Eigendecomposition-Free Sampling Set Selection for Graph Signals

Akie Sakiyama, *Member, IEEE*, Yuichi Tanaka, *Senior Member, IEEE*, Toshihisa Tanaka, *Senior Member, IEEE* and Antonio Ortega, *Fellow, IEEE*

Abstract—This paper addresses the problem of selecting an optimal sampling set for signals on graphs. The proposed sampling set selection (SSS) is based on a localization operator that can consider both vertex domain and spectral domain localization. We clarify the relationships among the proposed method, sensor position selection methods in machine learning, and existing SSS methods based on graph frequency. In contrast to the alternative graph signal processing-based approaches, the proposed method does not need to compute the eigendecomposition of a variation operator, while still considering (graph) frequency information. We evaluate the performance of our approach through comparisons of prediction errors and execution time.

Index Terms—Graph signal processing, sampling set selection, graph sampling theorem, localization operator, graph uncertainty principle

I. INTRODUCTION

A. Motivation

Graphs give intuitive and effective representations for visualizing or investigating large quantities of intricately interrelated data. Network topologies have been studied in graph theory for a long time. In the past half-decade, the theory of analyzing and processing data on the vertices of a graph as well as underlying graph topologies, namely, signal processing on graphs, has been developed rapidly [1]–[4]. This theory enables us to efficiently apply signal processing techniques to many practical problems, such as social [2], traffic [5], brain [6], [7], and sensor networks [8], [9], following approaches similar to those used for audio, image, or time domain signals in traditional signal processing. This paper considers sampling methods for graph signals, a key topic in the development of signal processing on graphs.

The sampling of graph signals is an essential task for treating big or complex-structured data in the real world [1], [2]. Handling such raw data consumes a significant amount of system resources, both storage and computation, and sampled versions capturing most of the relevant information in the data would thus be highly desirable.

A main challenge in graph signal sampling is that in general there is no such thing as "regular sampling," and thus the sampling set has to be optimized based on the topology of the underlying graph. Many different approaches have been proposed for sampling set selection (SSS) on graphs [10]–[18]. Unfortunately, most of these methods have high computational complexity, as they require eigendecompositions to compute the graph Fourier basis (or some of its vectors).

SSS can be classified into deterministic and random sampling methods. Deterministic approaches [10]–[17] select vertices (often one-by-one) such that a target cost function is minimized or maximized by each selection, whereas random methods [19], [20] select vertices randomly according to some pre-computed probability distribution. In this study, we focus on the deterministic approach because it has the following advantage with respect to random sampling.

In random sampling-based methods, nothing prevents selecting vertices that have similar importance (higher probability) but that happen to be close to each other. We observe that this often happens, especially when graphs have irregular degree distributions and therefore the distribution of probabilities is also biased. When a high-probability vertex is selected and then a nearby vertex is also chosen, the second vertex may not lead to improvements in reconstruction performance. Thus, in practice, random sampling methods may perform well on average, but often require more samples than deterministic methods to achieve the same reconstruction quality.

Deterministic SSS techniques are based on selecting vertices for minimizing the reconstruction error when signals are reconstructed from their samples. They have been studied in the context of sampling theorems for graph signals [14], [16], [21]. They define a cost function based on the assumption that the reconstruction is performed using ideal filters under different optimality criteria (e.g., average case or worst case).

Recently, a vertex-localized SSS was proposed [22]. This is a two-step algorithm, whereby vertices are first screened to obtain a *permissible set* of vertices, i.e., vertices that are far enough from those vertices that have already been selected. Then, an optimal vertex is selected from the permissible set. Although this is conceptually similar to our approach, ours is a one-step algorithm. To select sufficiently far vertices, we control the vertex/spectral spread using graph spectral filters other than the ideal filters.

The above SSS methods are summarized in Table I. The abbreviations of the deterministic methods are found in Section II-B, along with the cost functions used in each case.

In this paper, we propose a deterministic sampling method

TABLE I
COMPARISON OF GRAPH-BASED SAMPLING SET SELECTION METHODS

	Deterministic/ random	Kernel	Localization in vertex domain	Localization in graph freq. domain
Cumulative coherence [19]	Random	Ideal	✓ *	✓
Global/local uncertainty [20]	Random	Arbitrary	✓	✓
MaxCutoff [16]	Deterministic	λ^k $(k \in \mathbb{Z}_+)$		✓
MinSpec/MinTrac [14]	Deterministic	Ideal		✓
MinFrob/MaxFrob/MaxPVol [21]	Deterministic	Ideal		✓
Vertex screening [22]	Deterministic	Ideal	✓	✓
Proposed method	Deterministic	Arbitrary	✓	✓

* Localized in the vertex domain only if the ideal kernel is approximated by a polynomial

for graph signals based on the graph localization operator [20] and reveal the relationship among the sensor selection methods based on the Gaussian process [23]–[26], the existing graph sampling methods [14], [16], [21], and the proposed method. The localization operator is introduced in the context of the uncertainty principle of graph signals [20]. It is the vertex domain operator with consideration of the graph frequency domain information.

Our contributions in this paper are summarized as follows.

- Using the localized operator for SSS, the following benefits are obtained: a) graph frequency localization makes it possible to mimic the frequency-based SSS criteria of [14]–[16], b) vertex localization is used to enable distributed SSS, and c) polynomial localization operators lead to lower complexity, i.e., eigendecomposition-free algorithms (see also Table I). This makes the SSS algorithm significantly faster (see Section V).
- We provide a unifying framework for many SSS techniques proposed to date as special cases of the localization operator-based SSS with different kernels and different optimization criteria to minimize the error. Even methods that were not initially viewed from a graph perspective, e.g., methods based on entropy [23], [24] and mutual information [25], [26], are included as its special cases (see Section IV).

Our preliminary work [27], [28] partially solved the problem of sensor position selection of sensor networks [25], [29]–[31] using sampling theory for graph signals [14]–[16] and proposed a sensor selection method based on the localization operator. This paper adds many theoretical and practical implications. Specifically, we newly propose sampling approaches based on error minimization and clarify the relationships between the alternative sampling methods for graph signals and the proposed methods.

In the experiment, we present the execution time and prediction error comparisons to evaluate the performance of the proposed approach. The proposed method is faster and shows better performance than the alternative approaches [14], [16], [21].

The rest of this paper is organized as follows. The preliminaries and notation are summarized in Section I-B. Section II introduces the sensor position selection approaches and the alternative graph sampling methods based on the graph Fourier basis. Section III provides the signal reconstruction method and describes the proposed vertex and signal selection algorithm based on the graph localization operator.

The section also compares the computational complexities of the alternative and proposed methods. Section IV clarifies that the proposed method has a deep connection with the alternative approaches introduced in Section II. Section V shows the experimental results of SSS and predicts the signals on unobserved vertices. Finally, Section VI concludes the paper.

B. Preliminaries and Notation

A graph is represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ denote sets of vertices and edges, respectively. A graph signal is defined as $\boldsymbol{f} \in \mathbb{R}^N$, where N is the number of vertices. We will only consider a connected, finite, undirected graph with no multiple edges. The variation operators are used for frequency analysis of graph signals. Although this paper mainly uses the graph Laplacian, we can use any variation operators, such as the adjacency matrix.

