DEMIXING AND BLIND DECONVOLUTION OF GRAPH-DIFFUSED SPARSE SIGNALS

Fernando J. Iglesias[†], Santiago Segarra[‡], Samuel Rey-Escudero[†], Antonio G. Marques[†], and David Ramírez[,⋆]*

[†]Dept. of Signal Theory and Comms., King Juan Carlos University, Madrid, Spain
[‡]Inst. for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA
[*]Dept. of Signal Theory and Comms., University Carlos III of Madrid, Leganés, Spain
[⋆]Gregorio Marañón Health Research Institute, Madrid, Spain

ABSTRACT

This paper generalizes the classical joint problem of signal demixing and blind deconvolution to the realm of graphs. We investigate a setup where a single observation formed by the sum of multiple *graph signals* is available. The main assumption is that each individual signal is generated by an originally *sparse input* diffused through the graph via the application of a *graph filter*. In this context, we address the related problems of: 1) separating the individual graph signals, 2) identifying the unknown input supports, and 3) estimating the coefficients of the diffusing graph filters. We first consider the case where each signal – prior to mixing – is diffused in a different graph. We then particularize the results for the more challenging case where all the signals are diffused in the same graph. The corresponding demixing and blind graph-signal deconvolution problems are formulated, convex relaxations are presented, and recovery conditions are discussed. Numerical experiments in both the single and multiple graph cases show the capabilities of demixing in synthetic and biology-inspired graphs.

***Index Terms*—** Blind signal reconstruction, blind system identification, demixing, graph signal processing, source separation.

1. INTRODUCTION

In many contemporary applications, the signals of interest are defined on irregular domains that can be conveniently represented by graphs. Depending on the particular setup, the graph may correspond to an actual (electrical, biological, social) network where the signal is observed, or encode (pairwise) statistical relationships between the signal values. Under the assumption that the signal properties are related to the topology of the associated graph, the goal of graph signal processing (GSP) is to develop algorithms that fruitfully leverage this relational structure [1]. Examples of relevant problems that have been recently addressed using GSP tools include signal reconstruction [2–5], modeling and inference of diffusion processes [6, 7], and topology identification [8, 9], to name a few.

The work of three first authors was suported by the Spanish MINECO grants OMICRON (TEC2013-41604-R) and KLINILYCS (TEC2016-75361-R). S. Segarra was also supported by the MIT IDSS seed grant. The work of the last author was partly supported by the Spanish MINECO grants OTOSIS (TEC2013-41718-R) and COMONSENS Network (TEC2015-69648-REDC); by the Spanish MINECO and the European Commission (ERDF) grants ADVENTURE (TEC2015-69868-C2-1-R) and CAIMAN (TEC2017-86921-C2-2-R); and by the Comunidad de Madrid grant CASI-CAM-CM (S2013/ICE-2845).

In this paper we investigate how to generalize classical demixing and blind sparse deconvolution to graph signals. Formally, consider a graph with N nodes, and suppose that the observed signal $\mathbf{y} \in \mathbb{R}^N$ adheres to the model $\mathbf{y} = \sum_{p=1}^{P} \mathbf{y}_p$ with $\mathbf{y}_p = \mathbf{H}_p\mathbf{x}_p$. Each $\mathbf{y}_p$ represents a diffused graph signal generated by the application of a *graph filter* $\mathbf{H}_p \in \mathbb{R}^{N\times N}$ to a *sparse* input $\mathbf{x}_p \in \mathbb{R}^N$ with unknown support. For this model, our aim is to solve the following problem: Given the topology of the graphs, the observed signal $\mathbf{y}$ and side information on the inputs and the graph filters, find $\{\mathbf{x}_p\}_{p=1}^P$ and $\{\mathbf{H}_p\}_{p=1}^P$. Since graph filters implement local diffusion dynamics [7, 10], this model is of practical interest in applications such as opinion formation and source identification in social networks, inverse problems of biological signals supported on graphs, and modeling and estimation of diffusion processes in multi-agent networks. Signals adhering to this generative model will be referred to as *diffused sparse* signals [11].

