Sampling and Reconstruction of Diffused Sparse Graph Signals from Successive Local Aggregations

Samuel Rey-Escudero, Fernando J. Iglesias, Cristóbal Cabrera, and Antonio G. Marques, *Senior Member, IEEE*

***Abstract*—We analyze the sampling and posterior recovery of *diffused sparse graph signals* from observations gathered at a single node using an *aggregation sampling scheme*. Diffused sparse graph signals can be modeled as the output of a linear graph filter to a sparse input, and are useful in scenarios where a few seeding (source) nodes generate a non-zero input which is then diffused according to the network dynamics dictated by the filter. Instead of considering a traditional setup where the observations correspond to the signal values at a subset of nodes, here the observations are obtained locally at a single node via the *successive aggregation* of its own value and that of its neighbors. Depending of the particular application, the goal is to use the *local* observations to recover the diffused signal or (the location and values of) the seeds. Different sampling configurations are investigated, including those of known and unknown location of the sources as well as that of the diffusing filter being unknown.**

***Index Terms*—Diffused Graph Signals, Interpolation, Aggregation Sampling, Source identification, Active Sampling.**

I. Introduction

The prevalence of network science and big data has motivated the emergence of graph Signal Processing (SP), whose goal is to extend classical SP tools to signals defined on irregular domains represented by a graph. Graph SP problems which have recently received attention include the sampling of bandlimited graph signals (BGS) [1]–[4], filtering [5]–[9], and network topology inference [10]–[12], to name a few.

This paper investigates the recovery of diffused sparse graph signals (DSGS) from observations taken at a particular node using an aggregation sampling scheme (AGSS) [4]. DSGS are a class of signals that can be modeled as a sparse graph signal (zero everywhere except in a few seeding nodes) which is then diffused through the network via a graph filter. Our ultimate goal is to *reconstruct both* the observed signal and the seeds from the available observations. The AGSS is a sampling method for graph signals (introduced in [4]) where nodes successively aggregate the values of the signal in their neighborhood. Recovery can be guaranteed even if observations are gathered at a single node and it can be implemented distributively. Our *contribution* is the generalization of the AGSS, which was originally proposed for BGS, for DSGS. Additionally, we generalize existing results for support identification and blind deconvolution with observations from the AGSS. The algorithms presented in this paper are relevant also for distributed estimation and source localization. Sampling and recovery using as input the signal value at a subset of nodes were discussed in [1]–[3] for BGS, and in [13] for DSGS. Aggregation and space-shift sampling (a generalization of the AGSS considering multiple nodes) for BGS were investigated in [4]. Blind deconvolution and filter identification for DSGS in a centralized setup with access to the full signal (no sampling) were investigated in [14].

Paper organization: Sec. II reviews graph SP concepts and defines formally DSGS and the AGSS. Sec. III presents the main results, including the recovery schemes with known and unknown seeds, as well as unknown diffusing filter. The effect of noise, the design of the sampling matrix and consideration of more than one sampling node are also briefly analyzed. Succinct illustrative numerical results, including some showcasing practical relevance in real datasets, are presented in Sec. IV.

A shorter version of this technical report (with reduced simulations) has been submitted to IEEE Signal Processing Letters (March 2018). For citation purposes, please cite the version submitted to IEEE. Work supported by the Spanish grants TEC2013-41604-R, TEC2016-75361-R, and FPU17-04520. All authors are with the Dept. of Signal Theory and Comms., King Juan Carlos Univ. Emails: {samuel.rey.escudero, fernando.iglesias, cristobal.cabrera,antonio.garcia.marques}@urjc.es

II. Fundamentals of Graph SP

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ denote a directed graph, where $\mathcal{N}$ is the set of nodes, with cardinality N, and $\mathcal{E}$ is the set of edges, with $(i,j) \in \mathcal{E}$ if i is connected to node j. The set $\mathcal{N}_i := \{\, j \,|(j,i) \in \mathcal{E}\}$ denotes the incoming neighborhood of node i. For a given $\mathcal{G}$, the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ is sparse with non-zero elements A_{ij} if and only if $(j,i) \in \mathcal{E}$. If $\mathcal{G}$ is unweighted, the elements A_{ij} are binary. If the graph is weighted, then the value of A_{ij} captures the strength of the link between i and j. The focus of this paper is not on analyzing $\mathcal{G}$, but a graph signal defined on its set of nodes. Such a signal can be represented as a vector $\mathbf{x} = [x_1, \ldots, x_N]^T \in \mathbb{R}^N$ where the i-th entry represents the signal value at node i.

The graph-shift operator (GSO). The GSO $\mathbf{S}$ is defined as an $N \times N$ matrix whose entry S_{ij} can be non-zero only if $i = j$ or $(j,i) \in \mathcal{E}$. Common choices for $\mathbf{S}$ are $\mathbf{A}$ [5] and the graph Laplacian [10]. $\mathbf{S}$ represents a linear transformation that can be computed *locally*. Specifically, if $\mathbf{y} = [y_1, \ldots, y_N]^T$ is defined as $\mathbf{y} = \mathbf{S}\mathbf{x}$, then node i can compute y_i provided that it has access to the values of x_j at its neighbors $j \in \mathcal{N}_i$. We assume that $\mathbf{S}$ is diagonalizable so that there exists an $N \times N$ matrix $\mathbf{V}$ and a diagonal matrix $\mathbf{\Lambda}$ such that $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$.