The combinatorial graph Laplacian is defined as $\mathbf{L} := \mathbf{D} - \mathbf{A}$, where $\mathbf{A}$ is the adjacency matrix whose (m, n)th element is the weight of the edge between m and n if m and n are connected, and 0 otherwise, and a diagonal matrix $\mathbf{D}$ is the degree matrix whose mth diagonal element is $D(m, m) = \sum_n A(m, n)$. The ith eigenvalue of $\mathbf{L}$ is λ_i, which can be ordered, without loss of generality as: $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \ldots \leq \lambda_{N-1} = \lambda_{\max}$, and its eigenvector is $\boldsymbol{u}_i \in \mathbb{C}^N$.

The graph Fourier transform is defined as follows [32], [33]: $\bar{\boldsymbol{f}} = \mathbf{U}^H \boldsymbol{f}$, where $\mathbf{U} = [\boldsymbol{u}_0 \ldots \boldsymbol{u}_{N-1}]$. The inverse graph Fourier transform is $\boldsymbol{f} = \mathbf{U}\bar{\boldsymbol{f}}$. Let $g(\lambda_i)$ be the spectral kernel; then, the filtering in the graph frequency domain can be written as $\boldsymbol{f}_{\text{out}} = \mathbf{U}g(\boldsymbol{\Lambda})\mathbf{U}^H \boldsymbol{f}_{\text{in}}$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_0, \ldots, \lambda_{N-1})$.

The nth element of the localization operator on the center vertex i is defined as [20]

$$T_{g,i}(n) = \sqrt{N} \sum_{l=0}^{N-1} g(\lambda_l) u_l^H(i) u_l(n), \qquad (1)$$

where $g(\lambda)$ is an arbitrary filter kernel. The matrix arranging localization operator in a row is

$$\mathbf{T} = [\boldsymbol{T}_{g,0} \ \boldsymbol{T}_{g,1} \cdots \boldsymbol{T}_{g,N-1}] = \mathbf{U}g(\boldsymbol{\Lambda})\mathbf{U}^H. \qquad (2)$$

The other notation used in this paper is summarized in Table II.

Symbol	Description
$\mu_i(\mathbf{Y})$	ith eigenvalue of $\mathbf{Y}$ $\mu_{\min}(\mathbf{Y}) = \mu_0(\mathbf{Y}) \leq \cdots \leq \mu_{N-1}(\mathbf{Y}) = \mu_{\max}(\mathbf{Y})$
$\boldsymbol{v}_i(\mathbf{Y})$	eigenvector of $\mathbf{Y}$ corresponding to $\mu_i(\mathbf{Y})$
$\sigma_i(\mathbf{X})$	ith singular value of $\mathbf{X}$ $\sigma_{\min}(\mathbf{X}) = \sigma_0(\mathbf{X}) \leq \cdots \leq \sigma_{N-1}(\mathbf{X})$
$\det[\mathbf{Y}]$	determinant of $\mathbf{Y}$, $\det[\mathbf{Y}] = \prod_i \mu_i(\mathbf{Y})$
$\mathrm{tr}[\mathbf{Y}]$	trace of $\mathbf{Y}$, $\mathrm{tr}[\mathbf{Y}] = \sum_i \mu_i(\mathbf{Y})$
$\|\boldsymbol{x}\|$	$\mathrm{diag}(\mathrm{sgn}(\boldsymbol{x}))\boldsymbol{x}$
$\|\mathbf{X}\|$	$\mathrm{sgn}(\mathbf{X}) \circ \mathbf{X}$
$\|\mathcal{A}\|$	number of elements in $\mathcal{A}$
$\|\boldsymbol{x}\|_p$	ℓ_p norm of $\boldsymbol{x}$
$\|\mathbf{X}\|_F$	Frobenius norm of $\mathbf{X}$, $\|\mathbf{X}\|_F = \sqrt{\mathrm{tr}(\mathbf{X}^H \mathbf{X})} = \sqrt{\sum_i^{\max\{M,L\}} \sigma_i^2(\mathbf{X})}$
$\|\mathbf{X}\|_2$	spectral norm of $\mathbf{X}$, $\|\mathbf{X}\|_2 = \sigma_{\max}(\mathbf{X})$
$\boldsymbol{x}_{\mathcal{A}}$	restriction of $\boldsymbol{x}$ to its components indexed by $\mathcal{A}$
$\mathbf{X}_{\mathcal{A}\mathcal{B}}$	restriction of $\mathbf{X}$ to its rows by $\mathcal{A}$ and columns by $\mathcal{B}$
$\mathbf{X}_{\mathcal{A}}$	$\mathbf{X}_{\mathcal{A}\mathcal{A}}$
$1_{\mathcal{A}}$	$1_{\mathcal{A}}(m) = 1$ if $m \in \mathcal{A}$ and 0 otherwise
$\mathbf{X}^T$, $\boldsymbol{x}^T$	transpose of $\mathbf{X}$ or $\boldsymbol{x}$
$\mathbf{X}^H$, $\boldsymbol{x}^H$	conjugate transpose of $\mathbf{X}$ or $\boldsymbol{x}$
$\mathbf{X}^{-1}$	inverse of $\mathbf{X}$
$\mathbf{X}^+$	Moore-Penrose pseudoinverse of $\mathbf{X}$

II. RELATED WORKS FOR SAMPLING SET SELECTION

We briefly introduce the objective functions of the related works for selecting sensor locations and sampling points of graph signals. Their derivations are described in Appendices A and B. We consider the problem of selecting $|\mathcal{S}| = F$ points, where $\mathcal{S}$ is the set of selected locations (for sensor selections) or vertices (for graph sampling theories), out of $|\mathcal{V}| = N$ possible locations or vertices in the original graph. The related works and their objective functions are summarized in Table III.

A. Sensor Position Selection Based on Gaussian Process

Sensor position selection algorithms have been developed in the area of machine learning. One of the major methods assumes that the spatial phenomena are modeled as a Gaussian process (GP) and, therefore, the stochastic signal $\boldsymbol{f}$ has the following Gaussian joint zero-mean distribution [29]:

$$p(\boldsymbol{f}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det[\mathbf{K}]} \exp\left(-\frac{1}{2}\boldsymbol{f}^T \mathbf{K}^{-1} \boldsymbol{f}\right), \quad (3)$$

where $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the covariance matrix of all locations $\mathcal{V}$ whose (i,j)th element is $\mathcal{K}(i,j)$ with a symmetric positive-definite kernel function $\mathcal{K}(\cdot,\cdot)$. The benefit of the GP model is that, if the signal $\boldsymbol{f}$ is distributed according to a multivariate Gaussian, the marginal and conditional distributions of its subset signal $f(y)$, where $y \in \mathcal{V}$, are also Gaussian with conditional variance $\sigma_{y|\mathcal{S}}^2 = \mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}}\mathbf{K}_{\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S}y}$. Under this assumption, sensors are placed at the most informative locations.

