Aggregation Sampling of Graph Signals in the Presence of Noise

Santiago Segarra, Antonio G. Marques, Geert Leus, and Alejandro Ribeiro

Abstract—A scheme to sample bandlimited graph signals in the presence of noise is analyzed. Samples are aggregated at a single node by successive applications of the so-called graph-shift operator that encodes the local structure of the underlying graph. In contrast to the noiseless case, when noise is present the choice of the sampling node and the local sample-selection scheme play a major role in determining the interpolation error. We provide optimal sampling schemes for particular noise models. We also analyze and provide identifiability conditions for the case where the frequency support of the bandlimited signal is unknown. Finally, simulations with synthetic and real-world graph signals are used to illustrate the behavior of aggregation sampling in noisy scenarios.

Keywords—*Graph signal processing, Sampling, Interpolation, Bandlimited graph signals, Aggregation, Noise.*

I. INTRODUCTION

In recent years, classical signal processing tools, originally conceived to study time-varying signals, have been extended to the domain of signals defined on graphs [2]–[4]. This entails generalizing algorithms currently available for time-varying signals while gaining intuition on what concepts are preserved and lost when a signal is defined not in the classical time grid, but in a more general graph.

A problem that has received significant attention is that of sampling bandlimited graph signals, i.e., signals that have a sparse representation in a frequency domain. Most works implement a *selection* sampling approach, where one has access to the value of the signal at a subset of nodes and wants to recover the signal in the entire graph [5], [6]. Recently, an *aggregation* sampling approach has been proposed, where one has access to the value of the signal at a single node but at different time instants as given by the successive applications of a local graph-shift operator [7]. For the particular case of time-varying signals, the graph-shift operator represents time delays and both the selection and aggregation sampling strategies boil down to classical sampling. This paper studies the performance of aggregation sampling in noisy scenarios. An aggregation sampling scheme is fully specified by two elements: i) the sampling node, and ii) the shifted signals (time instants) observed. In the absence of noise, recovery from a given aggregation sampling scheme is either perfect or infeasible [7]. When noise is present, however, the choice of the sampling node and the sample-selection scheme have a major impact on the interpolation error. In Section III, an interpolator based on the Best Linear Unbiased Estimator (BLUE) is designed and the effect on the corresponding error covariance matrix of different noise models is discussed. In Section IV we assume that the signal to be sampled is bandlimited but of unknown frequency support and provide conditions under which the signal can be identified. This problem falls into the category of sparse signal reconstruction [8], [9] where the main idea is to leverage the structure of the observation matrix to facilitate recovery. Simulations illustrating the behavior of aggregation sampling in the presence of noise and uncertainty in the signal support

for both synthetic and real-world graph signals close the paper. The proofs not included in this paper can be found in an online appendix.[1]

II. PRELIMINARIES

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ denote a *directed* graph. The set of nodes $\mathcal{N}$ has cardinality N, and the set of links $\mathcal{E}$ is such that $(i, j) \in \mathcal{E}$ if and only if node i is connected to node j. The set $\mathcal{N}_i : \{ j \,|\, (j, i) \in \mathcal{E} \}$ contains all nodes with an incoming connection to i and is termed the incoming neighborhood of i. The focus of this paper is not on analyzing $\mathcal{G}$, but a graph signal defined on the set of nodes $\mathcal{N}$. 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 component represents the value of the signal at node i. The graph $\mathcal{G}$ is endowed with a *graph-shift operator* $\mathbf{S}$ defined as an $N \times N$ matrix whose entry (i, j), denoted as S_{ij}, can be nonzero only if $i = j$ or $(j, i) \in \mathcal{E}$. Common choices for $\mathbf{S}$ are the adjacency matrix of the graph [3], [10] and its Laplacian [2]. The intuitive interpretation of $\mathbf{S}$ is that it represents a linear transformation that can be computed locally at the nodes of the graph. We will assume that $\mathbf{S}$ is diagonalizable, i.e., that there exists matrices $\mathbf{V}$ and $\mathbf{\Lambda}$ such that $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{\Lambda}$ is diagonal with elements $\{\lambda_i\}_{i=1}^N$.

