Rethinking Sketching as Sampling: Efficient Approximate Solution to Linear Inverse Problems

Fernando Gama[†], Antonio G. Marques[‡], Gonzalo Mateos[*] and Alejandro Ribeiro[†]

***Abstract*—Sampling and reconstruction of bandlimited graph signals have well-appreciated merits for dimensionality reduction, affordable storage, and online processing of streaming network data. However, these parsimonious signals are oftentimes encountered with high-dimensional linear inverse problems. Hence, interest shifts from reconstructing the signal itself towards instead approximating the input to a prescribed linear operator efficiently. In this context, we propose a novel sampling scheme that leverages the bandlimitedness of the output as well as the transformation whose input we wish to approximate. We formulate problems to jointly optimize sample selection and a sketch of the target inverse mapping, so when the latter is affordably applied to the sampled output signal, the result is close to the desired input. The developed sampling plus reduced-complexity processing pipeline is particularly useful for streaming data, where the linear transform has to be applied fast and repeatedly to successive response signals.**

***Index Terms*—Sketching, sampling, graph signal processing, streaming, linear inverse problems**

I. Introduction

Graph Signal Processing (GSP) has emerged as a field that generalizes traditional signal processing tools to operate on signals defined on irregular domains that can be conveniently represented as a graph [1], [2]. Noteworthy advances in the context of linear problems include sampling and reconstruction of bandlimited graph signals, linear shift-invariant graph filtering and computation of the graph Fourier transform (GFT), just to name a few [3], [4]. Traditionally, linear models and signal representations have played a central role in science and engineering mostly because of their conceptual simplicity and mathematical tractability. However, as signals become increasingly high dimensional even linear transformations can be challenging to implement. This is particularly notorious in streaming contexts, where say this massive linear operator has to be repeatedly and efficiently applied to a sequence of signals [5].

These Big Data challenges motivated a recent body of work collectively addressing so-termed sketching problems [6], which seek computationally-efficient solutions to a subset of (typically inverse) linear problems. In a nutshell, the idea is to *draw a sketch* of the linear model such that the resulting linear transform is lower dimensional, while still offering quantifiable approximation error guarantees. The procedure consists in designing a fat random projection matrix to pre-multiply and reduce the dimensionality of the linear operator matrix, so that the resulting *matrix sketch* captures the quintessential structure of the model. Likewise, the input signal vector is adapted to this smaller *sketched* linear operator by performing the same random projection, often ignoring any additional information about the signal model such as its statistical structure. A different alternative to look at the problem is by exploiting results in the context of sampling. Sampling is, arguably, the simplest dimensionality-reduction tool. The classical approach for time-varying signals is to assume that the original signal is bandlimited and then use as information only a subset of its values. Under the GSP framework, the concept of bandlimitedness can be generalized to graph signals living on a low-dimensional subspace, not necessarily given by the Fourier basis. Using this interpretation, several sampling schemes have been recently proposed [7]–[10]. Sampling is particularly attractive in streaming contexts since selecting a subset of values of a signal entails no computational cost. In contrast, the cost of sketching each new signal using random projections is often nontrivial. Another advantage of sampling schemes is the exploitation of a priori knowledge of the signal (in particular, its sparse representation in the frequency domain), whereas the design of the random projection ignores any structural information on the input. On the other hand, sampling methods typically focus on reconstruction of the signal, without accounting for later processing the signal may be subject to; see [11], [12] for exceptions, the latter dealing with graph signals.

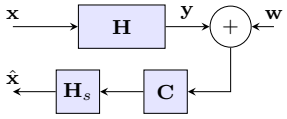

Fig. 1. Knowing the linear transformation $\mathbf{H}$ and having access to a stream of noisy measurements of the output, we want to design the sampling matrix $\mathbf{C}$ and the reduced linear transformation (sketch) $\mathbf{H}_s$ to form an estimator of the input $\hat{\mathbf{x}} = \mathbf{H}_s\mathbf{C}\mathbf{y}$, such that the predicted response $\mathbf{H}\hat{\mathbf{x}}$ approximates $\mathbf{y}$.

The contribution in this paper is to bring together the aforementioned two perspectives, proposing a low-complexity solution to high-dimensional linear inverse problems. To that end, we combine a sampling scheme for the observed output with a matrix sketch for the inverse transformation. To be rigorous, consider the scheme in Fig. 1. The postulated sketching problem is formulated as a joint design of a sampling matrix $\mathbf{C}$ and a matrix sketch $\mathbf{H}_s$ that operate on measurements $(\mathbf{y} + \mathbf{w})$ with the objective of providing an estimate $\hat{\mathbf{x}}$ of the input $\mathbf{x}$. By exploiting the graph-bandlimited structure of the output $\mathbf{y}$ (Section II), the online computational cost can be significantly reduced by just operating with $\mathbf{H}_s$ over a subset of values of the output $\mathbf{C}(\mathbf{y} + \mathbf{w})$. The joint design of the optimal selection scheme and matrix sketch can be solved off line. To gain insights, Section III-A presents first the problem formulation for the case where $\mathbf{w} = \mathbf{0}$, and then Section III-B addresses the general case where observation noise is present. In Section III-C suboptimal solutions with reduced complexity are discussed. Finally, in Section IV the proposed joint design is used to compute efficiently the frequency components of a graph signal. Conclusions are drawn in Section V.

