ESTIMATION OF NETWORK PROCESSES VIA BLIND GRAPH MULTI-FILTER IDENTIFICATION

Yu Zhu[†]*, Fernando J. Iglesias*[‡]*, Antonio G. Marques*[‡]*, and Santiago Segarra*[†]

[†]Dept. of Electrical and Computer Engineering, Rice University, Houston, TX, USA
[‡]Dept. of Signal Theory and Communications, King Juan Carlos University, Madrid, Spain

ABSTRACT

We study the problem of jointly estimating several network processes that are driven by the same input, recasting it as one of blind identification of a bank of graph filters. More precisely, we consider the observation of several graph signals – i.e., signals defined on the nodes of a graph – and we model each of these signals as the output of a different network process (represented by a graph filter) defined on a common *known* graph and driven by a common *unknown* input. Our goal is to recover the specifications of every network process by only observing the outputs. Since every process shares the same input, the estimation problems are coupled, and a joint inference method is proposed. We study two different scenarios, one where the orders of the filters are known, and one where they are not. For the former case we propose a least-squares approach and provide conditions for recovery. For the latter case, we put forth a sparse recovery algorithm with theoretical guarantees. Finally, we illustrate the methods here proposed via numerical experiments.

Index Terms— blind identification, graph filter, network process, sparse recovery

1. INTRODUCTION

Many biological [1], technological [2], and social systems [3] can be better understood when interpreted as networks, comprising a large number of individual components that interact with each other to generate a global behavior. Associated with these networks, an increasing amount of network data is being collected, primarily comprised of data attached to the nodes or agents that form the network. The modeling and analysis of this type of data have been the subject of study of graph signal processing (GSP). Under the assumption that the signal properties are related to the topology of the underlying network, the goal of GSP is to develop algorithms that fruitfully leverage this relational structure [4–6]. Examples of relevant problems that have been recently addressed using GSP tools include graph signal sampling [7–11], graph-based Wavelet and Fourier transforms [12, 13], and topology identification [14–17], to name a few.

In the past years, substantial effort has been devoted to the development and understanding of graph filters [18–21]. This thrust is driven by the fact that graph filters are versatile models for network processes, allowing us to apply signal processing tools to the study of network science problems. Noteworthy approaches include the optimal design of graph filters [19, 20], the consideration of nonlinear filters [22], and the blind identification of filters [23–25]. Of special interest to us is this last direction, where one tries to estimate the specification of a network process (i.e., the coefficients of the associated graph filter) by only observing one output. Nonetheless, in many scenarios we have access to the output of *multiple* related network processes excited by the same input. For example, different patterns of brain activity when an individual is presented with a common input such as the same visual stimulation. Alternatively, we can think of a single dynamic network process that is sensed at different points in time as corresponding to multiple network processes of different length driven by the same input. For both cases, the common input calls for a joint estimation formulation, which is the focus of the current paper. In classical discrete-time signal processing the related problem of multi-channel system identification is well studied and solutions have been provided based on different assumptions. In particular, some approaches assume statistical knowledge of the input and put forth second-order and higher-order methods for the recovery [26, 27] whereas others treat the inputs as unknown deterministic signals [28, 29]. In the current work we draw inspiration from this latter direction, and tackle the more general problem of graph filter multi-channel identification.

Contributions and paper outline. The paper contains two main contributions. First, for the scenario where the orders of the unknown filters are given, we present a least-squares approach along with necessary and sufficient conditions for recovery. Though related to the classical approach for time-invariant filters, the consideration of a general graph raises interesting technical issues. Second, for the scenario where the filter orders are not given, we propose a sparse recovery formulation and provide theoretical guarantees of performance. The remainder of this paper is organized as follows. In Section 2 we provide basic notions of GSP and classical filter identification needed to introduce our problem. In Section 3 we formulate the problem to be solved, and propose methods for its solution along with theoretical analysis of these methods. Numerical experiments illustrating our developments are presented in Section 4 and closing remarks are included in Section 5.

2. PRELIMINARIES

We first introduce core concepts in GSP that allow us to formulate our problem of joint estimation of network processes, followed by a quick introduction to the classical problem of blind *time-invariant* multi-channel identification.