Graph filters. Given a GSO $\mathbf{S} \in \mathbb{R}^{N\times N}$, the linear transformation represented by the matrix $\mathbf{H} \in \mathbb{R}^{N\times N}$ is called a linear shift-invariant graph filter if $\mathbf{H}$ can be written as [5]

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

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. To show the relevance of graph filters in the context of *local* linear transformations [6]–[8], define the l-th shifted signal $\mathbf{z}^{(l)} := \mathbf{S}^l\mathbf{x}$ and further define the $N \times N$ matrix

$$\mathbf{Z} := [\mathbf{z}^{(0)}, \mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(N-1)}] = [\mathbf{x}, \mathbf{S}\mathbf{x}, \ldots, \mathbf{S}^{N-1}\mathbf{x}], \quad (2)$$

that groups the signal $\mathbf{x}$ and the result of the first $N-1$ applications of the GSO. It is then clear that: a) the output of the graph filter can be found as $\mathbf{y} = \mathbf{H}\mathbf{x} = \mathbf{Z}\mathbf{h}$, with $\mathbf{h}$ being zero-padded if $L < N$; b) since $\mathbf{S}$ is a local operator, the l-th column of $\mathbf{Z}$ can be found locally from the $(l-1)$-th one as $\mathbf{z}^{(l)} = \mathbf{S}\mathbf{z}^{(l-1)}$; and c) as one moves right-wise in (2), the columns of $\mathbf{Z}$ can be viewed as the evolution of a process which is diffused linearly according to the local structure codified in $\mathbf{S}$. Indeed, it is known that graph filters are useful to encode linear network dynamics [8], [14].

DSGS. Given a GSO $\mathbf{S} \in \mathbb{R}^{N\times N}$, the signal $\mathbf{x} \in \mathbb{R}^N$ is called a diffused sparse graph signal of order S if it can be written as

$$\mathbf{x} = \mathbf{H}\mathbf{s}, \text{ where } \mathbf{H} = \textstyle\sum_{l=0}^{N-1} h_l\mathbf{S}^l \text{ and } \|\mathbf{s}\|_0 \leq S. \quad (3)$$

Clearly, signals in (3) can be viewed as the state reached after the diffusion process modeled by $\mathbf{H}$ is over, and the sparse input $\mathbf{s} \in \mathbb{R}^N$ has been spread through the graph.

Frequency domain representation. Use the eigenvectors of $\mathbf{S}$ to define the matrix $\mathbf{U} := \mathbf{V}^{-1}$, and its eigen-values to define the Vandermonde matrix $\mathbf{\Psi} \in \mathbb{R}^{N\times L}$, where $\Psi_{ij} := (\Lambda_{ii})^{j-1}$. Then, the frequency representations of a *signal* $\mathbf{x}$ and of a *filter* $\mathbf{h}$ are defined as $\tilde{\mathbf{x}} := \mathbf{U}\mathbf{x}$ and $\tilde{\mathbf{h}} := \mathbf{\Psi}\mathbf{h}$, with $\mathbf{U}$ and $\mathbf{\Psi}$ acting as Graph Fourier Transforms (GFT) [8], [15]. Exploiting those, the output $\mathbf{y} = \mathbf{H}\mathbf{x}$ in the frequency domain is given by

$$\tilde{\mathbf{y}} = \text{diag}\big(\mathbf{\Psi}\mathbf{h}\big)\mathbf{U}\mathbf{x} = \text{diag}\big(\tilde{\mathbf{h}}\big)\tilde{\mathbf{x}} = \tilde{\mathbf{h}} \circ \tilde{\mathbf{x}}. \quad (4)$$

with $\circ$ denoting the entry-wise product. Identity (4) is the counterpart of the convolution theorem for time signals [8].

A. Sampling BGS

Consider the sampling set $\mathcal{M} \subseteq \mathcal{N}$ with cardinality $M \leq N$ and suppose that we are interested in obtaining the M values of the vector $\mathbf{x}$ corresponding to the entries indexed by $\mathcal{M}$. Then, use $\mathcal{M}$ to define the (fat) selection matrix $\mathbf{C}_\mathcal{M} \in \{0,1\}^{M\times N}$ matrix whose elements satisfy: $\sum_j C_{\mathcal{M},ij} = 1$ for all i; and $\sum_i C_{\mathcal{M},ij} = 1$ if $j \in \mathcal{M}$ and $\sum_i C_{\mathcal{M},ij} = 0$ otherwise. With these notational conventions, the sampled version of the graph signal $\mathbf{x}_\mathcal{M}$ is then $\mathbf{x}_\mathcal{M} := \mathbf{C}_\mathcal{M}\mathbf{x}$. If the sampling set is $\mathcal{M} = \{1, ..., M\}$, then the selection sampling is denoted as simply $\mathbf{C}_M$ and corresponds to the first M rows of the identity matrix of size N. While in classical SP uniform sampling exhibits a number of advantages [16], it is not clear how to design good selection matrices $\mathbf{C}_\mathcal{M}$ for graph signals. For the particular case of BGS (which are signals $\mathbf{x}$ that can be written as $\mathbf{x} = \mathbf{V}_K\tilde{\mathbf{x}}_K$, with $\tilde{\mathbf{x}}_K \in \mathbb{C}^K$ representing the $K \leq N$ first frequency coefficients and $\mathbf{V}_K = \mathbf{V}\mathbf{C}_K^T \in \mathbb{C}^{N\times K}$ their associated eigenvectors), two sampling schemes, briefly presented next, have been proposed: selection and aggregation sampling [1], [2], [4]. While for time varying signals both schemes reduce to traditional sampling [4], for more general graph topologies, they produce different outcomes.