1) *Entropy [23], [24]:* The objective function is

$$\mathcal{S}^* = \arg\max_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \log\det[\mathbf{K}_{\mathcal{S}}]. \quad (4)$$

A greedy algorithm that adds sensor y^* satisfying following condition to $\mathcal{S}$ one by one is used for optimization:

$$y^* = \arg\max_{y \in \mathcal{S}_m^c} \mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}_m}\mathbf{K}_{\mathcal{S}_m}^{-1}\mathbf{K}_{\mathcal{S}_m y}, \quad (5)$$

where $\mathcal{S}_m$ are the already selected vertices in the mth iteration, $\mathcal{S}_m^c = \mathcal{V} \setminus \mathcal{S}_m$.

2) *Mutual Information (MI) [25], [26]:* The objective function is

$$\mathcal{S}^* = \arg\max_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \log\det[\mathbf{K}_{\mathcal{S}}] + \log\det[\mathbf{K}_{\mathcal{S}^c}], \quad (6)$$

where $\mathcal{S}^c = \mathcal{V} \setminus \mathcal{S}$. A greedy algorithm is also used for optimization:

$$y^* = \arg\max_{y \in \mathcal{S}_m^c} \frac{\mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}_m}\mathbf{K}_{\mathcal{S}_m}^{-1}\mathbf{K}_{\mathcal{S}_m y}}{\mathcal{K}(y,y) - \mathbf{K}_{y\overline{\mathcal{S}}_m}\mathbf{K}_{\overline{\mathcal{S}}_m}^{-1}\mathbf{K}_{\overline{\mathcal{S}}_m y}}, \quad (7)$$

where $\overline{\mathcal{S}}_m = \mathcal{V} \setminus (\mathcal{S}_m \cup y)$.

B. Graph Sampling Based on Graph Fourier Basis

Sampling methods based on graph frequency consider the problem of reconstructing bandlimited graph signals from their subsampled versions [12]–[15]. Note that we do not need to assume the GP model for the graph signal processing-based approaches (including the proposed approach).

Let us define ω- (for [16]) and $|\mathcal{F}|$- (for [14], [21]) bandlimited graph signals as the signals that have zero graph Fourier coefficients corresponding to the eigenvalues greater than ω and $\lambda_{|\mathcal{F}|-1}$, respectively: $\overline{f}(i) = 0$ for $\lambda_i > \omega$ or $i \geq |\mathcal{F}|$, where $\mathcal{F}$ is the set of indices associated with nonzero graph Fourier coefficients.

1) *Based on Cutoff Frequency (MaxCutoff):* The objective function [16] is

$$\mathcal{S}^* = \arg\max_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \mu_{\min}((\mathbf{L}^k)_{\mathcal{S}^c}). \quad (8)$$

The objective for a greedy optimization is represented as

$$y^* = \arg\max_{y \in \mathcal{S}_m^c} [v_{\min}^2((\mathbf{L}^k)_{\mathcal{S}_m^c})](y). \quad (9)$$

The signal recovered from the sampled one is calculated as

$$\widehat{\boldsymbol{f}} = \mathbf{U}_{\mathcal{V}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}}^+ \boldsymbol{f}_{\mathcal{S}}, \quad (10)$$

where $\mathcal{F}$ is the set of eigenvalues less than or equal to the estimated cutoff frequency $\Omega_k(\mathcal{S})$. If the original signal $\boldsymbol{f}$ is $\Omega_k(\mathcal{S})$-bandlimited, it can be perfectly recovered using (10).

2) Based on Error Minimization: [14] assumes that $|\mathcal{S}| \geq |\mathcal{F}|$ and uses (10) for the reconstruction. It proposes two objective functions for selecting optimal sampling sets:

- *MinSpec:*

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \max} \sigma_{\min}(\mathbf{U}_{\mathcal{S}\mathcal{F}}). \quad (11)$$

The greedy algorithm is used to optimize this problem:

$$y^* = \underset{y \in \mathcal{S}_m^c}{\arg \max} \sigma_{\min}(\mathbf{U}_{\mathcal{F}(\mathcal{S}_m \cup y)}). \quad (12)$$

- *MinTrac:*

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \min} \operatorname{tr}[(\mathbf{U}_{\mathcal{S}\mathcal{F}}^H \mathbf{U}_{\mathcal{S}\mathcal{F}})^{-1}]. \quad (13)$$

This also uses a greedy algorithm, which selects the vertex y^*:

$$y^* = \underset{y \in \mathcal{S}_m^c}{\arg \min} \operatorname{tr}[(\mathbf{U}_{\mathcal{F}(\mathcal{S}_m \cup y)}^H \mathbf{U}_{\mathcal{F}(\mathcal{S}_m \cup y)})^{-1}]. \quad (14)$$

3) Based on Localized Basis: [21] also assumes that $|\mathcal{S}| \geq |\mathcal{F}|$ and uses the following interpolation for recovering the original signal $\boldsymbol{f}$ from the sampled signal $\mathbf{D}_{\text{ver}}\boldsymbol{f}$:

$$\widehat{\boldsymbol{f}} = (\mathbf{D}_{\text{sp}}\mathbf{D}_{\text{ver}}\mathbf{D}_{\text{sp}})^+ \mathbf{D}_{\text{ver}}\boldsymbol{f}, \quad (15)$$

where $\mathbf{D}_{\text{ver}} = \operatorname{diag}(\mathbf{1}_{\mathcal{S}})$ and $\mathbf{D}_{\text{sp}} = \mathbf{U}\operatorname{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H$ are the sampling operator in the vertex domain and the bandlimiting operator in the graph frequency domain, respectively. This approach can perfectly recover the original signal $\boldsymbol{f}$ if it is $|\mathcal{F}|$-bandlimited.

There are three objective functions:

- *MinFrob:*

$$\begin{aligned} \mathcal{S}^* &= \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \min} \|(\mathbf{D}_{\text{sp}}\mathbf{D}_{\text{ver}}\mathbf{D}_{\text{sp}})^+\|_F \\ &= \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \min} \|(\operatorname{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H \mathbf{D}_{\text{ver}})^+\|_F \\ &= \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \min} \|(\mathbf{U}_{\mathcal{S}\mathcal{F}}^H)^+\|_F \end{aligned} \quad (16)$$

For this metric a greedy algorithm for optimization selects a vertex y^* at the mth step:

$$y^* = \underset{y \in \mathcal{S}_m^c}{\arg \min} \sum_{i=0}^{m-1} \frac{1}{\sigma_i(\mathbf{U}_{(\mathcal{S}_m \cup y)\mathcal{F}}^H)}. \quad (17)$$

- *MaxFrob:*

$$\begin{aligned} \mathcal{S}^* &= \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \max} \|\mathbf{D}_{\text{sp}}\mathbf{D}_{\text{ver}}\mathbf{D}_{\text{sp}}\|_F \\ &= \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \max} \|\mathbf{U}_{\mathcal{S}\mathcal{F}}^H\|_F. \end{aligned} \quad (18)$$

This can be solved by a simple strategy that selects the $|\mathcal{S}|$ columns of $\mathbf{U}_{\mathcal{S}\mathcal{F}}^H$ that have the maximum ℓ_2 norm.