Fundamentals of graph signal processing. Consider the directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ formed by the set $\mathcal{N}$ of N nodes and the set of edges $\mathcal{E}$, such that the pair (i, j) belongs to $\mathcal{E}$ if there exists an edge from node i to node j. Associated with a given $\mathcal{G}$, a graph signal can be represented as a vector $\mathbf{x} = [x_1, \ldots, x_N]^\top \in \mathbb{R}^N$, where the ith component, x_i, represents the signal value at node i. The network structure is captured by the graph shift operator (GSO) $\mathbf{S}$ [4], a sparse matrix that can take non-zero values if $(i, j) \in \mathcal{E}$ or $i = j$, that is, $[\mathbf{S}]_{ji} = 0$ for $(i, j) \notin \mathcal{E}$ and $i \neq j$. Notice that $\mathbf{S}$ reflects the

This work was supported in part by the Spanish grants MINECO Klinilycs TEC2016-75361-R and Instituto de Salud Carlos III DTS17/00158.

local connectivity of $\mathcal{G}$. Specifically, if $\mathbf{y} = \mathbf{S}\mathbf{x}$, node i can compute y_i as a linear combination of the signal values x_j where j belongs to the incoming neighborhood of i, i.e. $j \in \mathcal{N}_i = \{j \mid (j,i) \in \mathcal{E}\}$. The adjacency matrix [4] and the graph Laplacian [5] are usual choices for the GSO. Assuming that $\mathbf{S}$ is diagonalizable, the shift can be decomposed as $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{V}$ collects the eigenvectors of $\mathbf{S}$ as columns and the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ collects the eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_N]^\top$. Linear graph filters are defined as graph-signal operators that can be expressed as polynomials in $\mathbf{S}$ [4]

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

The filtering operation is thus given by $\mathbf{y} = \mathbf{H}\mathbf{x}$, where $\mathbf{y}$ is the filtered signal, $\mathbf{x}$ the input, and $\mathbf{h} = [h_0, \dots, h_{L-1}]^\top$ the filter coefficients. Moreover, we say that L is the *order* of the filter (i.e., the size of $\mathbf{h}$), and $L-1$ the *degree* of the filter (i.e., the degree of the polynomial $\mathbf{H}$).

Graph signals and filters can also be represented in the frequency (or Fourier) domain. Defining $\mathbf{U} = \mathbf{V}^{-1}$ and $\mathbf{\Psi}$ as a Vandermonde matrix of dimension $N \times L$ whose entry $\Psi_{ij} = \lambda_i^{j-1}$, then the frequency representations of a signal $\mathbf{x}$ and a filter $\mathbf{h}$ can be obtained as $\tilde{\mathbf{x}} = \mathbf{U}\mathbf{x}$ and $\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$. Correspondingly, graph filters can be rewritten as $\mathbf{H} = \mathbf{U}^{-1}(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l)\mathbf{U} = \mathbf{U}^{-1}\text{diag}(\mathbf{\Psi}\mathbf{h})\mathbf{U}$, from where it follows that the frequency representation of the output is given by

$$\tilde{\mathbf{y}} = \mathbf{U}\mathbf{y} = \mathbf{U}\mathbf{H}\mathbf{x} = \text{diag}(\mathbf{\Psi}\mathbf{h})\mathbf{U}\mathbf{x} = \text{diag}(\tilde{\mathbf{h}})\tilde{\mathbf{x}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}, \tag{2}$$

where $\circ$ denotes the elementwise (Hadamard) product. Notice that (2) is the graph equivalence of the classical convolution theorem.

Classical blind multi-channel identification. Following the problem formulation in [28], we have that $x(\cdot)$ represents a discrete-time input signal with corresponding output $y_i(\cdot)$ given by

$$y_i(k) = \sum_{j=0}^{L} h_i(j)x(k-j), \qquad i = 1, \dots, M. \tag{3}$$

More specifically, $y_i(\cdot)$ corresponds to the output of a FIR filter with channel impulse response $h_i(\cdot)$ when excited by the common input $x(\cdot)$. The blind identification problem can then be stated as follows: Given the channel outputs $\{y_i(k), i = 1, \dots, M; k = L, \dots, N\}$, determine the channel specifications $\{h_i(\cdot)\}_{i=1}^M$. Notice that (3) can be compactly written as $y_i(k) = h_i(k) * x(k)$, where $*$ denotes a convolution operation, from where it follows that

$$h_j(k) * y_i(k) = h_j(k) * h_i(k) * x(k) = h_i(k) * y_j(k). \tag{4}$$

In order to leverage the underlying algebraic structure of this problem, we may define the Hankel matrices $\{\mathbf{B}_i\}_{i=1}^M$ of dimension $(N-2L+1) \times (L+1)$, where the (m,n) element of $\mathbf{B}_i$ is $y_i(m+n+L-2)$. We recall the Data Selection Transform (DST) $\mathcal{D}$ [30] applied to $\{\mathbf{B}_i\}_{i=1}^M$, which is recursively given by $\mathcal{D}_2(\mathbf{B}_{(\cdot)}) = [\mathbf{B}_2, -\mathbf{B}_1]$ and

$$\mathcal{D}_m(\mathbf{B}_{(\cdot)}) = \left[\begin{array}{ccc|c} & \mathcal{D}_{m-1}(\mathbf{B}_{(\cdot)}) & & \mathbf{0} \\ \hline \mathbf{B}_m & & & -\mathbf{B}_1 \\ & \ddots & & \vdots \\ & & \mathbf{B}_m & -\mathbf{B}_{m-1} \end{array}\right]. \tag{5}$$

