackling them jointly.

While extending the feasibility problem in (3) to this new setup is straightforward [each output gives rise to a couple constraints as in (3)], generalizing the formulation in (8) requires more work. To this end, consider the $NP \times 1$ supervector of stacked output signals $\bar{\mathbf{y}} := [\mathbf{y}_1^T, ..., \mathbf{y}_P^T]^T$, and likewise for the unobserved inputs $\bar{\mathbf{x}} := [\mathbf{x}_1^T, ..., \mathbf{x}_P^T]^T$. Next, introduce the unknown rank-one matrices $\mathbf{Z}_p := \mathbf{x}_p\mathbf{h}^T$, $p = 1, ..., P$, and stack them: (i) vertically in $\bar{\mathbf{Z}}_v := [\mathbf{Z}_1^T, ..., \mathbf{Z}_P^T]^T = \bar{\mathbf{x}}\mathbf{h}^T \in \mathbb{R}^{NP\times L}$; and (ii) horizontally in $\bar{\mathbf{Z}}_h := [\mathbf{Z}_1, ..., \mathbf{Z}_P] \in \mathbb{R}^{N\times PL}$. Note that $\bar{\mathbf{Z}}_v$ is a rank-one matrix. Further, when all $\mathbf{x}_p$ share a common support, then so will all the row-sparse matrices $\mathbf{Z}_p$ (and hence $\bar{\mathbf{Z}}_h$). These observations motivate the following convex formulation [cf. (8)]

$$\min_{\{\mathbf{Z}_p\}_{p=1}^P} \ \|\bar{\mathbf{Z}}_v\|_* + \tau\|\bar{\mathbf{Z}}_h\|_{2,1} \tag{9}$$
$$\text{s. to } \ \bar{\mathbf{y}} = \Big(\mathbf{I}_P \otimes \Big(\mathbf{V}\big(\mathbf{\Psi}^T \odot \mathbf{U}^T\big)^T\Big)\Big)\,\text{vec}\big(\bar{\mathbf{Z}}_h\big)$$

where all P lifted bilinear constraints have been compactly expressed in terms of $\bar{\mathbf{y}}$ and $\text{vec}\big(\bar{\mathbf{Z}}_h\big)$ using a Kronecker product.

When the sparse support is not the same for all $\mathbf{x}_p$, matrix $\bar{\mathbf{Z}}_h$ is not row-sparse. In that case, $\|\bar{\mathbf{Z}}_h\|_{2,1}$ in (9) must be replaced with $\sum_{p=1}^P \|\mathbf{Z}_p\|_{2,1}$, possibly adjusting individual tuning parameters τ_p per signal. Either way, an efficient ADMM solver can be implemented for the multiple signal setting as well, and extensive numerical tests indicated that iteratively-reweighting as in Section 3.1 can yield markedly improved recovery performance (cf. Section 5).

5. NUMERICAL RESULTS

We illustrate the performance of the blind graph filter identification algorithm by solving (9) for different graphs $\mathcal{G}$ and varying the parameters L, S and P. Obtained estimates will be denoted by $\{\hat{\mathbf{x}}, \hat{\mathbf{h}}\}$.

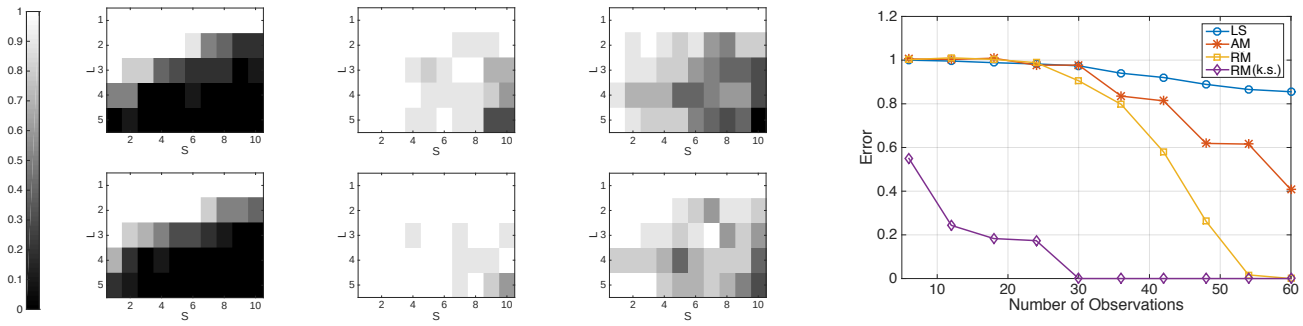

Fig. 1: (a) Rate of recovery of $\mathbf{x}$ and $\mathbf{h}$ as a function of S (non-zeros in $\mathbf{x}$) and L (filter length) in 50-node Erdős-Rényi graphs for $P = 1$ (top) and $P = 5$ (bottom) output signals. (b) Refinement over (a) via iteratively-reweighted optimization. (c) Counterpart of (b) for the brain network in [22]. (d) Recovery errors for several methods as a function of the number of observations in the brain network for $L = S = 3$.