- *MaxPVol:*

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F}{\arg \max} \det[\mathbf{U}_{\mathcal{S}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}}^H]. \quad (19)$$

This method also uses a greedy algorithm and the sampled vertex at the mth iteration is selected as

$$y^* = \underset{y \in \mathcal{S}_m^c}{\arg \max} \prod_{i=0}^{m-1} \mu_i(\mathbf{U}_{(\mathcal{S}_m \cup y)\mathcal{F}}\mathbf{U}_{(\mathcal{S}_m \cup y)\mathcal{F}}^H). \quad (20)$$

III. VERTEX SELECTION BASED ON LOCALIZATION OPERATOR

This section introduces the proposed SSS. First, we present the reconstruction method of missing graph signals based on the localization operator in (2). The sampled vertices are selected to minimize the reconstruction error or maximize the information corresponding to the localization operator. The computational complexities of the proposed and alternative methods are also discussed in this section.

A. Reconstruction Method

In our method, the missing values are reconstructed by a weighted linear combination of $\widetilde{\boldsymbol{T}}_{g,j} := (\mathbf{T}^k)_{j\mathcal{V}} = (\mathbf{U}g(\boldsymbol{\Lambda})^k\mathbf{U}^H)_{j\mathcal{V}}$ with arbitrary kernel $g(\cdot)$ which satisfies the condition described in the following theorem. That is, the sampled signal $\boldsymbol{f}_{\mathcal{S}}$ is recovered as follows:

$$\widehat{\boldsymbol{f}}_k = \sum_{j \in \mathcal{S}} \beta_j \widetilde{\boldsymbol{T}}_{g,j} = (\mathbf{T}^k)_{\mathcal{V}\mathcal{S}}\boldsymbol{\beta} = (\mathbf{T}^k)_{\mathcal{V}\mathcal{S}}((\mathbf{T}^k)_{\mathcal{S}})^+ \boldsymbol{f}_{\mathcal{S}}, \quad (21)$$

where $\boldsymbol{\beta} = ((\mathbf{T}^k)_{\mathcal{S}})^{-1}\boldsymbol{f}_{\mathcal{S}}$.

Theorem 1. *The $|\mathcal{F}|$-bandlimited signals with $|\mathcal{F}| \leq |\mathcal{S}|$ are perfectly recovered with (21) if k becomes large, i.e.,*

$$\lim_{k \to \infty} \widehat{\boldsymbol{f}}_k = \boldsymbol{f} \quad (22)$$

as long as the kernel of the localization operator satisfies $g(\lambda_i) > g(\lambda_j)$ for all $\lambda_i < |\mathcal{F}|$ and $\lambda_j \geq |\mathcal{F}|$.[1]

Proof. $\widehat{\boldsymbol{f}}_k$ can be rewritten as

$$\begin{aligned} \widehat{\boldsymbol{f}}_k &= (\mathbf{T}^k)_{\mathcal{V}\mathcal{S}}((\mathbf{T}^k)_{\mathcal{S}})^+ \boldsymbol{f}_{\mathcal{S}} \\ &= (\mathbf{U}g^k(\boldsymbol{\Lambda})\mathbf{U}^H)_{\mathcal{V}\mathcal{S}}((\mathbf{U}g^k(\boldsymbol{\Lambda})\mathbf{U}^H)_{\mathcal{S}})^+ \boldsymbol{f}_{\mathcal{S}} \\ &= \mathbf{U}g^k(\boldsymbol{\Lambda})\mathbf{U}_{\mathcal{S}\mathcal{V}}^H(\mathbf{U}_{\mathcal{S}\mathcal{V}}g^k(\boldsymbol{\Lambda})\mathbf{U}_{\mathcal{S}\mathcal{V}}^H)^+ \boldsymbol{f}_{\mathcal{S}} \\ &= \mathbf{U}g^{k/2}(\boldsymbol{\Lambda})(\mathbf{U}_{\mathcal{S}\mathcal{V}}g^{k/2}(\boldsymbol{\Lambda}))^+ \boldsymbol{f}_{\mathcal{S}} \\ &:= \mathbf{U}g^{k/2}(\boldsymbol{\Lambda})\widetilde{\boldsymbol{\alpha}}_k, \end{aligned} \quad (23)$$

where $\widetilde{\boldsymbol{\alpha}}_k := (\mathbf{U}_{\mathcal{S}\mathcal{V}}g^{k/2}(\boldsymbol{\Lambda}))^+ \boldsymbol{f}_{\mathcal{S}}$ and it is the estimation of $\boldsymbol{\alpha}_k = (g^{k/2}(\boldsymbol{\Lambda}))^+\mathbf{U}^H\boldsymbol{f}$ that is the modified graph Fourier coefficients by $g^{k/2}(\lambda)$. Because $\boldsymbol{f}$ is $|\mathcal{F}|$-bandlimited, $\alpha_k(m) = 0$ for $m > |\mathcal{F}|$ is always satisfied. Because $\boldsymbol{f}_{\mathcal{S}} = \mathbf{U}_{\mathcal{S}\mathcal{V}}g^{k/2}(\boldsymbol{\Lambda})\boldsymbol{\alpha}_k$, $\widetilde{\boldsymbol{\alpha}}_k = (\mathbf{U}_{\mathcal{S}\mathcal{V}}g^{k/2}(\boldsymbol{\Lambda}))^+ \boldsymbol{f}_{\mathcal{S}}$ is the estimation of $\boldsymbol{\alpha}_k$ only from $\boldsymbol{f}_{\mathcal{S}}$.

(23) is equivalently rewritten as

$$\widehat{\boldsymbol{f}}_k = \mathbf{U}\left(\frac{g(\boldsymbol{\Lambda})}{\beta}\right)^{k/2}\left(\mathbf{U}_{\mathcal{S}\mathcal{V}}\left(\frac{g(\boldsymbol{\Lambda})}{\beta}\right)^{k/2}\right)^+ \boldsymbol{f}_{\mathcal{S}}, \quad (24)$$

where $\beta = \min_{0 \leq i \leq |\mathcal{F}|-1} g(\lambda_i)$. Because $g(\lambda_i)/\beta < 1$ is satisfied for all $i \geq |\mathcal{F}|$, $\lim_{k \to \infty}(g(\lambda_i)/\beta)^{k/2} \to 0$. Then, for a sufficiently large k,

$$\mathbf{U}_{\mathcal{S}\mathcal{V}}\left(\frac{g(\boldsymbol{\Lambda}_{\mathcal{F}})}{\beta}\right)^{k/2} \approx \left[\left(\frac{g(\boldsymbol{\Lambda})}{\beta}\right)^{k/2}\mathbf{U}_{\mathcal{S}\mathcal{F}}^T \quad \mathbf{0}_{|\mathcal{S}|}\right]^T \quad (25)$$

[1]This condition indicates that $g(\lambda)$ is a low-pass filter.