Work in this paper is supported by the NSF award CCF-1217963 and Spanish MINECO TEC2013-41604-R. [†]Dept. of Electrical and Systems Eng., Univ. of Pennsylvania. [‡]Dept. of Signal Theory and Comms., King Juan Carlos Univ. [*]Dept. of Electrical and Comp. Eng., Univ. of Rochester. Emails: {fgama,aribeiro}@seas.upenn.edu, antonio.garcia.marques@urjc.es, gmateosb@ece.rochester.edu

II. Preliminaries

Let $N = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ be a network (graph) described by a set of n nodes $\mathcal{V}$, a set $\mathcal{E}$ of edges (i, j) and a weight function $\mathcal{W} : \mathcal{E} \to \mathbb{R}$ that gives weights to the directed edges. A graph signal $\mathbf{x} \in \mathbb{R}^n$ is defined on the nodes of such a network where each element of the vector $\mathbf{x} = [x_1, \ldots, x_n]^T$ represents a real value present at the node [1], [2]. A graph-shift operator $\mathbf{S} \in \mathbb{R}^{n \times n}$ is introduced in order to describe the impact of the structure of the network on the signal [13]. Matrix $\mathbf{S}$ is such that $[\mathbf{S}]_{i,j} = 0$ whenever $i \neq j$ and $(j, i) \notin \mathcal{E}$. The focus will be set on normal shifts, so that $\mathbf{S}$ can be decomposed as

$$\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H := \mathbf{V}\text{diag}\left([\lambda_1, \ldots, \lambda_n]^T\right)\mathbf{V}^H. \tag{1}$$

where the unitary matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{C}^{n \times n}$ contains the eigenvectors of $\mathbf{S}$ and the diagonal matrix $\mathbf{\Lambda} = \text{diag}\left([\lambda_1, \ldots, \lambda_n]^T\right) \in \mathbb{C}^{n \times n}$ the corresponding eigenvalues. Examples of commonly used normal shift operators include the Laplacian or the adjacency matrix of symmetric graphs [4], [14], the adjacency of a directed cycle for time signals, and the correlation or precision (inverse covariance) matrix of processes following a graphical model [15].

The eigendecomposition in (1) is used to define the Graph Fourier Transform (GFT) and the inverse GFT (iGFT) as $\tilde{\mathbf{x}} := \mathbf{V}^H\mathbf{x}$ and $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$, respectively [4]. Vector $\tilde{\mathbf{x}} = [\tilde{x}_1, \ldots, \tilde{x}_n]^T$ contains the frequency coefficients of $\mathbf{x}$. A key assumption made throughout this paper is that $\mathbf{x}$ is k-bandlimited on $\mathbf{S}$. This implies that there exists a constant $k \ll n$ such that $\tilde{x}_l = 0$ for all $l > k$. Then, the GFT vector $\tilde{\mathbf{x}}$ can be rewritten as $\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_k; \mathbf{0}_{n-k}]$, where $\tilde{\mathbf{x}}_k := [\tilde{x}_1, \ldots, \tilde{x}_k]^T$ contains the first k elements of $\tilde{\mathbf{x}}$. Collecting the k eigenvectors associated with the active frequencies in the matrix $\mathbf{V}_k := [\mathbf{v}_1, \ldots, \mathbf{v}_k] \in \mathbb{C}^{n \times k}$, the GFT-iGFT pairs can be rewritten as

$$\tilde{\mathbf{x}}_k = \mathbf{V}_k^H\mathbf{x}, \qquad \mathbf{x} = \mathbf{V}_k\tilde{\mathbf{x}}_k. \tag{2}$$

Assuming also $\mathbf{x}$ is a realization of a zero-mean random process, we specify next its covariance matrix $\mathbf{R}_x := \mathbb{E}[\mathbf{x}\mathbf{x}^T] \in \mathbb{R}^{n \times n}$. Since the realizations are bandlimited, we consider first the so-called frequency template $\mathbf{T} \in \mathbb{C}^{k \times k}$ defined as

$$\mathbf{T} \;:=\; \mathbb{E}\left[\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^H\right]. \tag{3}$$

This results in a singular positive semidefinite covariance matrix $\mathbf{R}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{V}_k\mathbb{E}[\tilde{\mathbf{x}}_k\tilde{\mathbf{x}}_k^H]\mathbf{V}_k^H = \mathbf{V}_k\mathbf{T}\mathbf{V}_k^H$ with rank k. If no information is available other than $\mathbf{x}$ being bandlimited, a reasonable choice is to set $\mathbf{T} = \mathbf{I}$ so that $\mathbf{R}_x = \mathbf{V}_k\mathbf{V}_k^H$. This type of spectral templates appear when $\mathbf{x}$ is *graph stationary with respect to* the shift operator $\mathbf{S}$ [15].