In selection sampling, the vector collecting the observations is $\mathbf{x}_\mathcal{M} = \mathbf{C}_\mathcal{M}\mathbf{V}_K\tilde{\mathbf{x}}_K$. The goal is to recover $\mathbf{x}$ from $\mathbf{x}_\mathcal{M}$, which can be done via $\mathbf{x} = \mathbf{V}_K\tilde{\mathbf{x}}_K = \mathbf{V}_K(\mathbf{C}_\mathcal{M}\mathbf{V}_K)^\dagger\mathbf{x}_\mathcal{M}$, with † denoting the pseudoinverse [2].

An equally valid, but less intuitive approach, is to fix a node, say i, and consider the sampling of the signal seen by this node as the GSO $\mathbf{S}$ is applied recursively. In other words, as the signal has been locally diffused according to $\mathbf{S}$, as described in (2). Then, using the definition of matrix $\mathbf{Z}$ in (2) and with $\mathbf{e}_i$ denoting the i canonical vector, the successively aggregated signal at node i is the i-th row of $\mathbf{Z}$, that is $\mathbf{z}_i := (\mathbf{e}_i^T\mathbf{Z})^T = \mathbf{Z}^T\mathbf{e}_i$. Sampling is now reduced to the selection of M out of the N elements of $\mathbf{z}_i$, that is $\mathbf{z}_{\mathcal{M},i} := \mathbf{C}_\mathcal{M}\mathbf{z}_i = \mathbf{C}_\mathcal{M}\big(\mathbf{Z}^T\mathbf{e}_i\big)$. Leveraging the results in [4], the signal $\mathbf{x}$ can be recovered from $\mathbf{z}_{\mathcal{M},i}$ as $\mathbf{x} = \mathbf{V}_K(\mathbf{C}_\mathcal{M}\mathbf{\Psi}^T\text{diag}(\boldsymbol{\upsilon}_i))^\dagger\mathbf{z}_{\mathcal{M},i}$ with $\boldsymbol{\upsilon}_i := [V_{i,1}, ..., V_{i,N}]^T$.

III. Aggregation Sampling of DSGS

This section considers the problem of reconstructing DSGS from *local* observations obtained from AGSS. In contrast with BGS, the interest when dealing with DSGS can be either in recovering $\mathbf{x}$ (signal reconstruction) or in recovering $\mathbf{s}$ (distributed source localization and estimation). For that reason, we start by analyzing the case where $\mathbf{H} = \mathbf{I}$ and $\mathbf{x} = \mathbf{s}$. After that, we discuss the more general case where the nodes sample the signal $\mathbf{x} = \mathbf{H}\mathbf{s}$, both for known and unknown $\mathbf{H}$. A collaborative setting where more than one node collects the samples closes the section. To help readability, a summary of the setups considered is provided in Table I.

Scenario	Sparse support	$\mathbf{H}$	$\mathbf{C}_\mathcal{M}$	Obs. matrix	Equation
Sparse recovery	Known	Known $\mathbf{H} = \mathbf{I}$	Fixed	$\mathbf{\Theta}$	(8)
Active sampling	Known	Known $\mathbf{H} = \mathbf{I}$	Flexible	$\mathbf{\Theta}$	(10)
Blind sparse recovery	Unknown	Known $\mathbf{H} = \mathbf{I}$	Fixed	$\mathbf{\Theta}$	(11)
Diffused recovery	Known	Known	Fixed	$\mathbf{\Xi}$	(8), replacing $\mathbf{\Theta}$ with $\mathbf{\Xi}$
Diffused active recovery	Known	Known	Flexible	$\mathbf{\Xi}$	(10), replacing $\mathbf{\Theta}$ with $\mathbf{\Xi}$
Blind diffused recovery	Unknown	Known	Fixed	$\mathbf{\Xi}$	(11), replacing $\mathbf{\Theta}$ with $\mathbf{\Xi}$
Blind deconvolution	Unknown	Unknown	Fixed	$\mathbf{\Phi}$	(14)

TABLE I: Compilation of the settings considered in Sec. III.

A. Aggregating the sparse input

A critical aspect to analyze the recovery $\mathbf{x} = \mathbf{s}$ from its aggregated samples is to write the relationship between the sampled signal $\mathbf{z}_\mathcal{M}$ and the sparse input $\mathbf{s}$. For the ease of exposition, we do that in the form of a lemma.

Lemma 1: Given the GSO $\mathbf{S} \in \mathbb{R}^{N\times N}$, the sampling matrix $\mathbf{C}_\mathcal{M}$ and a sparse input $\mathbf{x} = \mathbf{s}$, the shifted signal $\mathbf{z}_i$ and its sampled version $\mathbf{z}_{\mathcal{M},i}$ can be expressed as

$$\mathbf{z}_{\mathcal{M},i} = \mathbf{C}_\mathcal{M}\mathbf{z}_i \text{ and } \mathbf{z}_i = \mathbf{\Psi}^T\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{U}\mathbf{s} := \mathbf{\Theta}_i\mathbf{s}. \quad (5)$$

For the expression above, note that we have defined the observation matrix $\mathbf{\Theta}_i := \mathbf{\Psi}^T\text{diag}(\boldsymbol{\upsilon}_i)\mathbf{U}$, where we recall that $\mathbf{\Psi}$ is the GFT for filters, $\mathbf{U}$ is the GFT for signals, and $\boldsymbol{\upsilon}_i = [V_{i,1}, ..., V_{i,N}]^T$. While nontrivial, (5) can be derived after substituting (2) into (3), or by a minor modification

of the proof in [4, Lemma 2]. Interestingly, (5) reveals that $\mathbf{z}_{\mathcal{M},i}$ depends on how strongly the seeds express each of the frequencies (represented by $\mathbf{U}$s), how strongly the sampling node senses each of the frequencies (represented by the frequency pattern $\boldsymbol{\upsilon}_i$) and the spectral effect of the diffusion (powers of the GSO) captured in the Vandermode matrix $\boldsymbol{\Psi}$.