A. Aggregation sampling of bandlimited signals

Given a graph signal $\mathbf{x}$, define the l-th shifted signal $\mathbf{y}^{(l)} := \mathbf{S}^l \mathbf{x}$, which can also be written as $\mathbf{y}^{(l)} := \mathbf{S}\mathbf{y}^{(l-1)}$ with $\mathbf{y}^{(0)} = \mathbf{x}$. Note that the i-th entry of $\mathbf{y}^{(l)}$, denoted as $y_i^{(l)}$, can be computed by node i using the following local aggregation $y_i^{(l)} = \sum_{j \in \mathcal{N}_i} S_{ij} y_j^{(l)}$. The shifted signals are used to define the $N \times N$ matrix

$$\mathbf{Y} := [\mathbf{y}^{(0)}, \mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N-1)}] = [\mathbf{x}, \mathbf{S}\mathbf{x}, \ldots, \mathbf{S}^{N-1}\mathbf{x}]. \quad (1)$$

Since the i-th row of $\mathbf{Y}$ is associated with node i, we define the successively aggregated signal at i as $\mathbf{y}_i := (\mathbf{e}_i^T \mathbf{Y})^T = \mathbf{Y}^T \mathbf{e}_i$ where $\mathbf{e}_i$ is the i-th canonical vector (all entries are 0 except the i-th one, which is 1). Aggregation sampling consists in selecting K out of the N elements (rows) of $\mathbf{y}_i$. This is accomplished with a $K \times N$ binary selection matrix $\mathbf{C}$ to obtain $\bar{\mathbf{y}}_i := \mathbf{C}\mathbf{y}_i = \mathbf{C}(\mathbf{Y}^T \mathbf{e}_i)$, where $\mathbf{C}$ has only one 1 per row. We say that the signal $\bar{\mathbf{y}}_i$ samples $\mathbf{x}$ with *successive local aggregations*. Notice that $\mathbf{y}_i$ can be computed locally at node i using successive exchanges with neighboring nodes.

Recovery of the original signal $\mathbf{x}$ from its sampled version $\bar{\mathbf{y}}_i$ is possible under the assumption that $\mathbf{x}$ is bandlimited. More specifically, that $\mathbf{x}$ can be expressed as a linear combination of a *subset* of the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$, or, equivalently, that the vector $\hat{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ is sparse. In this context, vectors $\mathbf{v}_k$ are interpreted as the graph frequency basis and $\hat{x}_k$ as the corresponding signal frequency coefficients. It will be assumed that the active frequencies are the first K ones, associated with the eigenvalues of largest magnitude [5]. Under this assumption, it holds that $\hat{\mathbf{x}} = [\hat{x}_1, \ldots, \hat{x}_K, 0, \ldots, 0]^T$. For convenience, we define $\mathbf{V}_K := [\mathbf{v}_1, \ldots, \mathbf{v}_K]$ and $\hat{\mathbf{x}}_K := [\hat{x}_1, \ldots, \hat{x}_K]^T$ so that we may write $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_K^T, \mathbf{0}]^T$.

Define $\boldsymbol{v}_i := \mathbf{V}^T \mathbf{e}_i$, the $N \times N$ columwise Vandermonde matrix $\mathbf{\Psi}$ with entries $\Psi_{ij} := \lambda_j^{i-1}$, and the $N \times K$ matrix $\mathbf{\Psi}_i := \mathbf{\Psi}\mathrm{diag}(\boldsymbol{v}_i)\mathbf{E}_K$, where $\mathbf{E}_K := [\mathbf{e}_1, \ldots, \mathbf{e}_K]$ is a tall matrix

[1]http://www.seas.upenn.edu/~ssegarra

collecting the K first canonical vectors. Then, the sampled signal $\bar{\mathbf{y}}_i$ can be written in terms of $\hat{\mathbf{x}}$ as [7]

$$\bar{\mathbf{y}}_i = \mathbf{C}\mathbf{y}_i = \mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\hat{\mathbf{x}} = \mathbf{C}\boldsymbol{\Psi}_i\hat{\mathbf{x}}_K. \qquad (2)$$