If the signal $\mathbf{x}$ is k-bandlimited on $\mathbf{S}$ and the active frequencies are known, then the k values in $\tilde{\mathbf{x}}_k$ suffice to describe the n values in $\mathbf{x}$ [cf. (2)]. Hence, it is possible to recover the signal $\mathbf{x}$ by taking $p \geq k$ properly selected samples. In this context, different sampling schemes for graph signals have been proposed [8], [9]. Herein we consider the so-called selection sampling method [8], by which samples correspond to values of the signal $\mathbf{x}$ at a particular set of $p \geq k$ nodes. Specifically, the goal is to design a selection matrix $\mathbf{C} \in \mathcal{C}_{pn}$, where

$$\mathcal{C}_{pn} := \{\mathbf{C} \in \mathbb{R}^{p \times n} : \mathbf{C}\mathbf{1}_n = \mathbf{1}_n,\ \mathbf{C}^T\mathbf{1}_n \preceq \mathbf{1}_n,\ C_{ij} \in \{0, 1\}\},$$

so that samples $\mathbf{x}_s = \mathbf{C}\mathbf{x}$ contain enough information to accurately recover $\mathbf{x}$ via suitable interpolation. If samples of $\mathbf{x}$ can be acquired perfectly in the absence of noise, then any $\mathbf{C}$ for which $\mathbf{C}\mathbf{V}_k$ is nonsingular yields perfect recovery since

$$\mathbf{x} = \mathbf{V}_k(\mathbf{C}\mathbf{V}_k)^{-1}\mathbf{x}_s. \tag{4}$$

When samples are noisy, then there are several methods and algorithms for obtaining the optimal selection matrix $\mathbf{C}$; see [8], [9]. Working with $\mathbf{x}_s$ in lieu of $\mathbf{x}$ offers several advantages. Of particular interest here are the computational savings of processing $\mathbf{x}_s$ rather than $\mathbf{x}$, especially when $p \ll n$. This computational saving is particularly notorious when we realize that computing the least squares (LS) solution [16] entails $\mathcal{O}(mn^2)$ operations. Even through the use of sketching techniques (for each output measurement) this cost can be reduced to $\mathcal{O}(pn^2)$ plus the number of operations required to carry out the random projection [6]. By shifting this burden to the off-line design phase, we are left with only $\mathcal{O}(pn)$ online operations.

III. Sketching as Graph Signal Sampling

The problem of linear sketching is portrayed in Fig. 1, where the input $\mathbf{x} \in \mathbb{R}^n$ to a linear operator $\mathbf{H} \in \mathbb{R}^{m \times n}$ results in an output $\mathbf{y} \in \mathbb{R}^m$, that is $\mathbf{y} = \mathbf{H}\mathbf{x}$. We consider the output $\mathbf{y}$ to be a k-bandlimited graph signal corrupted by additive noise $\mathbf{w} \in \mathbb{R}^m$, and we further assume that a stream of outputs $\mathbf{z} := \mathbf{y} + \mathbf{w}$ is available. In this context, the objective is to estimate the input for each measurement of the output, given knowledge of the linear transformation $\mathbf{H} \in \mathbb{R}^{m \times n}$, where $m \geq n$ and both dimensions are large. To that end we wish to design a fixed sampling scheme $\mathbf{C} \in \mathcal{C}_{pm}$ and a fixed sketch $\mathbf{H}_s \in \mathbb{R}^{n \times p}$, such that with $p \leq m$ the signal $\mathbf{H}\hat{\mathbf{x}} := \mathbf{H}_s\mathbf{z}_s = \mathbf{H}_s\mathbf{C}(\mathbf{y} + \mathbf{w})$ resembles $\mathbf{H}\mathbf{x}$. The design is to be performed off line and applicable to all output measurements in the stream, assuming that these are realizations of a stationary graph signal process whose frequency template $\mathbf{T}$ (hence its covariance $\mathbf{R}_y$) is known. The noise $\mathbf{w}$ is zero mean, uncorrelated with respect to $\mathbf{y}$ and with known positive-definite covariance $\mathbf{\Sigma}_w = \mathbb{E}[\mathbf{w}\mathbf{w}^T] \succ \mathbf{0}$.

The design of $\mathbf{C}$ and $\mathbf{H}_s$ is performed jointly as the solution of the following minimization

$$\{\mathbf{C}^*, \mathbf{H}_s^*\} := \underset{\mathbf{C} \in \mathcal{C}_{pm}, \mathbf{H}_s}{\text{argmin}}\ \mathbb{E}\left[\left\|\mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{y} + \mathbf{w}) - \mathbf{y}\right\|_2^2\right]. \tag{5}$$

Observe that since (5) minimizes an ensemble mean-squared error (MSE) criterion of a linear function of $\mathbf{y}$, finding the optimal solution only requires knowledge of second-order statistics of $\mathbf{y}$. Hence, (5) can be solved off line, yielding an optimal sketch $\mathbf{H}_s^*$ and selection matrix $\mathbf{C}^*$ that will be the same for all the inputs under the stationarity assumption. As a result, during the online phase one must calculate $\mathbf{H}_s^*\mathbf{C}^*(\mathbf{y} + \mathbf{w})$ instead of solving the full problem, reducing the long-run computational cost.