		Objective	Objective w/ Localized Operator	Section
GP-based approach	Entropy [25]	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \log \det [\mathbf{K}_{\mathcal{S}}]$	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det [\mathbf{T}^{\mathrm{K}}_{\mathcal{S}}]$	II-A-1
	MI [25]	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \log \det [\mathbf{K}_{\mathcal{S}}] + \log \det [\mathbf{K}_{\mathcal{S}^c}]$	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det [\mathbf{T}^{\mathrm{K}}_{\mathcal{S}}] \det [\mathbf{T}^{\mathrm{K}}_{\mathcal{S}^c}]$	II-A-2
Graph frequency-based approach	MaxCutoff [16]	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \lambda_{\min}((\mathbf{L}^k)_{\mathcal{S}^c})$	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|(((\mathbf{T}^{\mathrm{L}})^k)_{\mathcal{S}^c})^{-1}\|_2$	II-B-1
	MinSpec [14]	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|\mathbf{U}^+_{\mathcal{S}\mathcal{F}}\|_2$	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|(\mathbf{T}^{\mathrm{I}}_{\mathcal{S}\mathcal{V}})^+\|_2$	II-B-2
	MinTrac [14]	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{U}^H_{\mathcal{S}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}})^{-1}]$	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{T}^{\mathrm{I}}_{\mathcal{S}})^{-1}]$	II-B-2
	MinFrob [21]	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|(\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+\|_F$	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{T}^{\mathrm{I}}_{\mathcal{S}})^{-1}]$	II-B-3
	MaxFrob [21]	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}}\|_F$	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{T}^{\mathrm{I}}_{\mathcal{S}})]$	II-B-3
	MaxPVol [21]	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det [\mathbf{U}^H_{\mathcal{S}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}}]$	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det [\mathbf{T}^{\mathrm{I}}_{\mathcal{S}}]$	II-B-3

and

$$\left(\mathbf{U}_{\mathcal{S}\mathcal{V}}\left(\frac{g(\mathbf{\Lambda})}{\beta}\right)^{k/2}\right)^+ \approx \left[\left(\frac{g(\mathbf{\Lambda}_{\mathcal{F}})}{\beta}\right)^{-k/2}\mathbf{U}^+_{\mathcal{S}\mathcal{F}}\quad \mathbf{0}_{|\mathcal{S}|}\right], \quad (26)$$

where $\mathbf{0}_{|\mathcal{S}|}$ is a $|\mathcal{S}| \times |\mathcal{S}|$ null matrix. From (26), (24) can be rewritten as

$$\widehat{\boldsymbol{f}}_k = \mathbf{U}_{\mathcal{V}\mathcal{F}}\mathbf{U}^+_{\mathcal{S}\mathcal{F}}\boldsymbol{f}_{\mathcal{S}}. \quad (27)$$

This coincides with (10), and therefore, (21) can perfectly recover the $|\mathcal{F}|$-bandlimited signals. $\qquad\square$

B. Reconstruction Error

Here, we consider the reconstruction error by (21) when the sampled signal contains additive noise, i.e., $\boldsymbol{o} = \boldsymbol{f}_{\mathcal{S}} + \boldsymbol{n}_{\mathcal{S}}$, where $\boldsymbol{f}$ is an $|\mathcal{F}|$-bandlimited signal and each element in the additive noise term $\boldsymbol{n} \in \mathbb{R}^N$ is an i.i.d. zero-mean Gaussian random variable with variance σ^2.

The error $\boldsymbol{e} := \boldsymbol{f} - \widehat{\boldsymbol{f}}_k$ is represented as

$$\boldsymbol{e} = \boldsymbol{f} - (\mathbf{T}^k)_{\mathcal{V}\mathcal{S}}((\mathbf{T}^k)_{\mathcal{S}})^+\boldsymbol{o} = (\mathbf{T}^k)_{\mathcal{V}\mathcal{S}}((\mathbf{T}^k)_{\mathcal{S}})^+\boldsymbol{n}_{\mathcal{S}}. \quad (28)$$

This can be rewritten as

$$\begin{aligned}
\boldsymbol{e} &= \mathbf{U}g^k(\mathbf{\Lambda})\mathbf{U}^H_{\mathcal{S}\mathcal{V}}(\mathbf{U}_{\mathcal{S}\mathcal{V}}g^k(\mathbf{\Lambda})\mathbf{U}^H_{\mathcal{S}\mathcal{V}})^+\boldsymbol{n}_{\mathcal{S}} \\
&= \mathbf{U}g^{k/2}(\mathbf{\Lambda})\mathbf{U}^H\mathbf{U}g^{k/2}(\mathbf{\Lambda})\mathbf{U}^H_{\mathcal{S}\mathcal{V}} \\
&\quad \times (\mathbf{U}_{\mathcal{S}\mathcal{V}}g^{k/2}(\mathbf{\Lambda})\mathbf{U}^H\mathbf{U}g^{k/2}(\mathbf{\Lambda})\mathbf{U}^H_{\mathcal{S}\mathcal{V}})^+\boldsymbol{n}_{\mathcal{S}} \\
&= \mathbf{T}^{k/2}((\mathbf{T}^{k/2})_{\mathcal{S}\mathcal{V}})^+\boldsymbol{n}_{\mathcal{S}}.
\end{aligned} \quad (29)$$

Then, the error covariance matrix is calculated as

$$\begin{aligned}
\mathbf{E} &= \boldsymbol{e}\boldsymbol{e}^H \\
&= \mathbf{T}^{k/2}((\mathbf{T}^{k/2})_{\mathcal{S}\mathcal{V}})^+\boldsymbol{n}_{\mathcal{S}}\boldsymbol{n}^H_{\mathcal{S}}((\mathbf{T}^{k/2})^H_{\mathcal{S}\mathcal{V}})^+\mathbf{T}^{k/2} \\
&= \sigma^2\mathbf{T}^{k/2}((\mathbf{T}^{k/2})^H_{\mathcal{S}\mathcal{V}}(\mathbf{T}^{k/2})_{\mathcal{S}\mathcal{V}})^+\mathbf{T}^{k/2}.
\end{aligned} \quad (30)$$

For minimizing the error, we should minimize or maximize the trace, determinant, or maximum eigenvalue of $\mathbf{E}$, depending on the optimization strategy. The approaches are summarized in Table IV, and their derivations are shown in Appendix C.

C. Vertex Selection Methods Based on Support of Localization Operator

While the cost function based on (30) can interpret various existing SSS approaches as its special cases (see Section IV), a naive realization of maximizing/minimizing the cost functions

Optimal Design	Objective
A-optimal	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[((\mathbf{T}^k)_{\mathcal{S}})^{-1}]$
D-optimal	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det[((\mathbf{T}^k)_{\mathcal{S}})^{-1}]$
E-optimal	$\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|((\mathbf{T}^{k/2})_{\mathcal{S}\mathcal{V}})^+\|_2$
T-optimal	$\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{T}^k)_{\mathcal{S}}]$

in Table IV needs eigendecomposition, which leads to high computational complexity.

We reconsider the intuition of the localization operator, which avoids the above-mentioned problem. Intuitively, the set of vertices $\mathcal{S}$ should be the most informative concerning the localization operator. The localized operator $\boldsymbol{T}_{g,i}$ is a filtered version of $\delta(i)$, where the filter kernel is $g(\lambda)$. Since $\delta(i)$ is completely localized at vertex i, the energy in $\boldsymbol{T}_{g,i}$ is also centered at vertex i.