Consequently, by defining $\mathbf{B} = \mathcal{D}_M(\mathbf{B}_{(\cdot)})$, all the possible cross relations of the form in (4) can be condensed in

$$\mathbf{B}\mathbf{h} = \mathbf{0}, \tag{6}$$

where $\mathbf{h}_i = [h_i(L), \cdots, h_i(0)]^\top$ and $\mathbf{h} = [\mathbf{h}_1^\top, \cdots, \mathbf{h}_M^\top]^\top$.

3. JOINT ESTIMATION OF MULTIPLE NETWORK PROCESSES

Leveraging the notions introduced in Section 2, we can formalize the problem to be solved. Based on a GSO $\mathbf{S}$, there exist M filters $\{\mathbf{H}^{(i)}\}_{i=1}^M$ defined as in (1) with associated filter coefficients $\{\mathbf{h}^{(i)}\}_{i=1}^M$. The outputs $\{\mathbf{y}^{(i)}\}_{i=1}^M$ are generated by these filters when excited by the same unknown input $\mathbf{x}$, i.e., $\mathbf{y}^{(i)} = \mathbf{H}^{(i)}\mathbf{x}$.

Problem 1 *Given* $\mathbf{S}$ *and the outputs* $\{\mathbf{y}^{(i)}\}_{i=1}^M$*, recover the filter coefficients* $\{\mathbf{h}^{(i)}\}_{i=1}^M$ *with no knowledge of the common input* $\mathbf{x}$*.*

First notice that the recovery sought here is up to a scalar ambiguity, since we can always multiply all the filter coefficients $\mathbf{h}^{(i)}$ by a scalar α and multiply $\mathbf{x}$ by $1/\alpha$ without modifying the observed outputs. To see why this problem bears practical relevance, note that graph filters can be used to represent linear diffusion dynamics that depend on the network topology [19, 20, 23]. Potential applications range from social networks where a rumor is spread across the network via local opinion exchanges, to brain networks where an epileptic seizure emanating from few regions is later diffused across the entire brain. Indeed, the different outputs $\mathbf{y}^{(i)}$ might correspond to different processes run on the same network, or a single dynamic process that is sensed at different points in time. In this way, the order of the filters – i.e., the size of the vectors $\{\mathbf{h}^{(i)}\}_{i=1}^M$ – indicates the length of the network processes considered. The approach towards solving Problem 1 is fundamentally different depending on whether this length information is given or not. Accordingly, in Section 3.1 we tackle the setting where the orders of the filters are known, whereas in Section 3.2 we consider the more challenging setting where this information is not given.

3.1. Known filter orders

The knowledge of the filter orders L_i allows us to define a series of Vandermonde matrices $\{\mathbf{\Psi}_i\}_{i=1}^M$ of dimension $N \times L_i$ whose (m,n) element is λ_m^{n-1} and a block diagonal matrix $\mathbf{\Psi}$ that contains M blocks respectively given by $\{\mathbf{\Psi}_i\}_{i=1}^M$. Moreover, we define the diagonal matrices $\tilde{\mathbf{Y}}_i = \text{diag}(\tilde{\mathbf{y}}_i)$, collecting the frequency representations of the outputs. Leveraging the classical DST transform [cf. (5)], we may define $\tilde{\mathbf{Y}} = \mathcal{D}_M(\tilde{\mathbf{Y}}_{(\cdot)})$, from where the following characterization of the true filter coefficients follows.