We first look at the optimal joint design when noise is not present (Section III-A) and then address its noisy counterpart (Section III-B). A number of alternatives to solve approximately the resultant optimizations are outlined in Section III-C.

A. Noiseless case

First, consider the case where $\mathbf{w} = \mathbf{0}$ in Fig. 1. In this noiseless scenario, the reduced-complexity estimate is given by $\hat{\mathbf{x}} = \mathbf{H}_s\mathbf{C}\mathbf{y}$.

Proposition 1 *Let $\mathbf{y} \in \mathbb{R}^m$ be a k-bandlimited signal with known spectral template $\mathbf{T} \in \mathbb{C}^{k\times k}$ and let $\mathbf{H} \in \mathbb{R}^{m\times n}$ be a linear transformation. Let $\mathbf{H}_s \in \mathbb{R}^{n\times p}$ be a reduced-input dimensionality sketch, $p \leq m$ and $\mathbf{C} \in \mathcal{C}_{pm}$ be a selection matrix. If $p = k$ and $\mathbf{C}^*$ is designed such that* $\text{rank}\{\mathbf{C}^*\mathbf{V}_k\} = p = k$*, then $\hat{\mathbf{x}} = \mathbf{H}_s^*\mathbf{C}^*\mathbf{y}$ yields the LS estimate $\hat{\mathbf{x}}_{LS} = \mathbf{A}_{LS}\mathbf{y} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$ provided that the sketch $\mathbf{H}_s^*$ is given by*

$$\mathbf{H}_s^* = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{V}_k(\mathbf{C}^*\mathbf{V}_k)^{-1}. \tag{6}$$

Proof: The MSE criterion [cf. (5)] is

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2\right] &= \mathbb{E}\left[\|\mathbf{H}(\mathbf{H}_s\mathbf{C}\mathbf{y})-\mathbf{y}\|_2^2\right] \\ &= \text{tr}\left[\mathbf{H}\mathbf{H}_s\mathbf{C}\mathbf{R}_y\mathbf{C}^T\mathbf{H}_s^T\mathbf{H}^T - 2\mathbf{H}\mathbf{H}_s\mathbf{C}\mathbf{R}_y + \mathbf{R}_y\right].\end{aligned} \tag{7}$$

Observe that because $\mathbf{C} \in \mathcal{C}_{pm}$ then $\text{rank}\{\mathbf{C}\} = p$. Now, optimizing (7) over $\mathbf{H}_s$ first and recalling that $\mathbf{R}_y = \mathbf{V}_k\mathbf{T}\mathbf{V}_k^H$, results in

$$\mathbf{H}_s^*\mathbf{C}\mathbf{V}_k\mathbf{T}\mathbf{V}_k^H\mathbf{C}^T = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{V}_k\mathbf{T}\mathbf{V}_k^H\mathbf{C}^T. \tag{8}$$

Now, if we set $p = k$ and choose $\mathbf{C}$ such that $\text{rank}\{\mathbf{C}\mathbf{V}_k\} = p = k$, then $(\mathbf{V}_k^H\mathbf{C}^T\mathbf{C}\mathbf{V}_k)^{-1}$ exists. Thus, by post multiplying both sides of (8) by the nonsingular $p \times p$ matrix $\mathbf{C}\mathbf{V}_k(\mathbf{V}_k^H\mathbf{C}^T\mathbf{C}\mathbf{V}_k)^{-1}\mathbf{T}^{-1}$, one arrives at

$$\mathbf{H}_s^*\mathbf{C}\mathbf{V}_k = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{V}_k.$$

Finally, because $\text{rank}\{\mathbf{C}\mathbf{V}_k\} = p = k$, then $(\mathbf{C}\mathbf{V}_k)^{-1}$ exists, so we obtain the closed-form solution for $\mathbf{H}_s^*$ given by (6). ■

All in all, in the absence of noise it suffices to first set $p = k$ to find a selection matrix $\mathbf{C} \in \mathcal{C}_{pm}$ such that $\text{rank}\{\mathbf{C}\mathbf{V}_k\} = p = k$, and then obtain $\mathbf{H}_s$ as per Proposition 1. This ensures that $\hat{\mathbf{x}}$ equals the LS solution using p samples of $\mathbf{y}$ via $\hat{\mathbf{x}} = \mathbf{H}_s\mathbf{C}\mathbf{y}$. Also note that $\text{rank}\{\mathbf{C}\mathbf{V}_k\} = k$ is the same condition for exact recovery with selection sampling [8]. This is expected, since in the noiseless case here the design of $\mathbf{C}$ decouples from that of $\mathbf{H}_s$. As a result, existing methods to determine the most informative nodes in sampling scenarios are also applicable here [12].