To proceed with the recovery of the sparse input, let us denote as $\mathcal{S}$ the support of $\mathbf{s}$, define $\mathbf{s}_{\mathcal{S}} := \mathbf{C}_{\mathcal{S}}\mathbf{s} \in \mathbb{R}^S$ as the vector collecting the non-zero values of $\mathbf{s}$, and suppose for now that this support is known. Then, we have that

$$\mathbf{z}_{\mathcal{M},i} = \mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{s} = \mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{C}_{\mathcal{S}}^T\mathbf{s}_{\mathcal{S}}. \tag{6}$$

This setup would correspond to the case where the indexes of the seeding nodes (identity of the influencers or location of the sources) is known, but their particular values are not. Clearly, for the recovery problem in (6) being identifiable, a necessary condition is M, the number of observations in $\mathbf{z}_{\mathcal{M},i}$, being no less than S, the number of unknowns in $\mathbf{s}_{\mathcal{S}}$. Consider first the extreme case $M = S$. It is then clear that the sparse signal $\hat{\mathbf{s}}^{(i)}$, with the superscript (i) denoting the index of the sampling node, can be recovered as

$$\hat{\mathbf{s}}^{(i)} = \mathbf{C}_{\mathcal{S}}^T\hat{\mathbf{s}}_{\mathcal{S}}^{(i)} \text{ and } \hat{\mathbf{s}}_{\mathcal{S}}^{(i)} = (\mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{C}_{\mathcal{S}}^T)^{-1}\mathbf{z}_{\mathcal{M},i}, \tag{7}$$

provided that the inverse exists, which will depend on the particular set of rows $\mathcal{M}$ and columns $\mathcal{S}$. For the standard case of $M > S$ and the observations being corrupted by additive noise, the *observed* signal $\mathbf{y}_i$ is given by $\mathbf{y}_i = \mathbf{z}_i + \mathbf{w}_i$. Assuming that the noise $\mathbf{w}_i$ is zero-mean, independent of $\mathbf{x}$, and with known covariance matrix $\mathbf{R}_w^{(i)} := \mathbb{E}[\mathbf{w}_i\mathbf{w}_i^H]$, the best linear unbiased estimator of the sparse inputs is [17]

$$\hat{\mathbf{s}}^{(i)} = \mathbf{C}_{\mathcal{S}}^T\hat{\mathbf{s}}_{\mathcal{S}}^{(i)} \text{ and } \hat{\mathbf{s}}_{\mathcal{S}}^{(i)} = (\mathbf{M}_{\mathcal{M}})^{\dagger}(\mathbf{R}_{w,\mathcal{M}}^{(i)})^{-1/2}\mathbf{y}_{\mathcal{M},i}, \text{ with } \tag{8}$$

$$\mathbf{R}_{w,\mathcal{M}}^{(i)} := \mathbf{C}_{\mathcal{M}}\mathbf{R}_w^{(i)}\mathbf{C}_{\mathcal{M}}^T \text{ and } \mathbf{M}_{\mathcal{M}} := (\mathbf{R}_{w,\mathcal{M}}^{(i)})^{-1/2}\mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{C}_{\mathcal{S}}^T,$$

provided that $\mathbf{R}_{w,\mathcal{M}}^{(i)}$ is non-singular. Note that if the noise is Gaussian, the estimator in (8) attains the Cramér-Rao bound.

Using (8), the error covariance matrix is [17]

$$\mathbf{R}_e^{(i)} := \mathbb{E}[(\mathbf{s}_{\mathcal{S}} - \hat{\mathbf{s}}_{\mathcal{S}}^{(i)})(\mathbf{s}_{\mathcal{S}} - \hat{\mathbf{s}}_{\mathcal{S}}^{(i)})^T] = (\mathbf{M}_{\mathcal{M}}^T\mathbf{M}_{\mathcal{M}})^{\dagger}, \tag{9}$$

which depends on the noise model, the spectrum of the GSO, the seed nodes, the node taking the observations, and the sample-selection scheme adopted. (9) can then be used to assess the performance of the estimation. The particular error metric depends on the application at hand [18]. The most commonly used is the Mean Square Error (MSE), which corresponds to minimizing $\mathrm{trace}(\mathbf{R}_e^{(i)})$, with other popular metrics including the spectral norm $\lambda_{\max}(\mathbf{R}_e^{(i)})$ and the log determinant $\log\det(\tilde{\mathbf{R}}_e^{(i)})$.