Proposition 1 *Defining* $\mathbf{h} = [\mathbf{h}^{(1)\top}, \mathbf{h}^{(2)\top}, \dots, \mathbf{h}^{(M)\top}]^\top$ *obtained by vertically concatenating the coefficients of the* M *unknown filters, the following expression holds*

$$\tilde{\mathbf{Y}}\mathbf{\Psi}\mathbf{h} = \mathbf{0}. \tag{7}$$

Proof: From (2) it follows that $\tilde{\mathbf{y}}_i = \tilde{\mathbf{h}}_i \circ \tilde{\mathbf{x}}$ for all i. Thus, in a spirit similar to (4) we may write

$$\tilde{\mathbf{h}}_j \circ \tilde{\mathbf{y}}_i = \tilde{\mathbf{h}}_j \circ \tilde{\mathbf{h}}_i \circ \tilde{\mathbf{x}} = \tilde{\mathbf{h}}_i \circ \tilde{\mathbf{h}}_j \circ \tilde{\mathbf{x}} = \tilde{\mathbf{h}}_i \circ \tilde{\mathbf{y}}_j, \tag{8}$$

which implies that $\tilde{\mathbf{h}}_j \circ \tilde{\mathbf{y}}_i - \tilde{\mathbf{h}}_i \circ \tilde{\mathbf{y}}_j = \mathbf{0}$. Leveraging the definition of the DST, we may write the difference between every pair i, j as $\tilde{\mathbf{Y}}\tilde{\mathbf{h}} = \mathbf{0}$. By noting that $\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$ [cf. (2)], the result follows. □

Proposition 1 reveals that whenever $\tilde{\mathbf{Y}}\mathbf{\Psi}$ is well behaved, we can recover the true filter coefficients $\mathbf{h}$. More precisely, if we have that $\text{rank}(\tilde{\mathbf{Y}}\mathbf{\Psi}) = \sum_{i=1}^M L_i - 1$ we may uniquely determine $\mathbf{h}$ up to a multiplicative constant (cf. discussion following Problem 1). In order to get further insights about when this rank

condition holds, it is instrumental to define the scalar polynomials $p^{(i)}(z) = \sum_{l=0}^{L_i-1} h_l^{(i)} z^l$, whose coefficients are given by the true filter coefficients. Moreover, let us define the index set $\Omega \subseteq \{1, 2, \ldots, N\}$ as being the largest possible set satisfying that $\tilde{x}_n \neq 0$ and every λ_n is distinct for all $n \in \Omega$. Based on these notions, we state sufficient conditions for identifiability of the true filter coefficients by solving (7).

Proposition 2 *With Ω being the index set associated with $\mathbf{x}$ and $\mathbf{S}$, (7) admits a unique solution (up to a scalar multiple) if:*
i) $|\Omega| \geq L_{\max} + L_{\min} - 1$, where $L_{\max}$ and $L_{\min}$ are the maximum and minimum values in $\{L_i\}_{i=1}^M$, and
ii) The polynomials $\{p^{(i)}(z)\}_{i=1}^M$ do not share common roots.

Proof: Let us denote by $\mathbf{h}'$ a generic solution of (7), the goal of this proof is to show that under conditions *i)* and *ii)* it must be that $\mathbf{h} = \alpha \mathbf{h}'$. Consider the polynomials $p'^{(i)}(z) = \sum_{l=0}^{L_i-1} h_l'^{(i)} z^l$ associated with the generic solution $\mathbf{h}'$. From (8) it follows that

$$p^{(i)}(\lambda_n) p'^{(j)}(\lambda_n) \tilde{x}_n = p^{(j)}(\lambda_n) p'^{(i)}(\lambda_n) \tilde{x}_n, \tag{9}$$

for all $i, j = 1, \ldots, M$ and for all $n = 1, \ldots, N$. Fix an arbitrary j and choose the index i as one satisfying $L_i = L_{\min}$. Moreover, consider the indices $n \in \Omega$. Since $\tilde{x}_n \neq 0$ and every λ_n is distinct, it follows that the polynomial $p_{ij}(z) = p^{(i)}(z) p'^{(j)}(z) - p^{(j)}(z) p'^{(i)}(z)$ must have at least $|\Omega|$ roots given by the eigenvalues λ_n for $n \in \Omega$. However, the degree of $p_{ij}(z)$ is at most $L_{\max} + L_{\min} - 2$, thus, from condition *i)* we have that $p_{ij}(z) = 0$ for all $z \in \mathbb{C}$. Equivalently, we have that

$$p^{(i)}(z) p'^{(j)}(z) = p^{(j)}(z) p'^{(i)}(z), \tag{10}$$

for all $z \in \mathbb{C}$. It follows that every root z_0 of $p^{(i)}(z)$ – i.e., a zero in the left hand side of (10) – must be a root of $p'^{(i)}(z)$, since by condition *ii)* there exists a j^* such that $p^{(j^*)}(z_0) \neq 0$, and we know that $p^{(j^*)}(z_0) p'^{(i)}(z_0) = 0$. For the case where $p^{(i)}(z)$ has distinct roots (the case of repeated roots can also be shown, but we omit it for brevity), this immediately implies that $p^{(i)}(z) = \alpha p'^{(i)}(z)$ for some scalar α. Finally, replacing this equality in (10) we get that $p^{(j)}(z) = \alpha p'^{(j)}(z)$ for all arbitrary j, from where $\mathbf{h} = \alpha \mathbf{h}'$. □