Contributions and related works. In Sec. 2 we investigate the general case where each signal has been generated via a diffusion in a different (possibly related) graph. Then, in Sec. 3 we look at the particular setup where all the signals are diffused in the same graph. For the latter case, separability becomes more challenging, and additional assumptions on $\{\mathbf{x}_p\}_{p=1}^P$ and $\{\mathbf{H}_p\}_{p=1}^P$ are required to guarantee recovery. In both sections, we formulate the joint demixing and blind deconvolution (non-convex) problem, propose suitable convex relaxations, and discuss recoverability. Blind reconstruction of graph signals lying on subspaces has been analyzed in [4, 12] for graph bandlimited signals (sparse in the frequency domain) and in [11] for graph diffused signals (sparse in the node domain). Blind deconvolution of graph signals has been studied in [10, 11] under the assumption that each observation corresponds to a single diffused signal. Here, we look at the more challenging problem where $\{\mathbf{y}_p\}_{p=1}^P$ are not known and only $\mathbf{y} = \sum_{p=1}^{P} \mathbf{y}_p$ is observed. Finally, for observations adhering to a *classical convolution* model, demixing and blind deconvolution have been recently studied in [13] (for non-sparse inputs) and in [14] (for sparse inputs). Our contribution in this case is on the consideration of graph signals and filters –which modify the definition of convolution–, as well as the inclusion of the setting with multiple generating graphs.

1.1. Fundamentals of graph signal processing

This section reviews preliminary GSP concepts that will be used in our problem formulation. To facilitate exposition, in this section we assume that $P = 1$ so that the subscript p can be dropped. Additional details on GSP can be found in, e.g., [1, 15]. Consider the

directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ formed by the set $\mathcal{N}$ of N nodes and the set of links $\mathcal{E}$, such that the pair (i, j) belongs to $\mathcal{E}$ if there exists a link from node i to node j. Associated with a given $\mathcal{G}$, a graph signal can be represented as a vector $\mathbf{x} = [x_1, \ldots, x_N]^T \in \mathbb{R}^N$, where the ith component, x_i, represents the signal value at node i. The network structure is captured by the graph shift operator $\mathbf{S}$ [8, 16], a sparse matrix with non-zero values if $(i, j) \in \mathcal{E}$ or $i = j$, that is, $[\mathbf{S}]_{ji} \neq 0$ for $(i, j) \in \mathcal{E}$ or $i = j$. The adjacency matrix [8, 16] and the graph Laplacian [1] are usual choices for the shift operator. Assuming that $\mathbf{S}$ is diagonalizable, the shift can be decomposed as $\mathbf{S} = \mathbf{V}\text{diag}(\boldsymbol{\lambda})\mathbf{V}^{-1}$, where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_N]^T$ collects the eigenvalues. Linear graph filters are defined as graph-signal operators of the form $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$, i.e., polynomials in $\mathbf{S}$ [16]. The filtering operation is thus given by $\mathbf{y} = \mathbf{H}\mathbf{x}$, where $\mathbf{y}$ is the filtered signal, $\mathbf{x}$ the input, $\mathbf{h} = [h_0, \ldots, h_{L-1}]^T$ the filter coefficients, and $L - 1$ the filter degree.

Frequency interpretation. As in classical SP, graph filters and signals may be represented in the frequency (or Fourier) domain. Defining the graph Fourier operator as $\mathbf{U} = \mathbf{V}^{-1}$, the graph Fourier transform (GFT) of the signal $\mathbf{x}$ is $\tilde{\mathbf{x}} = \mathbf{U}\mathbf{x}$. For graph filters, the definition of the GFT that maps the filter coefficients $\mathbf{h}$ to the frequency response of the filter, $\tilde{\mathbf{h}}$, is given by $\tilde{\mathbf{h}} = \boldsymbol{\Psi}\mathbf{h}$, where $\boldsymbol{\Psi}$ is a $N \times L$ Vandermonde matrix whose elements are $[\boldsymbol{\Psi}]_{i,j} = \lambda_i^{j-1}$ [7]. Note that while in classical signal processing both Fourier transforms are the same, here $\mathbf{U} \neq \boldsymbol{\Psi}$. With $\circ$ denoting the Hadamard (elementwise) product, the definitions of the GFT for signals and filters allow us to rewrite the filtering operation in the spectral domain as $\tilde{\mathbf{y}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}$. Finally, let $\odot$ denote the Khatri-Rao (or columnwise Kronecker) product and $\text{vec}(\cdot)$ the vectorization operator that stacks the columns of the matrix input. Then, for the purpose of joint input and filter identification, the dependence of $\mathbf{y}$ on $\mathbf{x}$ and $\mathbf{h}$ can be alternatively written as [10]

$$\mathbf{y} = \mathbf{V}(\boldsymbol{\Psi}^T \odot \mathbf{U}^T)^T \text{vec}(\mathbf{x}\mathbf{h}^T). \tag{1}$$