B. Noisy case

In the case when there is additive noise $\mathbf{w}$ present in the model, the noise statistics have to be taken into account in the minimization of (5). Thus, observe first that if $\mathbf{C} \in \mathcal{C}_{pm}$ is a $p \times m$ selection matrix, then $\mathbf{C}\mathbf{C}^T = \mathbf{I}_p$ is the identity matrix of size $p \times p$. Moreover, $\mathbf{C}^T\mathbf{C} = \text{diag}(\mathbf{c})$ where $\mathbf{c} \in \{0,1\}^m$ is a sampling vector containing p ones, located in the places corresponding to the nodes to be sampled. Now, define $\bar{\mathbf{C}}_\alpha := \text{diag}(\mathbf{c})/\alpha$ as the rescaled sampling vector in matrix form, where $\alpha > 0$ is the rescaling parameter. Observe that if $\mathbf{H}$ is full rank, then the matrix $(\mathbf{H}^T\mathbf{H})$ is nonsingular [17]. Then, we can define the auxiliary matrices $\mathbf{B} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}_y \in \mathbb{R}^{m\times m}$ and $\bar{\mathbf{\Sigma}}_\alpha := \mathbf{R}_y + \mathbf{\Sigma}_w - \alpha\mathbf{I}_m \in \mathbb{R}^{m\times m}$.

With these definitions, the following proposition shows that (5) is equivalent to a mixed-binary optimization problem, with linear objective and linear matrix inequality (LMI) constraints.

Proposition 2 *The solution of* (5) *is given by $\mathbf{C}^*$ and $\mathbf{H}_s^* = \mathbf{H}_s^*(\mathbf{C}^*)$, where*

$$\mathbf{H}_s^*(\mathbf{C}) = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}_y\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_y+\mathbf{\Sigma}_w)\mathbf{C}^T\right)^{-1} \tag{9}$$

and $\mathbf{C}^*$ can be obtained as the solution to the problem

$$\min_{\mathbf{C}\in\mathcal{C}_{pm}} \text{tr}\left[\mathbf{R}_y - \mathbf{B}\mathbf{C}^T\left(\mathbf{C}(\mathbf{R}_y+\mathbf{\Sigma}_w)\mathbf{C}^T\right)^{-1}\mathbf{C}\mathbf{B}^T\right]. \tag{10}$$

Furthermore, (10) *is equivalent to*

$$\begin{aligned}\min_{\mathbf{c}\in\{0,1\}^m,\mathbf{Y}} \ & \text{tr}\left[\mathbf{Y}\right] \\ \text{s.t.} \ & \bar{\mathbf{C}}_\alpha = \text{diag}(\mathbf{c})/\alpha \ , \ \mathbf{c}^T\mathbf{1}_m = p \\ & \begin{bmatrix}\mathbf{Y}-\mathbf{R}_y+\mathbf{B}\bar{\mathbf{C}}_\alpha\mathbf{B}^T & \mathbf{B}\bar{\mathbf{C}}_\alpha \\ \bar{\mathbf{C}}_\alpha\mathbf{B}^T & \bar{\mathbf{\Sigma}}_\alpha^{-1}+\bar{\mathbf{C}}_\alpha\end{bmatrix} \succeq \mathbf{0}\end{aligned} \tag{11}$$

where $\mathbf{Y} \in \mathbb{R}^{m\times m}$ is an auxiliary variable and $\alpha > 0$ is any scalar satisfying $\bar{\mathbf{\Sigma}}_\alpha = (\mathbf{R}_y + \mathbf{\Sigma}_w - \alpha\mathbf{I}_m) \succ \mathbf{0}$.

Proof: Due to the assumption of independence between $\mathbf{x}$ and $\mathbf{w}$, the objective function in (5) can be rewritten as

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{y}+\mathbf{w})-\mathbf{y}\|_2^2\right] \\ &= \text{tr}\left[\mathbf{R}_y - 2\mathbf{H}\mathbf{H}_s\mathbf{C}\mathbf{R}_y + \mathbf{H}\mathbf{H}_s\mathbf{C}(\mathbf{R}_y+\mathbf{\Sigma}_w)\mathbf{C}^T\mathbf{H}_s^T\mathbf{H}^T\right].\end{aligned} \tag{12}$$