Proposition 2 states sufficient conditions under which we can solve Problem 1 by simply choosing the unique solution to (7). Notice that the first condition is essentially encoding a notion of spectral richness of the observed process. Indeed, $|\Omega|$ is equal to the number of distinct frequencies in our underlying graph (i.e., distinct eigenvalues λ_n) that are excited by the unknown input ($\tilde{x}_n \neq 0$). Proposition 2 successfully parallels the classical result in [28, Th. 1]. Notice, however, that in [28] every filter was assumed to be of the same length, and the measure of richness used was essentially different since, in the absence of a graph, the richness was exclusively determined by the input. We also present necessary conditions for identifiability.

Proposition 3 *If (7) admits a unique solution (up to a scalar multiple) it must be that:*
i) $|\Omega| \geq L_{\max}$, and
ii) The polynomials $\{p^{(i)}(z)\}_{i=1}^M$ do not share common roots.

Proof: (sketch for brevity) To see that *i)* is indeed a necessary condition it suffices to show that it is a necessary condition for the simpler problem of recovering $\mathbf{h}$ when the common input $\mathbf{x}$ is known. For condition *ii)*, consider the case where there is a common root z_0 shared by all polynomials $\{p^{(i)}(z)\}_{i=1}^M$ so that we can write $p^{(i)}(z) = q^{(i)}(z)(z - z_0)$ for all i. In this case, it can be shown that the filter coefficients associated with the polynomials $q^{(i)}(z)(z - z_0')$ for any choice of z_0' also solve (7). □

Proposition 3 states that whenever the number of distinct frequencies excited by our input is smaller than the number of free parameters in the longest network process (highest-degree graph filter), then recovery is not possible. Similarly, whenever the responses of the filters are not different enough to distinguish them correctly, recovery cannot be achieved. Notice that Propositions 2 and 3 provide separate insights on either sufficient or necessary conditions for recovery. The derivation of necessary *and* sufficient conditions is a matter of current development. Lastly, real-world observations are usually noisy and modeled as $\mathbf{y}^{(i)} = \mathbf{H}^{(i)}\mathbf{x} + \boldsymbol{\epsilon}^{(i)}$, where $\boldsymbol{\epsilon}^{(i)}$ represents the noise of the ith observation. In practice (including the experiments in Section 4), our estimate $\hat{\mathbf{h}}$ for the filter coefficients is simply obtained as $\hat{\mathbf{h}} := \operatorname{argmin}_{\mathbf{h} | \|\mathbf{h}\|_2 = 1} \|\tilde{\mathbf{Y}} \boldsymbol{\Psi} \mathbf{h}\|_2$.

3.2. Unknown filter orders

The approach in Section 3.1 heavily relies on the knowledge of L_i in order to build matrix $\boldsymbol{\Psi}$. Whenever the orders L_i are unknown, our proposed approach is to overshoot them, and then incorporate a complexity regularizer that promotes shorter filters whenever they can appropriately explain the observations. To be more precise, we denote the assumed filter orders by Q_i, where we ensure that $Q_i \geq L_i$ for all i by selecting large enough orders Q_i as guided by available domain knowledge. We then define a series of Vandermonde matrices $\{\boldsymbol{\Theta}_i\}_{i=1}^M$ of dimension $N \times Q_i$ whose (m, n) element is λ_m^{n-1} and a block diagonal matrix $\boldsymbol{\Theta}$ whose M diagonal blocks are given by $\{\boldsymbol{\Theta}_i\}_{i=1}^M$. Notice that $\boldsymbol{\Theta} = \boldsymbol{\Psi}$ whenever $Q_i = L_i$ for all i. Based on the true filter coefficients $\mathbf{h}^{(i)}$, let us define the zero-padded vectors $\bar{\mathbf{h}}^{(i)} = [\mathbf{h}^{(i)\top}, \mathbf{0}_{(Q_i - L_i)\times 1}^\top]^\top$. Consequently, if we define $\bar{\mathbf{h}} = [\bar{\mathbf{h}}^{(1)\top}, \ldots, \bar{\mathbf{h}}^{(M)\top}]^\top$, it follows from Proposition 1 that