In all cases we define the graph-shift operator as the adjacency matrix of $\mathcal{G}$, $\mathbf{S} = \mathbf{A}$. The "true" vectors $\mathbf{x}_0$ and $\mathbf{h}_0$ are drawn from standard multivariate Gaussian distributions and are normalized to unit length. Given $\mathbf{x}_0$ and $\mathbf{h}_0$, synthetic observations $\mathbf{y}$ are generated from (4). The root-mean-square error RMSE $:= \|\tilde{\mathbf{x}}\tilde{\mathbf{h}}^T - \mathbf{x}_0\mathbf{h}_0^T\|_F$ is adopted as figure of merit to assess recovery performance.

Random graphs. Defining a successful recovery as one with RMSE smaller than 0.01, we empirically estimate the rate of successful recovery on Erdős-Rényi graphs ($N = 50$, $p = 0.3$) [23] as a function of L and S by averaging the success over 10 realizations for each parameter combination; see Fig. 1(a). We solve (9) for $P = 1$ [which reduces to (8); top figure] and $P = 5$ (bottom figure) output signals for $\tau = 20$. As expected, the difficulty of the problem increases when either L or S increase, depicted in the figure by the black area around the bottom-right corner. Moreover, we can appreciate the benefit of considering multiple output signals, e.g., for $P = 5$ the recovery rate when $L = 2$ and $S = 7$ is 0.8 whereas the corresponding rate for $P = 1$ is 0.5. These rates can be substantially improved via the iterative reweighing of the ℓ_2/ℓ_1 mixed norm; see Fig. 1(b). Even for $P = 1$ (top) the recovery rate is not null for every parameter combination, achieving a minimum of 0.3 for $L = 5$ and $S = 10$. Moreover, this minimum can be increased to 0.6 by setting $P = 5$ (bottom). For this latter setting, perfect recovery is achieved consistently for most combinations of L and S.

Brain graph. We now consider a weighted undirected graph $\mathcal{G}$ of the human brain, consisting of $N = 66$ nodes or regions of interest (ROIs) and whose edge weights are given by the density of anatomical connections between regions [22]. The level of activity of each ROI can be represented by a graph signal $\mathbf{x}$, thus successive applications of $\mathbf{S}$ model a linear evolution of the brain activity pattern. Supposing we observe a linear combination (filter) of the evolving states of an originally sparse brain signal, then blind identification amounts to jointly estimating which regions were originally active, the activity in these regions and the coefficients of the linear combination. We mimic the recovery rate analysis performed for Erdős-Rényi graphs and, due to the shown performance advantage, we report the results obtained when using the iterative reweighing refinement; see Fig. 1(c). As expected, the rates obtained for $P = 5$ (bottom) are markedly better than those for $P = 1$ (top). Furthermore, when comparing Figs. 1(b) and 1(c), it is immediate that, for fixed L and S, recovery in the brain network is more challenging than in Erdős-Rényi graphs. This can be explained by the marked structure of the brain network where nodes are divided into two weakly connected halves (hemispheres). Hence, the output signals in one hemisphere are not very informative about the input signals in the opposite hemisphere, increasing the difficulty of recovery.

So far we have assumed that we observe the entire output signal $\mathbf{y}$ when trying to infer $\mathbf{x}$ and $\mathbf{h}$. Nevertheless, it can be the case that we can only observe a subset of the nodes of the graph and try to recover the original signal and filter from this partial observation. Specifically, in Fig. 1(d), we fix $L = S = 3$ and $P = 1$ and analyze the error behavior – median errors across 50 realizations – as a function of the number of observations – accessible values of the output – for different recovery algorithms. Apart from our rank minimization (RM) algorithm we consider a naive least squares (LS) where we solve (4) via a pseudoinverse. Moreover, we consider an alternating minimization (AM) algorithm implemented in two iterative steps: i) Given $\mathbf{x}$, vector $\mathbf{h}$ is found as the least-squares solution of (2), and ii) Given $\mathbf{h}$, vector $\mathbf{x}$ is found by minimizing $\|\mathbf{x}\|_1$ subject to (2) followed by a thresholding operation to retain S non-zero values. These two steps are repeated until convergence, and the algorithm is initialized with the LS estimate of $\mathbf{h}$. Finally, we consider as a benchmark our RM method when the support of $\mathbf{x}$ is known (k.s.).