*1) **Active sampling**:* Critical for the error performance is the design of a good sampling matrix. This requires solving first the optimal scheme for a fixed node i (which is a relevant problem by itself) and then selecting the best node (provided that the system operating conditions yield such a possibility). Except for trivial cases, optimizing the sampling set is NP-hard. In particular, when the interest is in the MSE, the optimal sampling matrix $\mathbf{C}_{\mathcal{M}}^{(i)*}$ corresponds to

$$\min_{\mathbf{C}_{\mathcal{M}}} \ \mathrm{trace}\Big((\mathbf{C}_{\mathcal{S}}\boldsymbol{\Theta}_i^T\mathbf{C}_{\mathcal{M}}^T(\mathbf{C}_{\mathcal{M}}\mathbf{R}_w^{(i)}\mathbf{C}_{\mathcal{M}}^T)^{-1}\mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{C}_{\mathcal{S}}^T)^{\dagger}\Big)$$
$$\text{s.t. } C_{\mathcal{M},ij} \in \{0,1\}, \ \textstyle\sum_j C_{\mathcal{M},ij} = 1, \ \|\mathbf{C}_{\mathcal{M}}\|_0 = M, \tag{10}$$

which is a fractional high-order polynomial minimization over binary variables. While convex relaxations to approximate (10) are available, greedy schemes for related problems have been shown to work well [19], and are advocated here.

*2) **Blind (sparse) recovery**:* In many relevant applications (e.g., those dealing with inverse problems) the seeds in $\mathcal{S}$ are unknown. In that case, the noiseless recovery requires solving

$$\mathbf{s}^* := \arg\min_{\mathbf{s}} \ ||\mathbf{s}||_0 \ \text{ s.t. } \ \mathbf{z}_{\mathcal{M},i} = \mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{s}, \tag{11}$$

and then setting $\hat{\mathbf{x}}^{(i)} = \hat{\mathbf{s}}^{(i)} = \mathbf{s}^*$. Leveraging results from sparse recovery, it can be shown that identifiability needs $M \geq 2S$ and the observation matrix $\boldsymbol{\Theta}_i$ to be full spark [20]. The problem in (11), which does not account for noise, is NP-hard due to the presence of the ℓ_0 norm. A standard approach is to relax the equality with a least squares cost and relax the ℓ_0 norm with a (weighted) ℓ_1 regularizer. This yields

$$\mathbf{s}_1^{(i)} := \arg\min_{\mathbf{s}} \|(\mathbf{R}_{w,\mathcal{M}}^{(i)})^{-1/2}\big(\mathbf{y}_{\mathcal{M},i} - \mathbf{C}_{\mathcal{M}}\boldsymbol{\Theta}_i\mathbf{s}\big)\|_2^2 + \gamma||\mathbf{s}||_1,$$

where $(\mathbf{R}_{w,\mathcal{M}}^{(i)})^{-1/2}$ accounts for the colored noise and γ is the regularization parameter.

B. Aggregating the diffused sparse input

We now analyze the recovery of $\mathbf{x} = \mathbf{H}\mathbf{s}$ from its aggregated samples, assuming first that the filter $\mathbf{H}$ is known. The main difference with reference to the previous case is that now the relation between $\mathbf{s}$ and $\mathbf{z}_{\mathcal{M},i}$ is

$$\mathbf{z}_{\mathcal{M},i} = \mathbf{C}_{\mathcal{M}}\mathbf{z}_i \text{ and } \mathbf{z}_i = \boldsymbol{\Psi}^T\mathrm{diag}(\boldsymbol{\upsilon}_i \circ \tilde{\mathbf{h}})\mathbf{U}\mathbf{s} := \boldsymbol{\Xi}_i\mathbf{s}, \tag{12}$$

where $\boldsymbol{\Xi}_i := \boldsymbol{\Psi}^T\mathrm{diag}(\boldsymbol{\upsilon}_i \circ \tilde{\mathbf{h}})\mathbf{U}$. Compared with (5), we notice that the observations depend not only on the frequency pattern of the sampling node $\boldsymbol{\upsilon}_i$, but also on the frequency response of the diffusing filter $\tilde{\mathbf{h}}$. Intuitively, nodes with a frequency pattern more aligned with that of the diffusing filter (so that $|\det(\mathrm{diag}(\boldsymbol{\upsilon}_i \circ \tilde{\mathbf{h}}))|$ is large) are more likely to give rise to a better reconstruction in the presence of noise.

Apart from replacing the observation matrix $\boldsymbol{\Theta}_i$ with $\boldsymbol{\Xi}_i$, another important difference stems from the particular error to minimize. Since in this case $\mathbf{x}$ and $\mathbf{s}$ are different, depending on the application the focus can be on estimating $\mathbf{s}$ and minimizing the MSE associated with $\mathbf{R}_{e,s}^{(i)} := \mathbb{E}[(\mathbf{s}_{\mathcal{S}} - \hat{\mathbf{s}}_{\mathcal{S}}^{(i)})(\mathbf{s}_{\mathcal{S}} - \hat{\mathbf{s}}_{\mathcal{S}}^{(i)})^T]$ or on estimating $\mathbf{x}$ and minimizing the MSE associated with $\mathbf{R}_{e,x}^{(i)} := \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}^{(i)})(\mathbf{x} - \hat{\mathbf{x}}^{(i)})^T]$. This requires to premultiply (or not) the error terms in the objectives of the optimizations presented in the previous sections. Such a dichotomy was not an issue for BGS since the frequency coefficients are a byproduct and the ultimate goal was to recover $\mathbf{x}$. By contrast, for DSGS both $\mathbf{x}$ (signal reconstruction) and $\mathbf{s}$ (source localization) are meaningful on their own.