Solving for $\mathbf{H}_s$ for a given fixed $\mathbf{C}$ yields (9). Matrix $\mathbf{C} \in \mathcal{C}_{pm}$ is full rank since it selects p *distinct* nodes, then $\mathbf{C}(\mathbf{R}_y + \mathbf{\Sigma}_w)\mathbf{C}^T$ has rank p and thus it is invertible [17]. Substituting the expression for $\mathbf{H}_s^*(\mathbf{C})$ into (12), yields (10). The inverse in the objective of (10) can be written as [12]

$$\begin{aligned}&(\mathbf{C}(\mathbf{R}_y+\mathbf{\Sigma}_w)\mathbf{C}^T)^{-1} = \left(\alpha\mathbf{I}_p - \alpha\mathbf{I}_p + \mathbf{C}(\mathbf{R}_y+\mathbf{\Sigma}_w)\mathbf{C}^T\right)^{-1} \\ &= \alpha^{-1}\mathbf{I}_p - \alpha^{-2}\mathbf{C}(\bar{\mathbf{\Sigma}}_\alpha^{-1} + \alpha^{-1}\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T,\end{aligned} \tag{13}$$

where $\alpha \neq 0$ is a rescaling parameter, and we used the Woodbury Matrix Identity [18]. Note that α has to be such that $\bar{\mathbf{\Sigma}}_\alpha = (\mathbf{R}_y + \mathbf{\Sigma}_w - \alpha\mathbf{I}_m)$ is still invertible. Substituting (13) into (10) and recalling that $\mathbf{C}^T\mathbf{C} = \text{diag}(\mathbf{c})$, we have that

$$\begin{aligned}\min_{\mathbf{c},\bar{\mathbf{C}}_\alpha} \ & \text{tr}\left[\mathbf{R}_y - \mathbf{B}\bar{\mathbf{C}}_\alpha\mathbf{B}^T + \mathbf{B}\bar{\mathbf{C}}_\alpha\left(\bar{\mathbf{\Sigma}}_\alpha^{-1}+\bar{\mathbf{C}}_\alpha\right)^{-1}\bar{\mathbf{C}}_\alpha\mathbf{B}^T\right] \\ \text{s.t.} \ & \bar{\mathbf{C}}_\alpha = \text{diag}(\mathbf{c})/\alpha \ , \ \mathbf{c}^T\mathbf{1}_m = p \ , \ \mathbf{c} \in \{0,1\}^m.\end{aligned} \tag{14}$$

Note that in (14) we optimize over a binary vector $\mathbf{c} \in \mathbb{R}^m$ with exactly p nonzero entries, instead of a binary matrix $\mathbf{C} \in \mathcal{C}_{pm}$. The p nonzero elements in $\mathbf{c}$ indicate the nodes to be sampled. Problem (14) can be reformulated as

$$\begin{aligned}\min_{\mathbf{c}\in\{0,1\}^m,\mathbf{Y}} \ & \text{tr}\left[\mathbf{Y}\right] \\ \text{s.t.} \ & \mathbf{R}_y - \mathbf{B}\bar{\mathbf{C}}_\alpha\mathbf{B}^T + \mathbf{B}\bar{\mathbf{C}}_\alpha\left(\bar{\mathbf{\Sigma}}_\alpha^{-1}+\bar{\mathbf{C}}_\alpha\right)^{-1}\bar{\mathbf{C}}_\alpha\mathbf{B}^T \preceq \mathbf{Y} \\ & \bar{\mathbf{C}}_\alpha = \text{diag}(\mathbf{c})/\alpha \ , \ \mathbf{c}^T\mathbf{1}_m = p\end{aligned} \tag{15}$$

where $\mathbf{Y} \in \mathbb{R}^{m\times m}$, $\mathbf{Y} \succeq 0$ is an auxiliary optimization variable. Using the Schur-complement lemma for *positive definite* matrices [17, Thm. 7.7.6], problem (15) can be written as (11). Hence, to complete the proof we need to show that $\bar{\mathbf{\Sigma}}_\alpha^{-1} + \bar{\mathbf{C}}_\alpha \succeq \mathbf{0}$ so that the aforementioned lemma can be invoked. To that end, suppose first that $\alpha < 0$. Then we have that $\bar{\mathbf{\Sigma}}_\alpha \succ \mathbf{0}$ and $\bar{\mathbf{C}}_\alpha = \alpha^{-1}\text{diag}(\mathbf{c}) \preceq \mathbf{0}$, so that $\bar{\mathbf{\Sigma}}_\alpha^{-1} + \bar{\mathbf{C}}_\alpha \succeq \mathbf{0}$ may not be positive definite. Suppose now that $\alpha > 0$. Then $\bar{\mathbf{C}}_\alpha \succeq \mathbf{0}$ and there always exists a sufficiently small positive α such that $\bar{\mathbf{\Sigma}}_\alpha \succ \mathbf{0}$ since $\mathbf{\Sigma}_w \succeq \mathbf{0}$. This implies that if α is chosen such that $\alpha > 0$ and $\bar{\mathbf{\Sigma}}_\alpha = \mathbf{R}_y + \mathbf{\Sigma}_w - \alpha\mathbf{I}_n \succ \mathbf{0}$ (which are the conditions

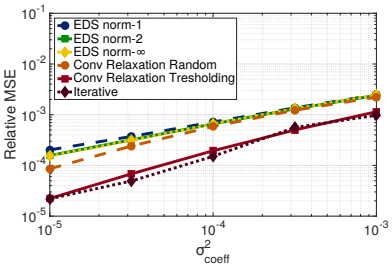

Fig. 2. Estimated MSE over 500 realizations as a function of noise for computing the GFT of a graph signal defined on an Erdős-Renyi graph of $m = 100$ nodes that has bandwidth $k = 10$.