1.2. Mixtures of diffused sparse signals

This paper considers an observation $\mathbf{y}$ formed as the sum of P signals $\{\mathbf{y}_p\}_{p=1}^P$, each of them generated as $\mathbf{y}_p = \mathbf{H}_p\mathbf{x}_p$, with $\mathbf{H}_p = \sum_{l=0}^{L_p-1} h_{p,l}\mathbf{S}_p^l$ and $\mathbf{x}_p$ having at most Q_p non-zero entries. Mathematically, this means that signal $\mathbf{y}_p$ belongs to a subspace of dimension (at most) Q_p spanned by a subset of the columns of $\mathbf{H}_p$. The particular subspace depends on the network topology encoded in $\mathbf{S}_p$, the filter coefficients $\mathbf{h}_p = [h_{p,0}, ..., h_{p,L_p-1}]^T$, and the support of $\mathbf{x}_p$.

To see why this model bears practical relevance, note that graph filters can be used to represent linear diffusion dynamics that depend on the network topology [7, 10, 11]. Potential applications range from social networks where a rumor originated by a small group of people 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.

2. DEMIXING WITH MULTIPLE DIFFUSING GRAPHS

Given an observation $\mathbf{y} = \sum_{p=1}^P \mathbf{y}_p$, we aim at identifying the P individual signals under the assumption that each of them can be modeled as a graph-diffused sparse input [cf. (1)] with *unknown* support. With $\boldsymbol{\Psi}_p$, $\mathbf{V}_p$ and $\mathbf{U}_p$ denoting the corresponding matrices associated with the graph shift $\mathbf{S}_p$, and leveraging the results presented in the previous sections, the joint demixing and blind deconvolution problem can be written as

$$\{\hat{\mathbf{x}}_p, \hat{\mathbf{h}}_p, \hat{\mathbf{y}}_p\}_{p=1}^P = \text{find} \quad \{\mathbf{x}_p, \mathbf{h}_p, \mathbf{y}_p\}_{p=1}^P \tag{2a}$$

$$\text{s. to} \quad \mathbf{y}_p = \mathbf{V}_p \left(\boldsymbol{\Psi}_p^T \odot \mathbf{U}_p^T\right)^T \text{vec}(\mathbf{x}_p\mathbf{h}_p^T) \tag{2b}$$

$$\mathbf{y} = \textstyle\sum_{p=1}^P \mathbf{y}_p, \qquad \|\mathbf{x}_p\|_0 \leq Q_p. \tag{2c}$$

When solving (2), the observation $\mathbf{y}$ as well as the matrices $\boldsymbol{\Psi}_p$, $\mathbf{V}_p$, and $\mathbf{U}_p$, and the parameters Q_p are assumed to be known. To handle the above demixing problem, note first that the constraint (2b) implies that signals $\mathbf{y}_p$ can be viewed as dummy variables fully determined by $\mathbf{x}_p$ and $\mathbf{h}_p$. Although this reduces considerably the size of the feasible set, the optimization is challenging for a number of reasons. Firstly, even if the support of the inputs were known, for the problem to be well-posed, the number of non-trivial unknowns $\sum_{p=1}^P (Q_p + L_p)$ needs to be less than the number of observations N. Secondly, each of the terms $\mathbf{x}_p\mathbf{h}_p^T$ is bilinear, which introduces a source of non-convexity and gives rise to an inherent scaling ambiguity. Thirdly, the presence of the ℓ_0 norm renders the optimization NP-hard.

The first step to design a tractable relaxation of (2), is to *lift* the problem by defining the $N \times L_p$ rank-one matrices $\mathbf{Z}_p = \mathbf{x}_p\mathbf{h}_p^T$, together with the $N \times NL_p$ transfer matrices $\mathbf{M}_p = \mathbf{V}_p \left(\boldsymbol{\Psi}_p^T \odot \mathbf{U}_p^T\right)^T$. With these notational conventions, and combining the *linear* constraints in (2b)-(2c) into a single one, the demixing problem can be equivalently written as

$$\{\hat{\mathbf{Z}}_p\}_{p=1}^P = \text{find} \quad \{\mathbf{Z}_p\}_{p=1}^P \tag{3}$$

$$\text{s. to} \quad \mathbf{y} = \textstyle\sum_{p=1}^P \mathbf{M}_p \text{vec}(\mathbf{Z}_p), \;\; \text{rank}(\mathbf{Z}_p) = 1, \;\; \|\mathbf{Z}_p\|_{2,0} \leq Q_p.$$

Note that $\|\mathbf{Z}_p\|_{2,0}$, defined as the number of non-zero rows of $\mathbf{Z}_p$, performs the role of $\|\mathbf{x}_p\|_0$ in (2). The vectors $\hat{\mathbf{x}}_p$ and $\hat{\mathbf{h}}_p$ are given by the scaled versions of the right and left principal *singular vectors* of the rank-one matrix $\hat{\mathbf{Z}}_p$, and each of the individual outputs is found as $\hat{\mathbf{y}}_p = \mathbf{M}_p \text{vec}(\hat{\mathbf{Z}}_p)$. Although the equivalent formulation (3) is still difficult to solve, it leads to a convex relaxation which is described in the ensuing section, after the following remark.