If matrix $\mathbf{C}\boldsymbol{\Psi}_i$ is invertible, we can interpolate $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{V}_K\hat{\mathbf{x}}_K = \mathbf{V}_K(\mathbf{C}\boldsymbol{\Psi}_i)^{-1}\bar{\mathbf{y}}_i. \qquad (3)$$

Conditions under which $\mathbf{C}\boldsymbol{\Psi}_i$ is invertible have been studied in [7, Prop. 1]. Those basically require the eigenvalues $\{\lambda_i\}$ to be distinct and the entries of $\boldsymbol{v}_i$ to be non-zero for the set of active frequencies. If the conditions are met and there is no noise, (3) can be used to perfectly interpolate $\mathbf{x}$ from the K samples in $\bar{\mathbf{y}}_i$. When noise is present, perfect recovery is in general infeasible, as we study next.

Remark 1: The conditions for $\mathbf{C}\boldsymbol{\Psi}_i$ being invertible given in [7, Prop. 1] leverage the fact of $\mathbf{C}\boldsymbol{\Psi}\mathbf{E}_K$ being the product of a diagonal and a Vandermonde matrix. To guarantee such a property, we define selection matrices of the form $\mathbf{C}_K(n_0, N_0) := [\mathbf{e}_{n_0}, \mathbf{e}_{n_0+N_0}, \ldots, \mathbf{e}_{n_0+(K-1)N_0}]^T$ and the *set* of admissible $\mathbf{C}$ as $\mathcal{C}_K := \{\mathbf{C}_K(n_0, N_0) \mid N_0 = 1, \ldots, N/K$ and $n_0 = 1, \ldots, N-N_0(K-1)\}$.

III. SAMPLING IN THE PRESENCE OF NOISE

Consider the shifted sampled signal $\mathbf{y}_i$ corrupted by additive noise, so that the *observed* signal $\mathbf{z}_i$ is given by $\mathbf{z}_i = \mathbf{y}_i + \mathbf{w}_i$. The noise $\mathbf{w}_i$ is assumed to be zero-mean, independent of the graph signal, and colored with a covariance matrix $\mathbf{R}_w^{(i)} := \mathbb{E}[\mathbf{w}_i\mathbf{w}_i^H]$. For notational convenience, we define also $\bar{\mathbf{w}}_i = \mathbf{C}\mathbf{w}_i$ and $\bar{\mathbf{R}}_w^{(i)} = \mathbf{C}\mathbf{R}_w^{(i)}\mathbf{C}^H$. Hence, the relation between the observed samples $\bar{\mathbf{z}}_i$ and the original signal $\mathbf{x}$ is given by

$$\bar{\mathbf{z}}_i = \mathbf{C}\boldsymbol{\Psi}_i\hat{\mathbf{x}}_K + \bar{\mathbf{w}}_i, \qquad \mathbf{x} = \mathbf{V}_K\hat{\mathbf{x}}_K. \qquad (4)$$

The BLUE estimator of $\hat{\mathbf{x}}_K$, which minimizes the least squares error, is given by [11]

$$\hat{\hat{\mathbf{x}}}_K^{(i)} = \left(\boldsymbol{\Psi}_i^H\mathbf{C}^H(\bar{\mathbf{R}}_w^{(i)})^{-1}\mathbf{C}\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i^H\mathbf{C}^H(\bar{\mathbf{R}}_w^{(i)})^{-1}\bar{\mathbf{z}}_i, \quad (5)$$

provided that the inverse in (5) exists. After obtaining $\hat{\hat{\mathbf{x}}}_K^{(i)}$, the time signal recovered at the i-th node $\hat{\mathbf{x}}^{(i)}$ can be found as

$$\hat{\mathbf{x}}^{(i)} = \mathbf{V}_K\hat{\hat{\mathbf{x}}}_K^{(i)}. \qquad (6)$$