Our proposed method clearly outperforms the naive LS and the AM approaches. E.g., for 60 observations, 6 less than the total number of nodes, our method achieves a median error of 0 while the errors for AM and LS are 0.43 and 0.86, respectively. Obtaining better results than LS is not surprising since this algorithm does not leverage the fact that $\mathbf{x}$ is sparse. Further, notice that our method outperforms AM even though the latter assumes that the value of S – but not the support of $\mathbf{x}$ – is known. The big gap between RM and RM (k.s.) represents the performance detriment associated to the support of $\mathbf{x}$ being unknown. For situations where partial information about the support is available, e.g. we know a priori that the input signal is null on a subset of nodes, intermediate curves are obtained.

6. CONCLUSIONS

We studied the problem of blind graph filter identification, which extends blind deconvolution of time (or spatial) domain signals to graphs. The developed rank minimization approach outperforms both the LS and AM algorithms and attains perfect reconstruction over a wide range of sparsity levels S and filter lengths L in synthetic and real-world graphs. Ongoing research addresses the issues of identifiability and theoretical guarantees of the convex nuclear-norm relaxation. Also of interest will be to investigate stable recovery in the presence of noise. Finally, a major step forward would be to also estimate the shift operator by bringing to bear methods of network topology inference [3]; see also [6] for a recent related approach to estimate the graph structure of graph signals.

7. REFERENCES

[1] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, Mar. 2013.

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

[3] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*, Springer, New York, NY, 2009.

[4] A. Sandryhaila and J.M.F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[5] S. Segarra, A. G. Marques, and A. Ribeiro, "Distributed implementation of linear network operators using graph filters," in *53rd Allerton Conf. on Commun. Control and Computing*, Univ. of Illinois at U-C, Monticello, IL, Sept. 30- Oct. 2, 2015.

[6] J. Mei and J.M.F. Moura, "Signal processing on graphs: Estimating the structure of a graph," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2015, pp. 5495–5499.

[7] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1711–1732, March 2014.

[8] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *arXiv preprint arXiv:1507.08364*, 2015.

[9] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Interpolation of graph signals using shift-invariant graph filters," in *European Signal Process. Conf. (EUSIPCO)*, Nice, France, Aug. 31 - Sept. 4 2015.

[10] S. Ling and T. Strohmer, "Self-calibration and biconvex compressive sensing," *arXiv preprint arXiv:1501.06864 [cs.IT]*, 2015.

[11] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. of American Control Conference*, 2001, vol. 6, pp. 4734–4739.

[12] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, "Simultaneously structured models with application to sparse and low-rank matrices," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2886–2908, May 2015.

[13] Y. Li, K. Lee, and Y. Bresler, "A unified framework for identifiability analysis in bilinear inverse problems with applications to subspace and sparsity models," *arXiv:1501.06120 [cs.IT]*, 2015.

[14] S. Segarra, G. Mateos, A. G. Marques, and A. Ribeiro, "Blind identification of graph filters with sparse inputs," in *IEEE Intl. Wrksp. Computat. Advances Multi-Sensor Adaptive Process. (CAMSAP)*, Cancun, Mexico, Dec. 13-16, 2015.

[15] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197–2202, Feb. 2003.

[16] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv preprint arXiv:1402.2637*, 2014.

[17] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, August 2010.

[18] J. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. on Information Theory*, vol. 51, pp. 1030–1051, Mar. 2006.

[19] M. Golbabaee and P. Vandergheynst, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2012, pp. 2741–2744.

[20] S. Boyd, N. Parikh, E. Chu, B.Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Machine Learn.*, vol. 3, pp. 1–122, 2011.

[21] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimzation," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, Dec. 2008.

[22] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. J. Honey, V. J Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, no. 7, pp. e159, 2008.

[23] B. Bollobás, *Random Graphs*, Springer, 1998.