The missing samples are interpolated by a weighted linear combination of the localization operators located at all observed samples. Thus, the "support" of a given $\boldsymbol{T}_{g,i}$ (the set of vertices where this function has nonzero values) can be regarded as the area (set of vertices) for which the ith vertex can contribute to interpolating unobserved samples. Thus, in order to improve our ability to reconstruct from samples, it would make sense to maximize the "covering area" for the sampling set $\mathcal{S}$, denoted $\#(|\boldsymbol{T}_{g,i}|, \epsilon)$, that is the number of elements in $|\boldsymbol{T}_{g,i}|$ that are larger than ϵ.

To maximize the support for $\mathcal{S}$, i.e., $\sum_{i\in\mathcal{S}} \#(|\boldsymbol{T}_{g,i}|, \epsilon)$, our goal is to select localized operators that satisfy the following condition: Each operator $\boldsymbol{T}_{g,i}$ has large energy in a first-order sense, i.e., $\|\boldsymbol{T}_{g,i}\|_1$ is large, since we use their weighted linear combination for reconstruction. Additionally, the overlapped areas of $\boldsymbol{T}_{g,i}$ and $\boldsymbol{T}_{g,j}$ $(i \neq j)$ should be small since two localization operators centered at vertices i and j can be redundant if they overlap with each other. Therefore, we select vertices such that the sum of $\|\boldsymbol{T}_{g,i}\|_1$ is large and overlapping area covered by both $\boldsymbol{T}_{g,i}$ and $\boldsymbol{T}_{g,j}$ $(i \neq j)$ is small:

$$\begin{aligned}
\mathcal{S}^* &= \underset{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F}{\arg\max} \eta\sum_{i\in\mathcal{S}}\langle\mathbf{1}_{N\times 1}, |\boldsymbol{T}_{g,i}|\rangle - \sum_{j\in\mathcal{S},j\neq i}\langle|\boldsymbol{T}_{g,i}|, |\boldsymbol{T}_{g,j}|\rangle \\
&= \underset{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F}{\arg\max} \sum_{i\in\mathcal{S}}\left\langle\left(\eta\mathbf{1}_{N\times 1} - \sum_{j\in\mathcal{S},j\neq i}|\boldsymbol{T}_{g,j}|\right), |\boldsymbol{T}_{g,i}|\right\rangle,
\end{aligned} \quad (31)$$

where $\eta \in \mathbb{R}_+$ is an arbitrary real value. To optimize the cost function, we use a greedy algorithm, which appends one vertex in the mth iteration by selecting a vertex y^* satisfying the following function:

$$y^* = \arg\max_{y \in \mathcal{S}_m^c} \left\langle R\left(\eta \mathbf{1}_{N \times 1} - \sum_{j \in \mathcal{S}_m} |\boldsymbol{T}_{g,j}|\right), |\boldsymbol{T}_{g,y}| \right\rangle,$$
(32)

where $R(\cdot)$ is the ramp function that satisfies $[R(\boldsymbol{x})](i) = x(i)$ if $x(i) \geq 0$ and 0 otherwise, and $\eta \in \mathbb{R}_+$ is an arbitrary real value.

In (32), we calculate the weighted norm of $\boldsymbol{T}_{g,y}$. A small weight is assigned to $T_{g,y}(i)$ if the ith vertex has already been covered: In this case, the weight of $\sum_{j \in \mathcal{S}_m} |\boldsymbol{T}_{g,j}|$ at the ith vertex is large. In each iteration, we avoid selecting vertices whose localization operators overlap with those of already-selected vertices, because the weight for $T_{g,y}(i)$ becomes 0 when $\sum_{j \in \mathcal{S}_m} |T_{g,j}(i)| \geq \eta$. In this study, we use $\eta = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{S}_m} |T_{g,j}(i)|$, which is experimentally determined.

If the kernel $g(\lambda)$ is a polynomial function, we can calculate (32) without an eigendecomposition of the graph Laplacian. This is because (2) is rewritten as $\mathbf{T} = \sqrt{N} g(\mathbf{L})$ when $g(\lambda)$ is a polynomial function. Therefore, localization operators can be obtained without the eigenvectors themselves. As a result, if the original kernel $g(\lambda)$ is a polynomial or the Chebyshev polynomial approximation is applied to $g(\lambda)$, an eigendecomposition is not required for the proposed SSS. Using the polynomial function, (32) can be rewritten as

$$y^* = \arg\max_{y \in \mathcal{S}_m^c} \left[|\mathbf{T}|\,\boldsymbol{w}\right](y),$$
(33)

where $\boldsymbol{w} = R\left(\eta \mathbf{1}_{N \times 1} - \sum_{j \in \mathcal{S}_m} |\boldsymbol{T}_{g,j}|\right)$.

D. Computational Complexity

Table V compares the computational complexities of the graph signal processing-based methods [10], [14], [16] and the proposed method shown in Section IV-B-2, where T_1 is the average number of iterations required for the convergence of a single eigen-pair, T_F is the number of iterations of convergence for the first F eigen-pair, k provides a trade-off between performance and complexity of the method proposed in [16], C is a constant, P is the approximation order of the Chebyshev polynomial approximation, and J is the number of nonzero elements in $\mathbf{T}$. We follow the notation in [16].

Note that $\mathbf{T}$ is a sparse matrix because its ith row has nonzero elements only at the columns corresponding to the ith vertex and several neighboring vertices. The calculation of the localization operator in the proposed method includes complexity for performing the Chebyshev polynomial approximation and filtering [33].

It can be seen that the calculation of the localization operator shows much lower complexity than those for calculating the eigen-pairs in the other approaches. MaxCutoff needs the calculation of eigen-pairs in each iteration whereas the other methods calculate the eigen-pairs or operator only once. Therefore, although the proposed method has a higher complexity

order than MaxCutoff in the sampling set search in Table V, its total execution time is usually lower than MaxCutoff.

IV. RELATIONSHIP BETWEEN PROPOSED METHOD AND ALTERNATIVE APPROACHES

The objective functions of the alternative methods can be rewritten using the localization operator with various kernels. They are summarized in Table III. Furthermore, we show that the existing approaches based on the graph Fourier basis are one of the proposed method based on the error minimization.

A. Sensor Position Selection Based on Gaussian Process

The sensor selection methods based on the GP model introduced in Section III-A can be viewed as the SSS approaches for graph signals that use the covariance matrix instead of the Laplacian matrix. In general, the graph Laplacian (precision matrix) and the covariance matrix have the following relationship [34]:

$$\mathbf{L} = \mathbf{K}^{-1} - \delta \mathbf{I}.$$
(34)

The parameter δ prevents the precision matrix from being singular. The precision matrix has the same set of eigenvectors $\{\boldsymbol{u}_i = \boldsymbol{v}_i(\mathbf{K})\}_{i=0, ..., N-1}$ with corresponding eigenvalues $\{\lambda_i = \frac{1}{\mu_i(\mathbf{K})} - \delta\}_{i=0, ..., N-1}$.