Remark 1 (Robust demixing) The demixing problem – in particular constraint (2b) – can be rewritten to account for noisy and incomplete observations. If the values of $\mathbf{y}$ are observed only at a subset of M nodes, both sides of (2b) must be left multiplied by the *sampling matrix* $\mathbf{C}_{\mathcal{M}} \in \{0, 1\}^{M \times N}$ whose rows correspond to canonical basis vectors identifying the indexes $\mathcal{M} = \{n_m\}_{m=1}^M$ of the observed entries. Presence of noise in the observations would require replacing the equality in (2b) with a term penalizing the difference between the two sides of (2b), with the penalty depending on the specific type of noise.

2.1. Convex relaxation and algorithmic approach

The main idea is to reformulate (3) as a *minimization* problem where the objective consists of tractable surrogates of the non-convex rank and $\ell_{2,0}$ constraints. To simplify exposition, consider replacing the rank with the nuclear norm, and the $\ell_{2,0}$ norm with the $\ell_{2,1}$ norm [17, 18]. The resultant optimization is

$$\{\hat{\mathbf{Z}}_p\}_{p=1}^P = \underset{\{\mathbf{Z}_p\}_{p=1}^P}{\text{argmin}} \quad \textstyle\sum_{p=1}^P \eta_p \|\mathbf{Z}_p\|_* + \sum_{p=1}^P \beta_p \|\mathbf{Z}_p\|_{2,1} \tag{4}$$

$$\text{s. to} \quad \mathbf{y} = \textstyle\sum_{p=1}^P \mathbf{M}_p \text{vec}\left(\mathbf{Z}_p\right),$$

where $\{\eta_p, \beta_p\}_{p=1}^P$ are tuning constants that, if no prior information is available, can be set to $\eta_p = 1$ and $\beta_p = \beta$ for all p. The optimization in (4) is convex, so that it can be handled with generic off-the-shelf solvers. When computational complexity is a concern, one can alternatively develop efficient algorithms tailored to the structure of (4), see, e.g., [19] for a related problem.

Recoverability conditions can be found by ignoring the low-rank nature of all the $\mathbf{Z}_p$ and focusing on the fact that the concatenation of the vectorized forms of all these matrices $\mathbf{z} = \text{vec}([\mathbf{Z}_1, \ldots, \mathbf{Z}_P])$ is effectively a sparse vector. In this way, one can think of recovering a sparse vector $\mathbf{z}$ satisfying $\mathbf{y} = [\mathbf{M}_1, \ldots, \mathbf{M}_P]\,\mathbf{z}$ and derive recoverability conditions in this basic setting [20]. More sophisticated conditions that take into account either the structure of the linear operators $\mathbf{M}_p$ – as done in [10] for blind deconvolution – or the row-sparse plus low-rank nature of $\mathbf{Z}_p$ are a matter of ongoing research.

Remark 2 (Non-convex relaxations) Alternative relaxations of the combinatorial constraints in (3) can be pursued. In particular, based on established sparse reconstruction techniques [17, 18, 21, 22], we may replace the cost function in (4) with

$$\sum_{p=1}^P \left(\eta_p \sum_{l=1}^{L_p} \log\left(\sigma_{p,l} + \epsilon_1\right) + \beta_p \sum_{n=1}^N \log\left(\|\mathbf{z}_{p,n}^T\|_2 + \epsilon_2\right)\right),$$

where $\sigma_{p,l}$ is the lth singular value of $\mathbf{Z}_p$; $\mathbf{z}_{p,n}^T$ is the nth row of $\mathbf{Z}_p$; and ϵ_1 and ϵ_2 are small positive constants. Two challenges of this approach are: 1) The resultant cost to minimize is concave and the problem must be addressed using an iterative majorization-minimization approach [21, 23]. 2) The minimization of $\sum_l \log\left(\sigma_{p,l} + \epsilon_1\right)$ with respect to the non-square matrix $\mathbf{Z}_p$ is not straightforward. One way to overcome this issue is to rely on the semidefinite embedding lemma as done in, e.g., [11, 21].