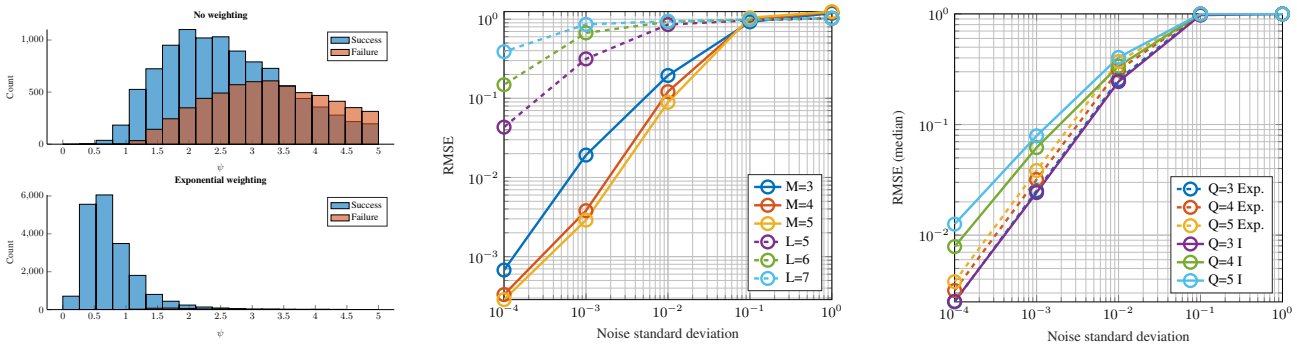
$$\tilde{\mathbf{Y}} \boldsymbol{\Theta} \bar{\mathbf{h}} = \mathbf{0}. \tag{11}$$

The results provided in Section 3.1 are still valid for (11). However, given that we have increased the number of free parameters by overshooting the filter orders, it becomes harder for the sufficient conditions in Proposition 2 to hold. This is an indication that the least-squares approach advocated in Section 3.1 – and in [28] for the classical setting – might not be appropriate for the scenario with unknown filter orders. By contrast, we propose to estimate $\bar{\mathbf{h}}$ by solving the following ℓ_1-analysis problem [31]

$$\hat{\bar{\mathbf{h}}} = \operatorname*{argmin}_{\bar{\mathbf{h}}} \|\boldsymbol{\Delta} \bar{\mathbf{h}}\|_1 \quad \text{s.t. } \tilde{\mathbf{Y}} \boldsymbol{\Theta} \bar{\mathbf{h}} = \mathbf{0},\ \bar{h}_1 = 1, \tag{12}$$

where the diagonal matrix $\boldsymbol{\Delta}$ contains positive predefined weights. More precisely, whenever $\boldsymbol{\Delta} = \mathbf{I}$, (12) seeks among all non-trivial solutions to $\tilde{\mathbf{Y}} \boldsymbol{\Theta} \bar{\mathbf{h}} = \mathbf{0}$ for the sparsest one where the ℓ_1 norm has been used as a convex surrogate of the non-convex ℓ_0 pseudo-norm. Moreover, given that the true $\bar{\mathbf{h}}$ is formed by concatenating multiple zero-padded vectors, when solving (12) we seek to strongly promote zeros for the filter coefficients associated with larger degrees. This can be achieved by setting increasing weights in $\boldsymbol{\Delta}$. Indeed, the results in Section 4 will show that there is a clear advantage associated with the consideration of weight matrices $\boldsymbol{\Delta}$ different from the identity. Lastly, the second constraint in (12) is simply enforced to avoid the trivial solution $\hat{\bar{\mathbf{h}}} = \mathbf{0}$. Given that recovery is always considered up to a scalar multiple, this constraint is only making the mild assumption that $\bar{h}_1^{(1)} \neq 0$ for the true filters.

A fundamental question is whether the solution to (12) indeed recovers the true filter coefficients $\bar{\mathbf{h}}$. In order to answer this, let



(a) (b) (c)

Fig. 1: Experimental results based on $20,000$ Monte Carlo simulations. (a) Validation of Proposition 4 and comparison of weight matrices $\mathbf{\Delta}$. (b) Estimation with known orders in the presence of noise, for varying orders L and number of network processes M. (c) Estimation with unknown orders in the presence of noise, for varying assumed orders Q and weight matrices $\mathbf{\Delta}$.