From (34), (3) indicates that the random signals have following distributions:

$$\begin{aligned} p(\boldsymbol{f}) &\propto \exp\left(-\boldsymbol{f}^T \mathbf{K}^{-1} \boldsymbol{f}\right) \\ &= \exp\left(-\boldsymbol{f}^T (\mathbf{L} + \delta \mathbf{I}) \boldsymbol{f}\right) \\ &= \exp\left(-\sum_i \sum_j A(i,j)(f(i) - f(j))^2 - \delta \sum_i f(i)^2\right), \end{aligned}$$
(35)

namely, vertices with similar signal values are connected by edges with large weights. This also indicates that the signals are smooth over the graph with the Laplacian obtained by (34).

The localization operator can rewrite the entropy criterion (4) as

$$\begin{aligned} \mathcal{S}^* &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}|=F} \log \det[\mathbf{K}_{\mathcal{S}}] \\ &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}|=F} \det[((\mathbf{L} + \delta \mathbf{I})^{-1})_{\mathcal{S}}] \\ &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}|=F} \det[\mathbf{T}_{\mathcal{S}}^{\mathrm{K}}], \end{aligned}$$
(36)

where $\mathbf{T}^{\mathrm{K}} = (\mathbf{L} + \delta \mathbf{I})^{-1}$, i.e., the localization operator with the kernel $g(\lambda) = 1/(\lambda + \delta)$. Similarly, (6) is

$$\begin{aligned} \mathcal{S}^* &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}|=F} \log \det[\mathbf{K}_{\mathcal{S}}] + \log \det[\mathbf{K}_{\overline{\mathcal{S}}\overline{\mathcal{S}}}] \\ &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}|=F} \det[\mathbf{T}_{\mathcal{S}}^{\mathrm{K}}] \det[\mathbf{T}_{\overline{\mathcal{S}}\overline{\mathcal{S}}}^{\mathrm{K}}]. \end{aligned}$$
(37)

Next, we clarify the characteristic of the greedy optimization step from a graph signal processing perspective. From the

TABLE V
COMPUTATIONAL COMPLEXITIES OF GRAPH SIGNAL PROCESSING-BASED APPROACHES

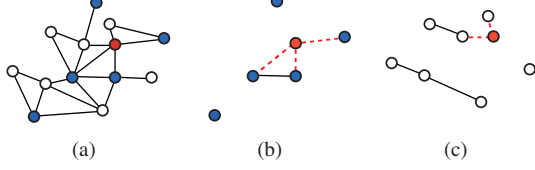|	MaxCutoff [16]	MinSpec [14]	MinFrob [21]	MaxFrob [21]	MaxPVol [21]	Proposed Method w/ CPA
Eigen-pair or operator computations	$O(k\|\mathcal{E}\|FT_1)$		$O((\|\mathcal{E}\|F + CF^3)T_F)$			$O((\|\mathcal{E}\| + N)P + J)$
Sampling set search	$O(NF)$	$O(NF^4)$	$O(NF^4)$	$O(NF)$	$O(F^3)$	$O(JF)$
Section	III-B-1	III-B-2	III-B-3-i	III-B-3-ii	III-B-3-iii	IV-B-2

Fig. 1. (a) Original graph. The blue vertices and red vertex indicate $\mathcal{S}$ and y, respectively. (b) $\mathbf{L}^y$. $L^y(y,y)$ is the total weight of the red dashed edges. (c) $\bar{\mathbf{L}}^y$. $\bar{L}^y(y,y)$ is the total weight of the red dashed edges.

block matrix inversion formula, the inversion of the covariance matrix can be represented as [13]:

$$
\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{K}_{\mathcal{S}^c} & \mathbf{K}_{\mathcal{S}^c\mathcal{S}} \\ \mathbf{K}_{\mathcal{S}\mathcal{S}^c} & \mathbf{K}_{\mathcal{S}} \end{bmatrix}^{-1}
$$
$$
= \begin{bmatrix} \mathbf{K}_{\mathcal{S}^c|\mathcal{S}}^{-1} & -(\mathbf{K}_{\mathcal{S}^c})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}\mathbf{K}_{\mathcal{S}|\mathcal{S}^c}^{-1} \\ -(\mathbf{K}_{\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}^{T}\mathbf{K}_{\mathcal{S}^c|\mathcal{S}}^{-1} & \mathbf{K}_{\mathcal{S}|\mathcal{S}^c}^{-1} \end{bmatrix},
$$
(38)

where $\mathbf{K}_{\mathcal{S}^c|\mathcal{S}} = \mathbf{K}_{\mathcal{S}^c} - \mathbf{K}_{\mathcal{S}^c\mathcal{S}}(\mathbf{K}_{\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}^{T}$ and $\mathbf{K}_{\mathcal{S}|\mathcal{S}^c} = \mathbf{K}_{\mathcal{S}} - \mathbf{K}_{\mathcal{S}\mathcal{S}^c}(\mathbf{K}_{\mathcal{S}^c})^{-1}\mathbf{K}_{\mathcal{S}\mathcal{S}^c}^{T}$. Using (34) and (38), the graph Laplacian and the covariance matrix have the following relationship:

$$
\mathbf{L}_{\mathcal{S}^c} + \delta\mathbf{I} = (\mathbf{K}_{\mathcal{S}^c} - \mathbf{K}_{\mathcal{S}^c\mathcal{S}}(\mathbf{K}_{\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}^{T})^{-1}.
$$
(39)

Fig. 1 considers a toy example that uses a synthesized simple graph for the sake of clarity. From (39), we can rewrite the entropy criterion in (5) as:

$$
y^* = \arg\max_{y \in \mathcal{S}_m^c} \frac{1}{L^y(y,y) + \delta^y},
$$
(40)

where $\mathbf{L}^y$ is the Laplacian matrix of the graph with the vertices $\mathcal{S}_m \cup y$ and the edges between these vertices (Fig. 1 (b)), and δ^y is the variance of $\boldsymbol{f}_{\mathcal{S}_m \cup y}$. It can be seen that the entropy criterion selects a vertex that has the minimum degree with the selected vertices, i.e., the vertex with the weakest connection with the already-selected vertices is selected. Because of this, the entropy criterion often places many vertices at the corners or boundaries of the space, as is well known.

The MI criterion in (7) can also be rewritten as

$$
y^* = \arg\max_{y \in \mathcal{S}_m^c} \frac{\bar{L}^y(y,y) + \bar{\delta}^y}{L^y(y,y) + \delta^y},
$$
(41)

where $\bar{\mathbf{L}}^y$ is the graph Laplacian containing the unselected vertices $\mathcal{S}_m^c$ and the edges in $\mathcal{S}_m^c$ (Fig. 1 (c)), and $\bar{\delta}^y$ is the variance of $\boldsymbol{f}_{\mathcal{S}_m^c}$. It can be observed that the MI criterion chooses the vertex that has the weakest connection with the selected vertices and strongest connection with the unselected vertices.

It is worth noting that the conventional entropy and MI criterion select vertices according to the edge information in the *graph vertex domain*, whereas the sampling methods based on the graph Fourier basis including the proposed method select vertices while considering the spectrum in the *graph frequency domain*.