C. Blind deconvolution

There may be scenarios where the diffusing filter $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$ is unknown. The recovery problem in this case is considerably more challenging, but it can be tackled provided that L, the order of the diffusing filter, is sufficiently small. To be specific, after some manipulations, the expression $\mathbf{x} = \mathbf{H}\mathbf{s}$ with $\mathbf{H}$ being a graph filter can be written as $\mathbf{x} = \mathbf{V}(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T \text{vec}(\mathbf{s}\mathbf{h}^T)$, where $\odot$ stands for the Khatri-Rao product (cf. [14]). We can then relate the samples $\mathbf{z}_{\mathcal{M},i}$ with the unknown $\mathbf{s}$ and $\mathbf{h}$ as

$$\mathbf{z}_{\mathcal{M},i} = \mathbf{C}_{\mathcal{M}} \mathbf{\Psi}^T \text{diag}(\boldsymbol{\upsilon}_i)(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T \text{vec}(\mathbf{s}\mathbf{h}^T), \quad (13)$$

which is a system of M *bilinear* equations. If the support $\mathcal{S}$ is known, the number of unknowns is $L + S$. If it is not, the number is $L + N$ and the constraint $\|\mathbf{s}\|_0 \leq S$ must be added, further complicating the problem. The resultant problem can be handled by an alternating scheme that iterates between optimizing $\mathbf{s}$ given $\mathbf{h}$ and optimizing $\mathbf{h}$ for the new $\mathbf{s}$. More sophisticated approaches include lifting techniques that define the lifted variable $\mathbf{\Sigma} := \mathbf{s}\mathbf{h}^T$, the observation matrix $\mathbf{\Phi}_i := \mathbf{\Psi}^T \text{diag}(\boldsymbol{\upsilon}_i)(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T$, and find an approximation to

$$\min_{\mathbf{\Sigma}} \|\mathbf{z}_{\mathcal{M},i} - \mathbf{C}_{\mathcal{M}} \mathbf{\Phi}_i \text{vec}(\mathbf{\Sigma})\|_2^2 + \gamma_1 \text{rank}(\mathbf{\Sigma}) + \gamma_2 \|\mathbf{\Sigma}\|_{2,0}, \quad (14)$$

where γ_1 and γ_2 are regularization parameters and $\|\mathbf{\Sigma}\|_{2,0}$ is defined as the number of non-zero rows of $\mathbf{\Sigma}$. The estimates $\hat{\mathbf{s}}^{(i)}$ and $\hat{\mathbf{h}}^{(i)}$ are then found the main left and right singular vectors of $\mathbf{\Sigma}^*$, which are subject to an inherent scaling ambiguity. See [14] for further justification and suitable relaxations.

In the case of selection sampling (SS), the expression analogous to (13) is

$$\mathbf{x}_{\mathcal{M}} = \mathbf{C}_{\mathcal{M}} \mathbf{V}(\mathbf{\Psi}^T \odot \mathbf{U}^T)^T \text{vec}(\mathbf{s}\mathbf{h}^T), \quad (15)$$

which is also a bilinear system similar to the previous one. Note that, as in the BGS case (cf. II-A), the role of $\mathbf{V}$ in SS is taken by $\mathbf{\Psi}^T \text{diag}(\boldsymbol{\upsilon}_i)$ in AGSS.

D. Space-shift sampling of diffused sparse signals

In many setups, access to more than one sampling node is available. This is useful to robustify the recovery and reduce the number of required samples per node, which is convenient because the conditioning number of the Vandermonde matrix $\mathbf{\Psi}$ (one of the factors in $\mathbf{\Theta}_i$) worsens as the samples per node increase. The resultant sampling scheme is referred to as space-shift sampling [4]. To particularize it to the setup at hand, define the vectorized version of $\mathbf{Z}$ as $\bar{\mathbf{z}}$ and then the $N^2 \times N$ matrix $\mathbf{\Upsilon} := [\text{diag}(\boldsymbol{\upsilon}_1), \ldots, \text{diag}(\boldsymbol{\upsilon}_N)]^T \text{diag}(\tilde{\mathbf{h}})$. With this conventions, $\bar{\mathbf{z}}$ can be written as $\bar{\mathbf{z}} = (\mathbf{I} \otimes \mathbf{\Psi}^T)\bar{\mathbf{\Upsilon}}\mathbf{U}\mathbf{s}$, where $\otimes$ stands for the Kronecker product. The sampled version in this case is given as $\bar{\mathbf{z}}_{\mathcal{M}} = \bar{\mathbf{C}}_{\mathcal{M}} \bar{\mathbf{z}}$, where $\bar{\mathbf{C}}_{\mathcal{M}}$ is a selection matrix of size $M \times N^2$. The results in the previous sections can be applied to this case as well provided that $\mathbf{\Theta}_i$ and $\mathbf{C}_{\mathcal{M}}$ are replaced with $\bar{\mathbf{\Theta}} := (\mathbf{I} \otimes \mathbf{\Psi}^T)\bar{\mathbf{\Upsilon}}\mathbf{U}$ and $\bar{\mathbf{C}}_{\mathcal{M}}$.

IV. Numerical experiments

Short simulations to illustrate and gain intuition about some of the results presented are shown here.