us introduce the following notation. We denote by $\mathbf{\Phi}$ the matrix obtained by dropping the first column of $\tilde{\mathbf{Y}}\mathbf{\Theta}$. We denote by $\bar{\mathbf{h}}_{(-1)}$ the vector obtained by dropping the first entry of $\bar{\mathbf{h}}$ and by $\mathbf{\Delta}_{(-1)}$ the matrix obtained by dropping the first column and the first row of $\mathbf{\Delta}$. Also, we denote by $\mathcal{I} = \text{supp}(\bar{\mathbf{h}}_{(-1)})$ the indices of the non-zero true coefficients (after dropping $\bar{h}_1$), and by $\mathcal{I}^c$ its complement. Finally, given any matrix $\mathbf{A}$ and a set of indices $\mathcal{S}$, the matrix $\mathbf{A}_{\mathcal{S}}$ is obtained by selecting the columns from $\mathbf{A}$ indicated by $\mathcal{S}$. With this notation in place, the following result holds.[1]

Proposition 4 *The solution to* (12) *coincides with the true filter coefficients* $\bar{\mathbf{h}}$ *if:*
i) $\text{rank}(\mathbf{\Phi}_{\mathcal{I}}) = |\mathcal{I}|$*, and*
ii) There exists a constant $\delta > 0$ *such that*

$$\psi = \|\mathbf{I}_{\mathcal{I}^c}^\top(\delta^{-2}\mathbf{\Delta}_{(-1)}^{-1}\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{\Delta}_{(-1)}^{-1} + \mathbf{I}_{\mathcal{I}^c}\mathbf{I}_{\mathcal{I}^c}^\top)^{-1}\mathbf{I}_{\mathcal{I}}\|_\infty < 1.$$

Proposition 4 provides sufficient conditions for the recovery of the coefficients $\bar{\mathbf{h}}$. The first condition is based on the support of the true filter coefficients and is needed to ensure the uniqueness of the solution to (12), a prerequisite for $\hat{\bar{\mathbf{h}}}$ to coincide with $\bar{\mathbf{h}}$. The second condition takes into account both the weight matrices $\mathbf{\Delta}$ and the size of the complement of the support of the true filters, i.e., how drastically Q_i overshoots L_i. Finally, in practical implementations where noise is present, we replace the first equality constraint in (12) by $\|\tilde{\mathbf{Y}}\mathbf{\Theta}\bar{\mathbf{h}}\|_2 \leq \epsilon$. Guarantees akin to those in Proposition 4 can be derived for this robust formulation but lie outside our present scope.

4. NUMERICAL EXPERIMENTS

All simulations consist of $20,000$ Monte Carlo runs. Inputs, filter coefficients, and noise realizations are drawn from standard Gaussian distributions, with filter coefficients normalized to have unit ℓ_2 norm. When the filter orders are known, the estimates $\hat{\mathbf{h}}$ are obtained from the singular value decomposition of $\tilde{\mathbf{Y}}\mathbf{\Psi}$ whereas a robust version of (12) is used when the filter orders are unknown (cf. last paragraphs of Sections 3.1 and 3.2, respectively).

We begin by examining the impact of parameter ψ in Proposition 4 on the estimation success. For this experiment we consider Erdős-Rényi (ER) graphs with $N = 50$ and edge probability $p = 0.1$, $L = 3$ coefficients per filter, $Q = 5$, and $M = 3$ network processes. In Fig. 1a we present the histograms of successful and failed recoveries as a function of ψ for two choices of $\mathbf{\Delta}$, namely, no weighting ($\mathbf{\Delta} = \mathbf{I}$) and weights that increase exponentially with order of each coefficient. We choose δ from $\{0.1, 1, 10\}$ and keep the one that achieves the lowest ψ (cf. Proposition 4). We first observe that, as expected, every realization for which $\psi < 1$ resulted in a successful recovery. In addition, the exponential weighting matrix improves the performance with respect to the absence of weighting. This improvement can be explained by the observed concentration of the empirical distribution of ψ around values smaller than 1.

We now analyze the effect that varying L and M has in recovering $\mathbf{h}$ when the orders of the filters are known. We first fix $M = 3$ and consider graphs drawn from a stochastic block model ($N = 54$, 3 blocks) with probabilities of intra- and inter-community edge appearance given by $p = 0.3$ and $q = 0.05$, respectively [32]. The dashed plots in Fig. 1b portray that the recovery becomes more challenging for larger values of L, where the RMSE of the recovered coefficients $\hat{\mathbf{h}}$ is computed as $\|\hat{\mathbf{h}} - \mathbf{h}\|_2/\|\mathbf{h}\|_2$. This is expected, since a larger L implies more unknowns to be estimated for the same number of observations. Next, we fix $L = 3$ and vary the number of filters M for ER graphs ($N = 50$, $p = 0.1$). The solid plots in Fig. 1b indicate that recovery is benefited by larger M. Intuitively, since we have a common input, the more filters we observe, the easier it is to perform joint estimation.