Finally, the error covariance matrices for the frequency and time estimators $\hat{\mathbf{R}}_e^{(i)} := \mathbb{E}[(\hat{\mathbf{x}}_K - \hat{\hat{\mathbf{x}}}_K^{(i)})(\hat{\mathbf{x}}_K - \hat{\hat{\mathbf{x}}}_K^{(i)})^H]$ and $\mathbf{R}_e^{(i)} := \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}^{(i)})(\mathbf{x} - \hat{\mathbf{x}}^{(i)})^H]$ are [11]

$$\hat{\mathbf{R}}_e^{(i)} = \left(\boldsymbol{\Psi}_i^H\mathbf{C}^H(\bar{\mathbf{R}}_w^{(i)})^{-1}\mathbf{C}\boldsymbol{\Psi}_i\right)^{-1}, \qquad (7)$$

$$\mathbf{R}_e^{(i)} = \mathbf{V}_K\hat{\mathbf{R}}_e^{(i)}\mathbf{V}_K^H. \qquad (8)$$

Note that the error covariance matrix $\mathbf{R}_e^{(i)}$ depends on the noise model, the frequencies of the graph (eigenvalues of the shift operator), the node taking the observations, and the sample-selection scheme adopted (cf. Remark 1).

The error covariance matrix can be used to assess the performance of the estimation. However, there exist multiple alternatives to quantify the error. Common approaches include the minimization of the trace of the covariance matrix, which corresponds to the minimizing of the Mean Square Error (MSE), as well as the minimization of its largest eigenvalue, its log determinant and the inverse of the trace of its inverse. We summarize these error metrics as follows

$$e_1 := \text{trace}(\mathbf{R}_e^{(i)}), \qquad e_2 := \lambda_{\max}(\mathbf{R}_e^{(i)}), \qquad (9)$$

$$e_3 := \log\det(\hat{\mathbf{R}}_e^{(i)}), \qquad e_4 := \left[\text{trace}\left(\hat{\mathbf{R}}_e^{(i)-1}\right)\right]^{-1}.$$

Notice that the error metrics e_3 and e_4 are computed based on the error covariance matrix for the frequency estimator $\hat{\mathbf{R}}_e^{(i)}$ instead of the time estimator since $\mathbf{R}_e^{(i)}$ is a singular matrix [cf. (8)].

Expressions (7)-(8) hold for a general noise covariance matrix $\mathbf{R}_w^{(i)}$. We now present three particular noise models of interest: i) White noise in the observed signal $\mathbf{z}_i$. This implies that $\mathbf{w}_i$ is white and therefore $\mathbf{R}_w^{(i)} = \sigma^2\mathbf{I}$, with σ^2 denoting the noise power. Consequently,

$$\bar{\mathbf{R}}_w^{(i)} = \sigma^2\mathbf{I}. \qquad (10)$$

ii) White noise in the original signal $\mathbf{x}$. With $\mathbf{w}$ denoting the white additive noise present in $\mathbf{x}$, we can use the linear observation model to write $\mathbf{w}_i = \boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\mathbf{V}^{-1}\mathbf{w}$. This implies that $\mathbf{R}_w^{(i)} = \sigma^2\boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\mathbf{V}^{-1}(\mathbf{V}^{-1})^H\text{diag}(\boldsymbol{v}_i)^H\boldsymbol{\Psi}^H$ and, when the shift $\mathbf{S}$ is a normal matrix, i.e. $\mathbf{V}^{-1} = \mathbf{V}^H$, this expression reduces to

$$\bar{\mathbf{R}}_w^{(i)} = \sigma^2\mathbf{C}\boldsymbol{\Psi}|\text{diag}(\boldsymbol{v}_i)|^2\boldsymbol{\Psi}^H\mathbf{C}^H. \qquad (11)$$

iii) White noise in the active frequency coefficients $\hat{\mathbf{x}}_K$. With $\hat{\mathbf{w}}_K$ denoting the white additive noise present in $\hat{\mathbf{x}}_K$, we can use the linear observation model to write $\mathbf{w}_i = \boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\mathbf{E}_K\hat{\mathbf{w}}_K = \boldsymbol{\Psi}_i\hat{\mathbf{w}}_K$. This implies that

$$\bar{\mathbf{R}}_w^{(i)} = \sigma^2\mathbf{C}\boldsymbol{\Psi}_i\boldsymbol{\Psi}_i^H\mathbf{C}^H. \qquad (12)$$

This noise model can arise when the signal to be sampled has been previously processed with a low-pass graph filter [10], [12].