Test case 1. First, we consider a stochastic block model (SBM) graph with N nodes and B communities with $N_b = N/B$ nodes each [21]. Edges exist with probability $p_{bb} = 0.4$ if the incident nodes are in the same community and with probability $p_{bb'} = 0.1$ if they are not. The remaining parameters are given in the caption of Fig. 1. The index of the seeding nodes is chosen uniformly at random and the seed value is drawn from a zero-mean unit variance Gaussian (ZMUVG). The filter taps have length $L = 6$ and each of them is drawn from a ZMUVG. Fig. 1 depicts the recovery rate, defined as the proportion of simulations for which the seeds are correctly identified and the ℓ_2-norm of the error is less than 0.1, as the number of seeds S increases. All sampling nodes are considered, and the median error is reported. The 10 scenarios (lines) in the figure consider if: 1) $\mathbf{C}_{\mathcal{S}}$ is known or not ("Known"/"Unk"); 2) the sampling node i is in the same community than one of the seeds or is a random node ("Comm"/"Rand"); 3) the sampled signal is either $\mathbf{s}$ or $\mathbf{x}$ ("$\hat{s}$"/"$\hat{x}$"); and 4) the sampling scheme is AGSS or SS ("AGSS"/"SS") . The results confirm that recovery is harder as S increases, that blind schemes are not able to recover the signal if $S > M/2 = 4$, and that knowledge of $\mathbf{C}_{\mathcal{S}}$ notably facilitates the recovery. We also observe that if the other two criteria are fixed, AGSS always outperform SS, confirming that AGSS are more robust and less sensitive to the sampling configuration [4]. Similarly, "Comm" is always better than "Rand". This is not surprising since the (diffused) seed values reach the sampling node faster if the node belongs to the same community. This also explains why

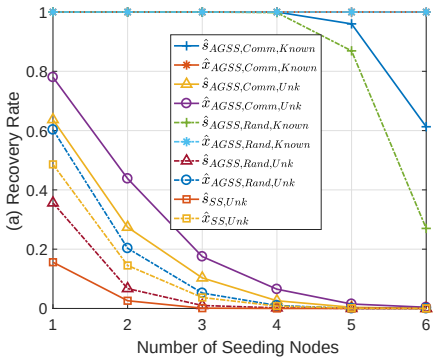

Fig. 1: Recovery rate of DSGS in SBM graphs. Signals are recovered via the ℓ_1-norm relaxation using the Laplacian as the GSO. 500 simulations with different graphs: $N = 50, B = 5, M = 8, \mathbf{R}_{w,\mathcal{M}}^{(i)} = 10^{-5}\mathbf{I}$.

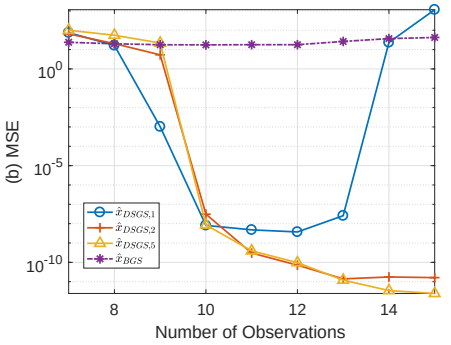

Fig. 2: Median MSE of recovered signals defined over 95 real-world graphs using a blind diffused recovery scheme.

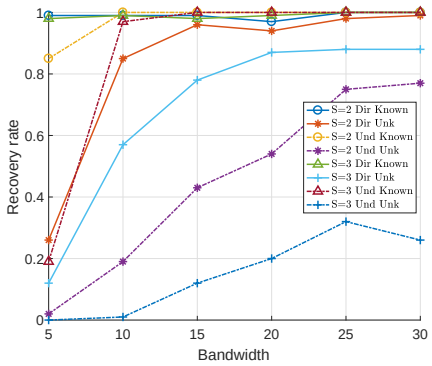

Fig. 3: Recovery rate of DSGS in directed and undirected SBM graphs for varying filter bandwidth length. Signals are recovered via the pseudoinverse (known $\mathbf{C}_{\mathcal{S}}$) and the ℓ_1-norm relaxation (unknown $\mathbf{C}_{\mathcal{S}}$) using the adjacency matrix as the GSO. Directed graphs with non-diagonalizable adjacency matrices are discarded. 100 graph-realizations for each type of graph, selecting the sampling node i from all the N nodes as the one leading to the smallest ℓ_2-norm of $(\mathbf{s}-\hat{\mathbf{s}})$. The remaining parameters are: $N=30, B=2, p_b=0.25, p_{bb'}=0.05, M=4, \mathbf{R}_{w,\mathcal{M}}^{(i)}=10^{-5}\mathbf{I}$

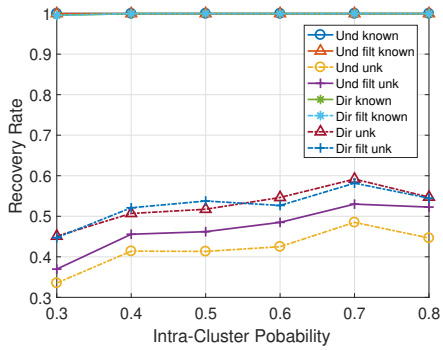

Fig. 4: Recovery rate of sparse and DSGS in directed and undirected SBM graphs for different probabilities of inter-cluster and intra-cluster links. Signals are recovered via the pseudoinverse (known $\mathbf{C}_{\mathcal{S}}$) and the ℓ_1-norm relaxation (unknown $\mathbf{C}_{\mathcal{S}}$) using the adjacency matrix as the GSO. The tested intra-cluster probability are $[0.3, 0.4, 0.5, 0.6, 0.7, 0.8]$, as shown in the horizontal axis. The corresponding inter-cluster probabilities are $[0.1, 0.15, 0.2, 0.25, 0.3]$. For each point in the figure, 100 graph-realizations are considered and, for each of those realizations, the $N=30$ nodes are tested, so that the rates shown correspond to averages across 3000 trials. The remaining parameters are: $B=3, M=9, \mathbf{R}_{w,\mathcal{M}}^{(i)}=10^{-6}\mathbf{I}$.

recovering $\mathbf{x}$ seems to be always easier than recovering (non-diffused) signal $\mathbf{s}$.