3. DEMIXING WITH A SINGLE DIFFUSING GRAPH

If all the signals are diffused in the same graph and then combined, then the joint demixing and deconvolution problem is inherently more challenging. To explain this more rigorously, note that if $\mathbf{S}_p = \mathbf{S}$ for all p, then the transfer matrix $\mathbf{M}_p = \mathbf{M}$ is equivalent for all the sources. As a result, the observation equation $\mathbf{y} = \sum_{p=1}^P \mathbf{M}_p \text{vec}(\mathbf{Z}_p)$ can be rewritten as $\mathbf{y} = \mathbf{M}\text{vec}(\sum_{p=1}^P \mathbf{Z}_p)$, which introduces a new source of ambiguity. To be more specific, consider the minimization form of the feasibility problem in (3), i.e., the original non-relaxed precursor of (4),

$$\{\hat{\mathbf{Z}}_p\}_{p=1}^P = \underset{\{\mathbf{Z}_p\}_{p=1}^P}{\text{argmin}} \quad \sum_{p=1}^P \eta_p \,\text{rank}(\mathbf{Z}_p) + \sum_{p=1}^P \beta_p \|\mathbf{Z}_p\|_{2,0}$$
$$\text{s. to} \quad \mathbf{y} = \mathbf{M}\,\text{vec}\Big(\sum_{p=1}^P \mathbf{Z}_p\Big) \tag{5}$$

Suppose for simplicity that $P = 2$, denote by $\mathbf{Z}_1^*$ and $\mathbf{Z}_2^*$ the true rank-one demixing matrices, and use these to define $\mathbf{Z}_1' = \mathbf{Z}_1^* + \mathbf{Z}_2^*$ and $\mathbf{Z}_2' = \mathbf{0}$. It readily follows that: 1) both $\{\mathbf{Z}_1^*, \mathbf{Z}_2^*\}$ and $\{\mathbf{Z}_1', \mathbf{Z}_2'\}$ are feasible solutions[1] of (5); and 2) from the triangle inequality we have that $\text{rank}(\mathbf{Z}_1')+\text{rank}(\mathbf{Z}_2') = \text{rank}(\mathbf{Z}_1') \leq \text{rank}(\mathbf{Z}_1^*)+\text{rank}(\mathbf{Z}_2^*)$. In words, the true matrices $\{\mathbf{Z}_1^*, \mathbf{Z}_2^*\}$ never achieve a smaller (rank) cost than $\{\mathbf{Z}_1', \mathbf{Z}_2'\}$. Moreover, if $\eta_1 \neq \eta_2$ then a solution where either $\mathbf{Z}_1' = \mathbf{0}$ or $\mathbf{Z}_2' = \mathbf{0}$ always achieves a smaller rank cost. A similar argument holds for the $\ell_{2,0}$ norm. This demonstrates that the demixing problem with a single diffusing graph is challenging because the individual $\mathbf{Z}_p$ are in general non-identifiable and only $\mathbf{Z} = \sum_{p=1}^P \mathbf{Z}_p$ could be recovered. Thus, we propose a two-step approach in which we first efficiently obtain the true $\mathbf{Z}^* = \sum_{p=1}^P \mathbf{Z}_p^*$ and then provide conditions for which the constituent $\mathbf{Z}_p^*$ can be uniquely recovered.

3.1. Convex relaxation and algorithmic approach

As done in the Section 2.1, a natural step to design an efficient algorithm for (5) is to replace the combinatorial pseudo-norms in its objective with their convex surrogates $\|\cdot\|_*$ and $\|\cdot\|_{2,1}$. This leads to an instance of (4) with $\mathbf{M}_p = \mathbf{M}$ for all p. Using arguments similar to those provided after (5), the solutions to this problem $\{\hat{\mathbf{Z}}_p\}_{p=1}^P$ also suffer from a linear ambiguity.[2] An alternative is to reformulate (4) in terms of $\mathbf{Z} = \sum_{p=1}^P \mathbf{Z}_p$ as follows

$$\hat{\mathbf{Z}} = \underset{\mathbf{Z}}{\text{argmin}} \;\; \|\mathbf{Z}\|_* + \beta\|\mathbf{Z}\|_{2,1} \quad \text{s. to} \;\; \mathbf{y} = \mathbf{M}\text{vec}\,(\mathbf{Z})\,. \tag{6}$$

The computational complexity of this reformulation is lower since it effectively reduces the number of optimization variables by a factor of P. Moreover, the formulation in (6) is similar to that of blind deconvolution (without demixing) with the exception that the true $\mathbf{Z}^*$ to be recovered, albeit being low-rank, is not rank-one in general. Nonetheless, the results in [10, Theorem 1] can still be leveraged to obtain probabilistic guarantees of recovering the true $\mathbf{Z}^*$ in terms of the matrices $\mathbf{\Psi}$ and $\mathbf{U}$ associated with the common diffusing graph.