A. Selection of the sampling set

The two elements that define the samples in $\bar{\mathbf{y}}_i$ are: the node i that samples (aggregates) the information and the sample-selection scheme specified by $\mathbf{C}$. Here, we discuss how to design these two elements and how that design depends on the properties of $\mathcal{G}$ (spectrum of $\mathbf{S}$) and $\mathbf{x}$ (set of active frequencies).

1) Selection of the sampling node: In a noiseless scenario, the outcome of running aggregation sampling at a specific node i is binary: it leads to either perfect or infeasible recovery [7]. However, when noise is present, the error covariance matrix $\mathbf{R}_e^{(i)}$ is different for each i. Hence, it is reasonable to select as a sampling node one leading to a small error. Selecting the best one requires the computation of N closed-form expressions, which involve matrix inversions. If computational complexity is a limiting factor, the structure of the noise correlation and the interpolation matrix (which depends on the spectrum of $\mathbf{S}$), can be exploited to reduce the computational burden.

For the case where white noise is present in $\hat{\mathbf{x}}_K$, when substituting (12) into (7) and (8), it follows that

$$\hat{\mathbf{R}}_e^{(i)} = \sigma^2\mathbf{I}, \quad \mathbf{R}_e^{(i)} = \sigma^2\mathbf{V}_K\mathbf{V}_K^H. \qquad (13)$$

Consequently, for this particular noise model, the estimator performance is independent of the node choice for every error metric [cf. (9)]. The result is intuitive: given that the noise and the signal are present in the same frequencies, it is irrelevant if a node amplifies or attenuates a particular frequency. By contrast, if the white noise is present in $\bar{\mathbf{y}}_i$, the following result holds.

Proposition 1: *If the noise covariance is given by* (10) *and the selection matrix is* $\mathbf{C} = \mathbf{C}_K(n_0, N_0)$; *then, the sampling node* i^* *that minimizes* e_4 *in* (9) *can be found as*

$$i^* = \arg\max_i \sum_{k=1}^{K} |[\boldsymbol{v}_i]_k|^2 \frac{|\lambda_k|^{2n_0} - |\lambda_k|^{2(n_0+N_0K)}}{1 - |\lambda_k|^{2N_0}}. \qquad (14)$$

The optimal sampling node i^* will be one with large values of $|[\boldsymbol{v}_i]_k|$ for the active frequencies $k \leq K$, i.e. a node that strongly expresses

the active frequencies. The relative importance of frequency k is given by the fraction in (14), which depends on the modulus of the associated eigenvalue and the structure of the selection matrix $\mathbf{C}$ (values for n_0 and N_0).

2) Design of the sample-selection scheme $\mathbf{C}$*:* The error covariance matrix, and hence the different error metrics presented in (9), depend on $\mathbf{C}$. As was the case for the sampling node selection, in some cases the spectral properties of $\mathbf{S}$ as well as the structure of $\mathcal{C}_K$ and the noise covariance, can be exploited to determine the optimal observation strategy. E.g., for the case where white noise is present in $\widehat{\mathbf{x}}_K$, it is immediate to see that the performance is independent of the sample-selection scheme [cf. (13)]. For the case where white noise is present in $\bar{\mathbf{y}}_i$, the following result holds.

Proposition 2: *If the noise covariance is given by* (10)*, the selection matrix is* $\mathbf{C} = \mathbf{C}_K(n_0, N_0)$ *and* N_0 *is fixed; then, the* n_0^* *minimizing* e_3 *in* (9) *is the same for any node* i *and can be found as*

$$n_0^* = \begin{cases} 1 & \text{if } \prod_{k=1}^{K} |\lambda_k|^2 \leq 1, \\ N - N_0(K-1) & \text{otherwise.} \end{cases} \quad (15)$$

The expression for n_0^* shows that if one application of $\mathbf{S}$ has an overall effect of amplification in the active frequencies, then we should aim to apply $\mathbf{S}$ as many times as possible, whereas if the opposite is true, we should avoid its application. The problem in (15) can be solved for different values of N_0 to get the overall optimal (n_0^*, N_0^*) pair.