In our last experiment, we assume that the filter orders $L = 3$ are unknown and we vary $Q \in \{3, 4, 5\}$ for a fixed value of $M = 3$. Moreover, we analyze the performance of the same two weight matrices $\mathbf{\Delta}$ considered in Fig. 1a on graphs randomly generated from a Barabási-Albert model ($N = 50$, ER seed with $n = 8$ and $p = 0.5$) [32]. In general, Fig. 1c shows that the accuracy decays as Q becomes larger, owing to the additional degrees of freedom. Finally, it becomes apparent that the exponential weighting is more robust to the level of overshooting when $Q > L$.

5. CONCLUSIONS

We have formulated the problem of joint estimation of network processes as a blind multi-channel identification of graph filters. This allowed us to leverage classical signal processing approaches towards the characterization of network processes. We showed that the problem is inherently different when the filter orders are known and when they are not. For the former case we put forth a least-squares approach and specified the conditions for recovery. For the latter case, we proposed a sparse recovery method with theoretical guarantees of performance. Future research avenues include estimation methods for setups where the process specification is partially shared across graph filters, and the identification of non-linear network processes.

[1]Proof can be found at `http://ss187.blogs.rice.edu/files/2018/10/ICASSP19_with_app-1oqs58s.pdf`

6. REFERENCES

[1] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, and R. Chen, "Topological structure analysis of the protein–protein interaction network in budding yeast," *Nucleic Acids Research*, vol. 31, no. 9, pp. 2443–2450, 2003.

[2] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sept. 1999.

[3] A. Garas, P. Argyrakis, C. Rozenblat, M. Tomassini, and S. Havlin, "Worldwide spreading of economic crisis," *New Journal of Physics*, vol. 12, no. 11, pp. 113043, 2010.

[4] 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.

[5] 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, 2013.

[6] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.

[7] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[8] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.

[9] L. F. O. Chamon and A. Ribeiro, "Greedy sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 34–47, Jan. 2018.

[10] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, Jul. 2016.

[11] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 764–778, Feb. 2017.

[12] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Comput. Harmonic Anal.*, vol. 30, no. 2, pp. 129 – 150, 2011.

[13] R. Shafipour, A. Khodabakhsh, G. Mateos, and E. Nikolova, "A digraph Fourier transform with spread frequency components," in *Global Conf. Signal and Info. Process. (GlobalSIP)*, Nov. 2017, pp. 583–587.

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

[15] V. Kalofolias, "How to learn a graph from smooth signals," in *Intl. Conf. Artif. Intel. Stat. (AISTATS)*, 2016, pp. 920–929.

[16] S. Segarra, A.G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Aug. 2017.

[17] Y. Shen, B. Baingana, and G. B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2503–2516, Feb. 2017.

[18] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.

[19] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4117–4131, Aug. 2017.

[20] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Autoregressive moving average graph filtering," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 274–288, Jan. 2017.

[21] O. Teke and P. P. Vaidyanathan, "Extending classical multirate signal processing theory to graphs – Part I: Fundamentals," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 409–422, Jan. 2017.

[22] S. Segarra, A. G. Marques, G. R. Arce, and A. Ribeiro, "Design of weighted median graph filters," in *IEEE Intl. Wrksp. Computat. Advances Multi-Sensor Adaptive Process. (CAMSAP)*, Dec. 2017, pp. 1–5.

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

[24] D. Ramírez, A. G. Marques, and S. Segarra, "Graph-signal reconstruction and blind deconvolution for diffused sparse inputs," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Mar. 2017, pp. 4104–4108.

[25] F. J. Iglesias, S. Segarra, S. Rey-Escudero, A. G. Marques, and D. Ramírez, "Demixing and blind deconvolution of graph-diffused sparse signals," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Apr. 2018, pp. 4189–4193.

[26] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Info. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[27] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Info. Theory*, vol. 36, no. 2, pp. 312–321, Mar. 1990.

[28] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[29] M. I. Gurelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 134–149, Jan. 1995.

[30] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Process.*, vol. 50, no. 1-2, pp. 83–99, Apr. 1996.

[31] H. Zhang, M. Yan, and W. Yin, "One condition for solution uniqueness and robustness of both l1-synthesis and l1-analysis minimizations," *Adv. in Comp. Math.*, vol. 42, no. 6, pp. 1381–1399, Dec. 2016.

[32] M. Newman and A.-L. Barabasi ans D. J. Watts, *The Structure and Dynamics of Networks*, vol. 19, Princeton University Press, 2011.