As previously mentioned, recovering the $\mathbf{Z}_p^*$ from $\mathbf{Z}^*$ is not feasible in general. Hence, one way to ensure recoverability is to impose additional conditions on the filter coefficients and the input signals. To this end we consider the following two scenarios:

(As1) $\mathbf{x}_p^T\mathbf{x}_{p'} = 0$ for all $p \neq p'$ or $\tilde{\mathbf{x}}_p^T\tilde{\mathbf{x}}_{p'} = 0$ for all $p \neq p'$.

(As2) $\mathbf{h}_p^T\mathbf{h}_{p'} = 0$ for all $p \neq p'$ or $\tilde{\mathbf{h}}_p^T\tilde{\mathbf{h}}_{p'} = 0$ for all $p \neq p'$.

Note that the orthogonality required in these scenarios holds true if, for example, the vectors involved are sparse and their support is non-overlapping. In this way, two sparse inputs with non-overlapping nodes satisfy the first requirement in (As1). Similarly, two bandpass filters defined in non-overlapping bands satisfy the second requirement in (As2). For future reference, we define the matrix $\mathbf{T}_\mathbf{x} = \mathbf{I}$ if the first requirement in (As1) holds and $\mathbf{T}_\mathbf{x} = \mathbf{U}$ if the second requirement is true. Similarly, we define $\mathbf{T}_\mathbf{h} = \mathbf{I}$ if the first requirement in (As2) is true, and $\mathbf{T}_\mathbf{h} = \mathbf{\Psi}$ if the second one holds.

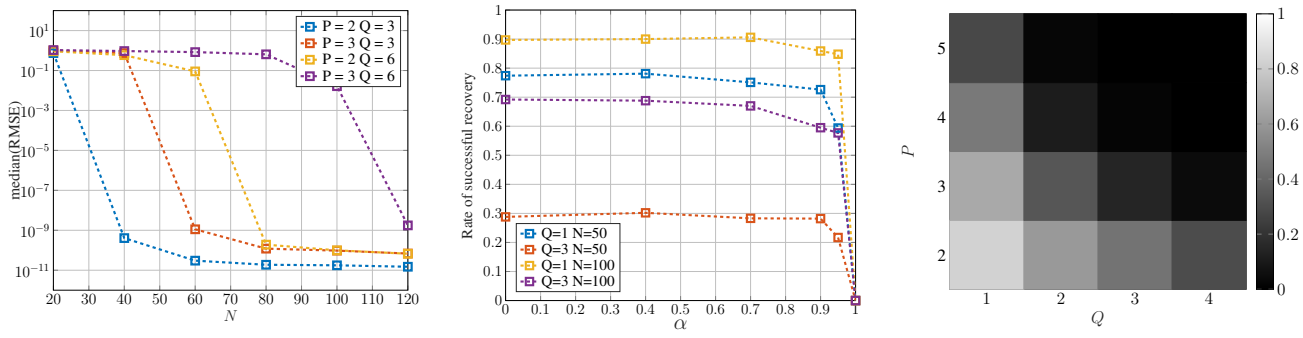
Proposition 1 *If the input signals $\{\mathbf{x}_p\}_{p=1}^P$ satisfy (As1), the diffusing graph filters $\{\mathbf{h}_p\}_{p=1}^P$ satisfy (As2) and $\|\mathbf{T}_\mathbf{x}\mathbf{x}_p\|^2\|\mathbf{T}_\mathbf{h}\mathbf{h}_p\|^2 \neq \|\mathbf{T}_\mathbf{x}\mathbf{x}_{p'}\|^2\|\mathbf{T}_\mathbf{h}\mathbf{h}_{p'}\|^2$ for all $p \neq p'$; then $\{\mathbf{Z}_p^*\}_{p=1}^P$ can be recovered from $\mathbf{Z}^*$.*

The proposition can be proved by construction. First, we form the matrix $\mathbf{T} = \mathbf{T}_\mathbf{x}\mathbf{Z}^*\mathbf{T}_\mathbf{h}^T$ and factorize it using the singular value decomposition (SVD) as $\mathbf{T} = \mathbf{L}\text{diag}(\boldsymbol{\sigma})\mathbf{R}^T$. Then, we set the estimated $\mathbf{Z}_p^*$ as

$$\widehat{\mathbf{Z}}_p^* = \sigma_p\,(\mathbf{T}_\mathbf{x}^{-1})\mathbf{l}_p\mathbf{r}_p^T(\mathbf{T}_\mathbf{h}^\dagger)^T, \tag{7}$$

[1] Note that any combination of the form $\{\mathbf{Z}_1^\theta = \theta_1\mathbf{Z}_1^* + \theta_2\mathbf{Z}_2^*, \mathbf{Z}_2^\theta = (1-\theta_1)\mathbf{Z}_1^* + (1-\theta_2)\mathbf{Z}_2^*\}$ is feasible as well.