One can also look at $\mathbf{C} \notin \mathcal{C}_K$. In that case, the problem can be formulated as a binary optimization over $\mathbf{C}$, which is typically challenging. If the size of the space search (N choose K) is not too large, the problem can be solved by exhaustive search. For more general cases, a reasonable approach is to formulate the problem, relax it, and exploit its structure to find a good approximate solution. Although of interest, developing approximate algorithms to design $\mathbf{C} \notin \mathcal{C}_K$ is out of the scope of this paper and is left as future work.

IV. IDENTIFYING THE SUPPORT OF THE GRAPH SIGNAL

In the previous sections, it has been assumed that the frequency support of $\widehat{\mathbf{x}}$ corresponded to the K principal eigenvectors. A related but more challenging problem is to design the sampling and interpolation procedures when the frequency support $\mathcal{K}$ is not known.

A. Noiseless joint recovery and support identification

Consider aggregation sampling in a noiseless scenario, where we know that $\widehat{\mathbf{x}}$ is K-sparse but we do not know the support of the K nonzero entries. Then, we may recover $\widehat{\mathbf{x}}$ by solving

$$\widehat{\mathbf{x}}^* := \arg\min_{\widehat{\mathbf{x}}} \quad ||\widehat{\mathbf{x}}||_0 \quad (16)$$
$$\text{s.t.} \quad \mathbf{C}\mathbf{y}_i = \mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\widehat{\mathbf{x}},$$

where $|| \cdot ||_0$ denotes the 0-norm, and the constraint follows from the relationship between the original and the observed signal [cf. (2)]. Problem (16) indeed recovers $\widehat{\mathbf{x}}$ when the following conditions hold.

Proposition 3: *Let* $\mathbf{x}$ *and* $\mathbf{C}$ *be, respectively, a bandlimited graph signal with at most* K *non-zero frequency components and a selection matrix with* $2K$ *rows of the form* $\mathbf{C} = \mathbf{C}_{2K}(n_0, N_0)$*. Then, if all the entries in* $\boldsymbol{v}_i$ *are non-zero and all the eigenvalues of* $\mathbf{S}$ *are non-zero and satisfy that* $\lambda_k^{N_0} \neq \lambda_{k'}^{N_0}$ *for all* $k \neq k'$*, it holds that*
i) the solution to (16) *is unique; and*
ii) the original graph signal can be recovered as $\mathbf{x} = \mathbf{V}\widehat{\mathbf{x}}^*$*.*

The conditions for joint recovery and identification support in Proposition 3 leverage the fact that $\boldsymbol{\Psi}$ is a Vandermonde matrix,

which is a distinct feature of the aggregation sampling scheme. From a computational perspective, the presence of the 0-norm in (16) renders the optimization non-convex, thus challenging to solve. A straightforward way to convexify it is to replace the 0-norm with a 1-norm. Note that if such a process finds a feasible solution, call it $\widehat{\mathbf{x}}_1^*$, such that $||\widehat{\mathbf{x}}_1^*||_0 = K$, then it holds that $\widehat{\mathbf{x}}^* = \widehat{\mathbf{x}}_1^*$. Conditions under which this process is guaranteed to identify the support can be found by analyzing either the coherence or the restricted isometry property (RIP) of the matrix $\mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)$ [8]. The former is easier to find since it depends on the most similar pair of eigenvalues of $\mathbf{S}$. However, the sparsity bound given by the matrix coherence is oftentimes too restrictive [8].

B. Noisy joint recovery and support identification

If noise is present *and* the frequency support of the signal is unknown, the (K-sparse) least squares estimate of $\widehat{\mathbf{x}}$ can be found as the solution to the following optimization problem

$$\widehat{\mathbf{x}}^* := \arg\min_{\widehat{\mathbf{x}}} ||(\bar{\mathbf{R}}_w^{(i)})^{-1/2}(\mathbf{C}\mathbf{y}_i - \mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\widehat{\mathbf{x}})||_2^2 \quad (17)$$
$$\text{s.t.} \quad ||\widehat{\mathbf{x}}||_0 \leq K$$

where the matrix multiplication $(\bar{\mathbf{R}}_w^{(i)})^{-1/2}$ in the objective accounts for the fact that the noise is colored. As in the noiseless case, an approach to convexify (17) is to replace the 0-norm with the 1-norm and solve the problem

$$\widehat{\mathbf{x}}_1^* := \arg\min_{\widehat{\mathbf{x}}} ||(\bar{\mathbf{R}}_w^{(i)})^{-1/2}(\mathbf{C}\mathbf{y}_i - \mathbf{C}\boldsymbol{\Psi}\text{diag}(\boldsymbol{v}_i)\widehat{\mathbf{x}})||_2^2 + \gamma||\widehat{\mathbf{x}}||_1,$$
$$(18)$$

for different values of the parameter γ.