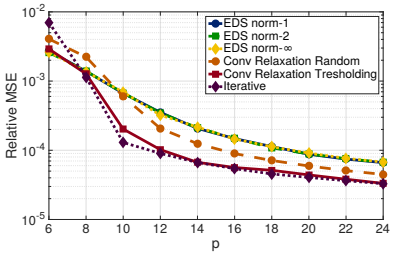

Fig. 3. Estimated MSE over 500 realizations as a function of the number of samples for computing the GFT of a graph signal defined on an Erdős-Renyi graph of $m = 100$ nodes that has bandwidth $k = 10$.

stated in the proposition), then $\bar{\mathbf{\Sigma}}_\alpha^{-1} + \bar{\mathbf{C}}_\alpha$ is positive definite and problems (14) and (11) are equivalent. ■

In words, the problem in (5) can be solved in two steps. First the optimal sketch $\mathbf{H}_s^*$ is found as a function of $\mathbf{C}$ via (9), and then this $\mathbf{H}_s^*(\mathbf{C})$ is substituted into (5) to formulate the optimization in (11), which depends only on $\mathbf{C}$. From an algorithmic perspective, the order is reversed. First, we find $\mathbf{C}$ by "solving" the binary optimization in either (10) or (11), and then the resulting $\mathbf{C}^*$ is substituted into (9) to find the *globally optimum* $\mathbf{H}_s^*$ as $\mathbf{H}_s^*(\mathbf{C}^*)$. Different heuristics to tackle the non-convex binary optimization over $\mathbf{C}$ are discussed next.

C. Heuristic approaches

Considering that $\mathbf{C}$ is binary, solving optimally the problems in (10) or (11) requires $\binom{m}{p}$ evaluations of the objective function. Since this may be computationally infeasible, this section provides two alternatives to find an approximate solution.

The first approach is to capitalize on the equivalent formulation in (11), relaxing the binary constraints as

$$\mathbf{c} \in \{0,1\}^m \underset{\text{Relaxation}}{\Rightarrow} \mathbf{0}_m \preceq \mathbf{c} \preceq \mathbf{1}_m, \tag{16}$$

so that (11) is transformed into a convex SDP problem that can be solved efficiently using $\mathcal{O}((n+m)^{3.5})$ operations. Once a solution to the relaxed problem is obtained, recovery of a binary vector $\mathbf{c}$ can be carried out in two ways. One alternative is to set the largest p elements to 1 and the rest to 0; this deterministic strategy is called tresholding. Another one normalizes the relaxed solution so that it adds up to one, and regards it as a probability distribution over the nodes. Then, the sampled nodes are drawn at random from this distribution, see [19]. Although not pursued here, also pertinent are formulations that penalize the objective with $\|\mathbf{c}\|_1$ and leverage ℓ_1-norm minimization advances to promote sparsity on $\mathbf{c}$.

The second approach relies on a greedy iterative scheme. Specifically, instead of searching over all possible $\binom{m}{p}$ sampling configurations, we run p iterations and in each of them we try nodes individually and retain the one yielding the best solution. This way only $(m)_p = m(m-1)\cdots(m-(p-1)) < m^p$ evaluations of the objective function (10) are required, so that the overall computational cost of this approximation is $\mathcal{O}((m)_p)$. Greedy algorithms have well-documented merits for sample selection, even for non-submodular objectives like the one in (10).

IV. Numerical experiments

In this section we consider the problem of computing the frequency coefficients of a graph signal. The numerical simulations carried out illustrate the performance of the sketching and sampling method described in this paper. We consider an Erdős-Rényi graph with $m = 100$ nodes and link activation probability of $p = 0.2$. We consider a k-bandlimited signal $\mathbf{y}$ on this graph with $k = 10$ and we wish to estimate the frequencies $\tilde{\mathbf{y}}_k$ that generated this signal through the linear model $\mathbf{y} = \mathbf{V}_k\tilde{\mathbf{y}}_k$. Thus, by setting $\mathbf{H} = \mathbf{V}_k$ as the linear transform and $\mathbf{x} = \tilde{\mathbf{y}}_k$, the method described here can be used to calculate the frequency coefficients of the signal in an efficient way. The signals are contaminated with white noise $\mathbf{w}$ independent of $\mathbf{y}$ with covariance matrix $\mathbf{\Sigma}_w = \sigma^2\mathbf{I}$ where σ^2 is the noise power calculated as $\sigma^2 = \sigma^2_{\text{coeff}} \cdot \|\tilde{\mathbf{y}}_k\|$. The different heuristics proposed in section III-C are compared to three experimental design sampling methods (EDS) proposed in [20], [21]. Specifically, we implement sampling with replacement with weights according to the: 1) ℓ_∞ norm of the rows of $\mathbf{V}$; 2) ℓ_1 norm of the rows of $\mathbf{V}$; and 3) ℓ_2 norm of the rows of $\mathbf{V}$. This latter case is the method proposed in [19]. In all cases, after obtaining the sampling matrix $\mathbf{C}$, the sketch $\mathbf{H}_s^*(\mathbf{C})$ is calculated as in (9) for each $\mathbf{C}$.