[2] Since both the nuclear and the $\ell_{2,1}$ norms are absolutely homogeneous (the rank and the $\ell_{2,0}$ penalties are not), the ambiguity is more severe in the convex formulation. To see this, note that any solution of the form $\mathbf{Z}_p = \theta_p\mathbf{Z}^*$ with $\theta_p \geq 0$ and $\sum_{p=1}^P \theta_p = 1$ is feasible and achieves the same $\sum_{p=1}^P \|\mathbf{Z}_p\|_*$. Hence, the postulated optimization will have multiple solutions with the same cost than that of the actual $\mathbf{Z}_p^*$ that we aim to identify.



(a) (b) (c)

Fig. 1: Blind demixing results for scenarios with: (a) single random graph and different diffusing filters; (b) two random graphs coupled via the "similarity" parameter α (when $\alpha = 0$ the two graphs are statistically independent, when $\alpha = 1$ the graphs are the same) ; and (c) between 2 and 6 brain graphs (the rate of successful recovery is represented).

where σ_p is the pth largest singular value of $\mathbf{T}$, $\mathbf{l}_p$ and $\mathbf{r}_p$ its associated left and right singular vectors, and † denotes the pseudoinverse. Then $\widehat{\mathbf{Z}}_p^* = \mathbf{Z}_p^*$ for all p follows from the fact that, under the conditions in the proposition, the constituent vectors of the rank-one matrix $\mathbf{T}_p = \mathbf{T}_\mathbf{x}\mathbf{x}_p^*(\mathbf{T}_\mathbf{h}\mathbf{h}_p^*)^T$ are orthogonal to those of any other $\mathbf{T}_{p'}$ with $p' \neq p$. Moreover, the condition $\|\mathbf{T}_\mathbf{x}\mathbf{x}_p\|^2\|\mathbf{T}_\mathbf{h}\mathbf{h}_p\|^2 \neq \|\mathbf{T}_\mathbf{x}\mathbf{x}_{p'}\|^2\|\mathbf{T}_\mathbf{h}\mathbf{h}_{p'}\|^2$ ensures that there is no rotation ambiguity in the SVD decomposition, thus guaranteeing recovery.

Remark 3 (A priori information) Identifiability and recovery benefit when prior information on $\{\mathbf{y}_p, \mathbf{h}_p, \mathbf{x}_p\}_{p=1}^P$ exists. This is particularly helpful in the single-graph scenario, where the inherent ambiguity is larger. A natural way to incorporate *probabilistic* information is to augment the objective of our optimization with tractable (log) prior distributions. Alternatively, we may have *deterministic* knowledge of some of (the entries of) the signals or the filter taps. This can be incorporated in the form of linear equality constraints. A particular case that simplifies considerably the structure of the optimization is that of $\{\mathbf{h}_p\}_{p=1}^P$ being known. In such a case, the bilinearity disappears and the resultant problem can be handled using classical sparse signal reconstruction tools. Another scenario with practical relevance is when K_p values of the signal $\mathbf{x}_p$ are known. Since we optimize over $\mathbf{Z}_p$ and not over $\mathbf{x}_p$, exploiting this information requires considering the slightly more involved constraints $\mathbf{z}_{p,i}^T x_{p,i+1} = \mathbf{z}_{p,i+1}^T x_{p,i}$ for $i \in \mathcal{K}_p$, where $\mathcal{K}_p = \{n_i\}_{i=1}^{K_p}$ collects the indexes of the entries of $\mathbf{x}_p$ that are known; see [11] for details.