V. NUMERICAL EXPERIMENTS

A. Recovery when the frequency support in unknown

Consider a 20-node undirected graph $\mathcal{G}$ generated as an Erdõs-Rényi graph with edge probability 0.20 [13]. With $\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}_A\mathbf{V}^H$ denoting its adjacency matrix, three different graph-shift operators are considered: $\mathbf{S}_1 = \mathbf{A}$, $\mathbf{S}_2 = \mathbf{I} - \mathbf{A}$, and $\mathbf{S}_3 = 0.5\mathbf{A}^2$. Notice that, even though the support of $\mathbf{S}_3$ differs from that of $\mathbf{S}_1$ and $\mathbf{S}_2$, the shift $\mathbf{S}_3$ still preserves the notion of locality as defined by a two-hop neighborhood. Note also that the three shift operators share the same set of eigenvectors $\mathbf{V}$, but they have a different set of eigenvalues.

Let $\mathbf{x}$ be a 3-bandlimited graph signal of unknown support, i.e., we know that $\widehat{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ contains $K = 3$ nonzero components, but we do not know the indices of these K active frequencies. In this case, $2K = 6$ samples are needed to guarantee identifiability and, provided that the conditions in Proposition 3 are satisfied, the signal can be recovered by solving problem (16) using any of the three shift operators. However, when solving a relaxed version of problem (16), accurate signal recovery depends on the specific graph, signal and node selected for reconstruction. Moreover, the recovery rate depends on the choice of the graph-shift operator $\mathbf{S}$. To assess recovery better, Fig. 1a plots the success rate – fraction of realizations for which the actual signal was recovered – for graph-shifts $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$, and different number of observations. Each point in the plots represents an average across all nodes in the graph, 5 signal realizations and 10 random graph realizations. The recovery rate for $\mathbf{S}_3 = 0.5\mathbf{A}^2$ is consistently higher than that for the other two shift operators. This could be explained because when squaring the adjacency matrix to generate $\mathbf{S}_3$, the dissimilarity between any pair of eigenvalues is increased, which reduces the matrix coherence associated with $\mathbf{S}_3$ and facilitates sparse recovery (cf. last paragraph in Section IV-A).
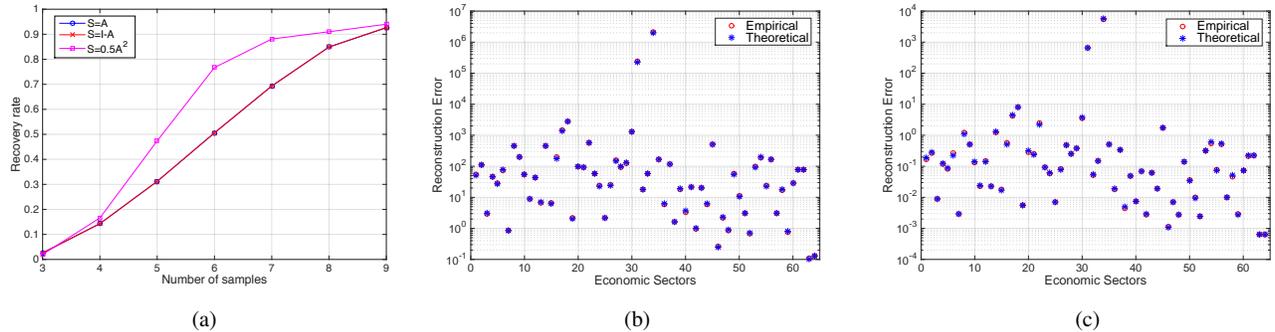
Fig. 1: (a) Recovery rate of bandlimited signals with unknown support as a function of the number of samples. (b) Empirical and theoretical reconstruction errors for different sampling nodes when white noise is added to the observed signal. (c) Empirical and theoretical reconstruction errors when white noise is added directly to the signal $\mathbf{x}_4$.