For the first simulation, we consider that $p = k = 10$ samples were taken and we estimate the relative MSE for varying levels of noise, from $\sigma^2_{\text{coeff}} = 10^{-5}$ to $\sigma^2_{\text{coeff}} = 10^{-3}$. For each value of σ^2_{coeff}, this MSE is obtained from averaging the relative error over 500 realizations. As can be seen from Fig. 2 the heuristic solutions to problem in Proposition 2 yield better performance than traditional sampling methods when computing the GFT of the graph signal efficiently. Similar levels of performance are achieved when the level of noise is fixed to $\sigma^2_{\text{coeff}} = 10^{-4}$ and the number of samples varies from $p = 6$ to $p = 24$ as illustrated in Fig. 3.

V. Conclusions

We studied the problem of computing efficient approximate solutions to linear inverse problems via sketching. The proposed method exploits sampling theory of bandlimited graph signals to minimize the online computational cost of obtaining an approximate solution, by designing a matrix sketch that operates only on a sampled version of the output. The joint design of the sampling scheme and the matrix sketch is carried out off line. The minimized online cost renders the proposed method ideal for streaming applications. Since the resultant problem was non-convex, different suboptimal schemes were proposed and their performance compared in the context of fast estimation of frequency components of a graph signal.

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 Processing Magazine*, vol. 30, no. 3, pp. 83–98, May 2013.

[2] 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 Processing Magazine*, vol. 31, no. 5, pp. 80–90, September 2014.

[3] S. K. Narang and A. Ortega, "Perfect Reconstruction Two-Channel Wavelet FIlter Banks for Graph Structured Data," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2786–2799, June 2012.

[4] A. Sandyhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3042–3054, June 2014.

[5] D. K. Berberidis, V. Kekatos, and G. B. Giannakis, "Online Censoring for Large-Scale Regressions with Application to Streaming Big Data," *IEEE Transactions on Signal Processing*, vol. 64, no. 15, pp. 3854–3867, August 2016.

[6] D. P. Woodruff, "Sketching as a Tool for Numerical Linear Algebra," *Foundations and Trends® in Theoretical Computer Science*, vol. 10, no. 1-2, pp. 1–157, 2014.

[7] I. Z. Pesenson and M. Z. Pesenson, "Sampling, Filtering and Sparse APproximations on Combinatorial Graphs," *Journal of Fourier Analysis and Applications*, vol. 16, no. 6, pp. 921–942, December 2010.

[8] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete Signal Processing on Graphs: Sampling Theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, December 2015.

[9] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of Graph Signals With Successive Local Aggregations," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, April 2016.

[10] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, "Signals on Graphs: Uncertainty Principle and Sampling," *arXiv:1507.08822v3 [cs.DM]*, 17 May 2016. [Online]. Available: http://arxiv.org/abs/1507.08822

[11] R. M. Gray, "Quantization in Task-Driven Sensing and Distributed Processing," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (ICASSP 2006)*. IEEE, 14-19 May 2006.

[12] S. Liu, S. P. Chepuri, M. Fardad, E. Masazade, G. Leus, and P. K. Varshney, "Sensor Selection for Estimation with Correlated Measurement Noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3509–3522, July 2016.

[13] M. Püschel and J. M. F. Moura, "Algebraic Signal Processing Theory: Foundation and 1-D Time," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3575–3585, August 2008.

[14] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, April 2013.

[15] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Stationary Graph Processes and Spectral Estimation," *arXiv:1603.04667v1 [cs.SY]*, 14 March 2016. [Online]. Available: http://arxiv.org/abs/1603.04667

[16] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed., ser. Wiley Series in Probability and Statistics. New York, NY: Wiley, 1973.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 1st ed. Cambridge, UK: Cambridge University Press, 1985.

[18] W. W. Hager, "Updating the Inverse of a Matrix," *SIAM Review*, vol. 31, no. 2, pp. 221–239, June 1989.

[19] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random Sampling of Bandlimited Graph Signals," *arXiv:1511.05118v2 [cs.SI]*, 20 May 2016. [Online]. Available: http://arxiv.org/abs/1511.05118

[20] R. Varma, S. Chen, and J. Kovačević, "Spectrum-Blind Signal Recovery on Graphs," in *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 13-16 December 2015, pp. 81–84.

[21] S. Chen, R. Varma, A. Singh, and J. Kovačević, "Signal recovery on graphs: Random versus experimentally designed sampling," in *2015 International Conference on Sampling Theory and Applications (SampTA)*. IEEE, 25-29 May 2015, pp. 337–341.