4. NUMERICAL RESULTS

We present the recovery performance for three blind demixing scenarios. Two of them consider Erdős-Rényi random graphs [24] with edge probability 0.1. The third one considers graphs representing regions of the human brain and their density of anatomical connections for several individuals [25]. In all three cases the adjacency matrix is used as shift, and the signals $\mathbf{x}_p$ as well as the coefficients $\mathbf{h}_p$ are drawn from a standard multivariate Gaussian. The vectors $\mathbf{x}_p$ and $\mathbf{h}_p$ are normalized to have unit norm, and the number of filter taps is $L = 3$. For the optimization we implemented the log-det relaxation [11,21]. Comparisons with other source localization algorithms are of interest, but, due to the space limitations, they are left as future work. Results are obtained from 1000 realizations. The demixing performance is measured via the median of the RMSE, defined as $\text{RMSE} = (\sum_{p=1}^P(\|\hat{\mathbf{Z}}_p - \mathbf{Z}_p\|_F^2)/\|\mathbf{Z}_p\|_F^2)^{1/2}$, or via the rate of successful demixing, i.e., $\Pr\left[\text{RMSE} < 10^{-3}\right]$.

Fig. 1a shows the results of demixing with a single diffusing graph for varying N. Vectors $\mathbf{x}_p$ are normalized to have the same ℓ_1 norm but different ℓ_2 norm, so that Prop. 1 is satisfied. A few settings with different number of filters P and number of non-zero input nodes Q are analyzed. As expected, the performance worsens for larger values of P and Q. Moreover, we observe that the recovery for $(P=3, Q=3)$ is harder than for $(P=2, Q=6)$, even when the latter has more coefficients to estimate, suggesting that Q is the critical parameter for recovery. The second scenario explores the impact of graph similarity in multi-graph deximing with two graphs. These two graphs are coupled so that α percent of their edges are the same. Note that $\alpha = 1$ leads to demixing with a single diffusing graph. Fig. 1b shows that as expected the multi-graph demixing strategy fails for $\alpha = 1$, although the performance does not considerably drop until $\alpha = 0.7$ (or even until $\alpha > 0.9$ for $N = 50$ and $Q = 3$). Finally, Fig. 1c depicts the successful recovery rate using brain graphs and $N = 66$. The results show that demixing is indeed feasible, although the performance decreases quickly as Q and P increase. This is not surprising since the brain graphs at hand exhibit strong similarities. A compelling observation stems from comparing the results between $(Q=3, P=2)$ in Fig. 1c and $(\alpha=0, Q=3, N=50)$ in Fig. 1b. The former, and presumably harder, problem of demixing with non-random graphs presents better performance than the latter; possibly due to the larger number of nodes (66 vs. 50). One final comment is in order. The scenarios analyzed in Figs. 1b and 1c suggest that the graph topologies have an important role in producing amenable blind deximing problems. An analytical characterization of the recovery performance is ongoing research and will be reported elsewhere.

5. CONCLUSIONS

This paper generalized the problems of joint demixing and blind deconvolution to graph signals. Two setups were investigated: one where each of the individual signals was diffused in a different graph, and a more challenging one where the graph was the same for all of them. The resultant non-convex recovery problems were relaxed and the recovery in both settings was discussed. Numerical results demonstrated the practical relevance of our algorithms in synthetic graphs and in biological-inspired networks implementing local diffusion dynamics. Ongoing work includes assessing how the properties of the graphs, signals, and filters affect the recovery performance.

6. 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, 2013.

[2] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, 2014, pp. 3864–3868.

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

[4] 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.

[5] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on graphs: Uncertainty principle and sampling," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4845–4860, Sept. 2016.

[6] A. Loukas, A. Simonetto, and G. Leus, "Distributed autoregressive moving average graph filters," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1931–1935, 2015.

[7] 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.

[8] 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, Jun. 2014.

[9] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary graph processes and spectral estimation," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5911–5926, Nov. 2017.

[10] 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, March 2017.

[11] D. Ramírez, A. G. Marques, and S. Segarra, "Graph-signal reconstruction and blind deconvolution for diffused sparse inputs," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 4104–4108.

[12] R. Varma, S. Chen, and J. Kovacević, "Spectrum-blind signal recovery on graphs," in *IEEE Intl. Wrksp. Computat. Advances Multi-Sensor Adaptive Process. (CAMSAP)*, Cancun, Mexico, Dec. 2015, pp. 81–84.

[13] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *IEEE Trans. Info. Theory*, vol. 63, no. 7, pp. 4497–4520, July 2017.

[14] A. Flinth, "Sparse blind deconvolution and demixing through $\ell_{1,2}$-minimization," *Adv. Comput. Math.*, Apr. 2017.

[15] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.

[16] 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.

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

[18] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Info. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

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

[20] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, New York, NY, USA: Springer, 2013.

[21] M. Fazel, H. Hindi, and S. P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *American Control Conf.*, 2003.

[22] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[23] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

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

[25] 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.