B. Recovery in the presence of noise

Consider a 64-node graph representing sectors of the U.S. economy and the shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ such that $\bar{\mathbf{S}}_{ij} = (U(i,j) + U(j,i))/2$, where $U(i,i')$ represents how much of the production of sector i (in dollars) was used as an input of sector i' during years 2008, 2009, and 2010 [14]. The 64 nodes consist of 62 industrial sectors plus two artificial sectors, namely: the added value generation (AV) and the market of final users (FU). Consider the signal $\mathbf{x}$ on the mentioned graph containing the total production of each sector (including AV and FU) during year 2011. Since signal $\mathbf{x}$ is approximately bandlimited, we define $\mathbf{x}_4 = \mathbf{V}_4\hat{\mathbf{x}}_4$ as the approximation of $\mathbf{x}$ obtained by keeping the first 4 frequency coefficients.

We perform aggregation sampling at different economic sectors (nodes) of multiple noisy versions of $\mathbf{x}_4$. We first consider adding noise to the observed signal as described in (10). In Fig. 1b we plot the empirical average reconstruction error at different nodes across 1,000 noisy realizations of $\mathbf{x}_4$ and compare it with the theoretical average error, i.e., the trace of $\mathbf{R}_e^{(i)}$ in (8) [cf. e_1 in (9)]. We first observe that the computed theoretical error indeed coincides with the average empirical error across realizations. Moreover, notice that the reconstruction performance is highly node dependent. The error is minimized for the reconstruction based on the artificial sectors AV and FU. This is reasonable since these two nodes – unlike other sectors – are closely related to every other sector of the economy. Furthermore, the sectors achieving the worst reconstruction errors are 'Publishing Industries' and 'Ground Passenger Transportation' corresponding to positions 34 and 31. This can be explained by observing that vectors $\bar{\mathbf{v}}_{34} = \mathbf{v}_{34}\mathbf{E}_4$ and $\bar{\mathbf{v}}_{31}$ contain elements very close to zero, increasing the sensitivity of the reconstruction in the presence of noise.

We then investigate the reconstruction performance when the noise is added to $\mathbf{x}_4$, following the model described in (11) for a fixed σ^2. As was the case in the previous section, the average empirical error (across 1,000 realizations) matches closely our theoretical estimates; see Fig. 1c. Moreover, the specific nodes that lead to a good (bad) interpolation performance are very similar to those in the previous noise model. Indeed, sectors 34 and 31 lead to the highest errors whereas AV and FU attain the best reconstructions.

Finally, we consider adding white noise only to the 4 active frequencies, as described in (12). The empirical reconstruction error associated with each node – averaged over 1,000 realizations – is the same. This validates the analysis in (13), which stated that, for this noise model, the quality of the reconstruction is node independent.

VI. Conclusion

The behavior of aggregation sampling in the presence of noise was analyzed. Optimal choices for the sampling node and the local sample-selection scheme were discussed, since these have a major impact on the interpolation error. Moreover, the interpolation performance for graph signals of unknown support was studied, and identifiability conditions were described.

References

[1] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *arXiv preprint arXiv:1504.04687*, 2015.

[2] 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, May 2013.

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

[4] X. Zhu and M. Rabbat, "Approximating signals supported on graphs," in *Proc. of IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2012, pp. 3921–3924.

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

[6] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *arXiv preprint arXiv:1503.05432*, 2015.

[7] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Sampling of graph signals: successive local aggregations at a single node," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2015.

[8] 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.

[9] X. Zhu and M. Rabbat, "Graph spectral compressed sensing for sensor networks," in *Proc. of IEEE Intl. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2012, pp. 2865–2868.

[10] 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.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

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

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

[14] Bureau of Economic Analysis, "Input-output accounts: the use of commodities by industries before redefinitions," *U.S. Dept. of Commerce*, 2011, http://www.bea.gov/iTable/index_industry.cfm.