Test case 2. Fig. 2 tests our schemes in the D&D protein structure database [22], where nodes account for amino acids, links capture similarity, and signals are the expression level of the amino acids. We assume that the data can be accurately modeled as DSGS and try to recover the full signal following the blind diffused recovery scheme (label "DSGS,1"), its space-shift counterpart with 2 and 5 nodes ("DSGS,2", "DSGS,5"), and AGSS modeling the data not as DSGS but as bandlimited ("BGS"). The median error of all graphs is reported and all sampling nodes are considered, selecting the 25th error percentile. The main observations are: 1) BGS yields the worst performance, pointing out that the DSGS model is a good fit for the information in the D&D database; and 2) for the "DSGS,1" the median MSE increases when the number of observations is high. This stems from the conditioning number of as explained in Sec. III-D. In contrast, the "DSGS,2" and "DSGS,5" schemes are more robust.

Test case 3. We test our schemes in the ETEX dataset [23], which contains $\{\mathbf{y}_t\}_{t=0}^{29}$ graph signals whose nodes correspond to different locations and t represents time. We use as GSO the Adjacency of the geographical graph [24], the seed is set as $\mathbf{s}=\mathbf{y}_0$, and the signal to be sampled and recovered is $\mathbf{x}=\mathbf{y}_t$ for all $t>0$. Using $M=16$ samples and the same approach than in the second test case, we run the experiment for 29 different signals (one per $t, t>0$), obtaining MSE of $3\cdot10^{-5}$ and $1.5\cdot10^{-5}$ for "DSGS,1" and "DSGS,2" respectively.

Test case 4. In Fig. 3 we analyze the impact on the recovery of two factors: a) the bandwidth of the diffusing filter and b) the directivity of the supporting graph. To this end, let us consider a bandpass filter $\tilde{\mathbf{h}}$ whose non-zero band consists of W elements randomly drawn from a ZMUVG. Moreover, we consider two types of SBM graphs: one where links are directed (denoted as "Dir" in the figure) and antoher one with undirected links ("Und"). For this test case, the adjacency matrix is chosen as GSO. The remaining parameters are detailed in the caption of Fig. 3. The plotted results reveal that successful interpolation from DSGS samples is more amenable in directed than undirected graphs. The additional information about the edge direction contained in the GSO, central during both filtering and the AGSS, helps identifying the seeds in $\mathcal{S}$. Furthermore, lower W hinders the diffusion of the seeds, making the recovery harder. Indeed, in the extreme case of $W=0$ the factor $\text{diag}(\boldsymbol{\upsilon}_i\circ\tilde{\mathbf{h}})$ in (12) renders the observations zero.

Test case 5. Fig. 4 studies the impact of the density of the graph on the recoverability of the signals. In this experiment, both the intra-cluster and the inter-cluster probability vary in the same direction, as explained in the caption of the figure. The results show that the recovery rate tends to improve when the graphs are denser. With a higher link probability, the chances that any node is close to the seeds increases, so that they can access the information related with the non-zero elements of the sampled signal. As a result, the fraction of nodes able to recover the signal increases. In addition, the plot confirms that directed graphs ("Dir") have a better recovery rate than undirected graphs ("Und"), which is consistent with the results presented for the test case 3.

REFERENCES

[1] 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)*, May 2014, pp. 3864–3868.

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

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

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

[6] D. I. Shuman, P. Vandergheynst, and P. Frossard, "Distributed signal processing via Chebyshev polynomial approximation," *arXiv preprint arXiv:1111.5239*, 2011.
[7] A. Sandryhaila, S. Kar, and J. Moura, "Finite-time distributed consensus through graph filters," in *IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, May 2014, pp. 1080–1084.
[8] 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.
[9] 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, 2017.
[10] 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, Dec. 2016.
[11] J. Mei and J. M. F. Moura, "Signal processing on graphs: Causal modeling of unstructured data," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2077–2092, Apr. 2017.
[12] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Ieee trans. signal inf. process. netw." *IEEE Trans. Signal Inf. Process. Netw*, vol. 3, no. 3, pp. 467 – 483, Sept. 2017.
[13] D. Ramirez, 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)*, March 2017, pp. 4104–4108.
[14] 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.
[15] A. Sandryhaila and J. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, June 2014.
[16] M. Unser, "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
[17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
[18] F. Pukelsheim, *Optimal Design of Experiments*. SIAM, 1993, vol. 50.
[19] 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.
[20] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
[21] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, vol. 84, no. 6, p. 066106, Dec. 2011.
[22] P. D. Dobson and A. J. Doig, "Distinguishing enzyme structures from non-enzymes without alignments," *Journal of molecular biology*, vol. 330, no. 4, pp. 771–783, 2003.
[23] K. Nodop, R. Connolly, and F. Girardi, "The field campaigns of the european tracer experiment (ETEX): Overview and results," *Atmospheric Environment*, vol. 32, no. 24, pp. 4095–4108, 1998.
[24] D. Thanou, X. Dong, D. Kressner, and P. Frossard, "Learning heat diffusion graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 484–499, 2017.