B. Graph Sampling Based on Fourier Basis

1) Based on Cutoff Frequency (MaxCutoff): The reconstruction algorithm in (21) with the ideal kernel, i.e., $\mathbf{T} = \mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H$, coincides with that of the alternative approaches in (10):

$$
\begin{aligned}
\widehat{\boldsymbol{f}}_k &= (\mathbf{T}^k)_{\mathcal{V}\mathcal{S}}((\mathbf{T}^k)_{\mathcal{S}})^+ \boldsymbol{f}_{\mathcal{S}} \\
&= \mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}_{\mathcal{S}\mathcal{V}}^H(\mathbf{U}_{\mathcal{S}\mathcal{V}}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}_{\mathcal{S}\mathcal{V}}^H)^+ \boldsymbol{f}_{\mathcal{S}} \\
&= \mathbf{U}_{\mathcal{V}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}}^H(\mathbf{U}_{\mathcal{S}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}}^H)^+ \boldsymbol{f}_{\mathcal{S}} \\
&= \mathbf{U}_{\mathcal{V}\mathcal{F}}\mathbf{U}_{\mathcal{S}\mathcal{F}}^+ \boldsymbol{f}_{\mathcal{S}}.
\end{aligned}
$$
(42)

Furthermore, the objective function in (8) can be rewritten as

$$
\begin{aligned}
\mathcal{S}^* &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \mu_{\min}(\mathbf{L}^k)_{\mathcal{S}^c} \\
&= \arg\min_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \mu_{\max}(((\mathbf{L}^k)_{\mathcal{S}^c})^{-1}) \\
&= \arg\min_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \|(((\mathbf{T}^L)^k)_{\mathcal{S}^c})^{-1}\|_2,
\end{aligned}
$$
(43)

where $\mathbf{T}^L = \mathbf{L}$ is the localization operator matrix with $g(\lambda_i) = \lambda_i$.

2) Based on Error Minimization:

- *MinSpec:* (11) is rewritten as

$$
\begin{aligned}
\mathcal{S}^* &= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \sigma_{\min}(\mathbf{U}_{\mathcal{S}\mathcal{F}}) \\
&= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \sigma_{\min}(\mathbf{U}_{\mathcal{F}\mathcal{S}}^H) \\
&= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \sigma_{\min}(\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathrm{diag}(\mathbf{1}_{\mathcal{S}})) \\
&= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \sigma_{\min}(\mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathrm{diag}(\mathbf{1}_{\mathcal{S}})) \\
&= \arg\max_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \sigma_{\min}(\mathbf{T}_{\mathcal{V}\mathcal{S}}^I) \\
&= \arg\min_{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = F} \|(\mathbf{T}_{\mathcal{V}\mathcal{S}}^I)^+\|_2
\end{aligned}
$$
(44)

where $\mathbf{T}^I = \mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H$ which is the localization operator matrix with the ideal filter: $g(\lambda_i) = 1$ for $\lambda_i \in \mathcal{F}$ and 0 otherwise.

- *MinTrac:* (13) is also rewritten as

$$\begin{aligned} \mathcal{S}^* &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{U}_{\mathcal{SF}}\mathbf{U}_{\mathcal{SF}}^H)^{-1}] \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[((\mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H)_{\mathcal{S}})^{-1}] \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{T}_{\mathcal{S}}^{\mathrm{I}})^{-1}]. \end{aligned} \quad (45)$$

3) Based on Localized Basis:

(i) *MinFrob:* (16) is rewritten as

$$\begin{aligned} \mathcal{S}^* &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|(\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathbf{D}_{\mathrm{ver}})^+\|_F \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \sum_{i=0}^{|\mathcal{F}|} \frac{1}{\sigma_i(\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathrm{diag}(\mathbf{1}_{\mathcal{S}}))} \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \sum_{i=0}^{|\mathcal{F}|} \frac{1}{\sigma_i(\mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathrm{diag}(\mathbf{1}_{\mathcal{S}}))} \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \sum_{i=0}^{|\mathcal{F}|} \frac{1}{\sigma_i(\mathbf{T}_{\mathcal{VS}}^{\mathrm{I}})} \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|(\mathbf{T}_{\mathcal{VS}}^{\mathrm{I}})^+\|_F \\ &= \arg\min_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[(\mathbf{T}_{\mathcal{S}}^{\mathrm{I}})^{-1}]. \end{aligned} \quad (46)$$

(ii) *MaxFrob:* Similar to (i), (18) is rewritten as

$$\begin{aligned} \mathcal{S}^* &= \arg\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathbf{D}_{\mathrm{ver}}\|_F \\ &= \arg\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \|\mathbf{T}_{\mathcal{VS}}^{\mathrm{I}}\|_F \\ &= \arg\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \mathrm{tr}[\mathbf{T}_{\mathcal{S}}^{\mathrm{I}}]. \end{aligned} \quad (47)$$

(iii) *MaxPVol:* (19) is rewritten as

$$\begin{aligned} \mathcal{S}^* &= \arg\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det[\mathbf{U}_{\mathcal{SF}}\mathbf{U}_{\mathcal{SF}}^H] \\ &= \arg\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det[(\mathbf{U}\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H)_{\mathcal{S}}] \\ &= \arg\max_{\mathcal{S}\subset\mathcal{V}:|\mathcal{S}|=F} \det[\mathbf{T}_{\mathcal{S}}^{\mathrm{I}}]. \end{aligned} \quad (48)$$

The objective functions represented by the localization operators are summarized in Table III. Because $(\mathbf{T}^{\mathrm{I}})^k = \mathbf{T}^{\mathrm{I}}$, the alternative SSS methods shown in (45)–(48) coincide with the proposed SSS based on the error minimization, which is introduced in Table IV, in the case of using the ideal kernel for the localization operator. Furthermore, MaxCutoff, as shown in (43), can also be viewed as the objective function for minimizing the error covariance matrix caused by the reconstruction shown in (21).

Proposition 1. *MaxCutoff can be viewed as the error minimization for signal reconstruction using (21) with* $\mathbf{T} = (\mathbf{L} + \delta\mathbf{I})^{-1}$*, in the case in which δ goes to zero.*[2]

The proof is shown in Appendix D.

In summary, all of the existing SSS methods introduced in this paper can be viewed as special cases of the proposed SSS based on the error minimization with the different optimal criteria and kernels.

---

[2]δ prevents the precision matrix from being singular.

TABLE VI
SPEEDUP FACTOR OF OUR METHOD WITH RESPECT TO ALTERNATIVE
APPROACHES FOR $N = 2000$

Methods	Speedup factor
Entropy	1.68
MI	262.85
MaxCutoff	120.24
MinSpec	833.01
MinFrob	1898.37
MaxFrob	0.94
MaxPVol	1843.99
RandSamp	0.14

V. EXPERIMENTAL RESULTS

A. Setup

In the experiments, we used the following six graphs:

- Random sensor graph.
- Random graph with Erdős–Rényi model (ER graph): The edge connecting probability was set to 0.05.
- Random regular graph: Each vertex connects to six vertices.
- Random graph with Barabási–Albert model (BA graph): The initial connected graph has six vertices.
- Community graph: 11 communities with random sizes are yielded.
- Minnesota Traffic graph.

For the comparison of execution time, we selected the random sensor graph with different numbers of vertices. For the comparison of prediction errors, we used all six graphs with the number of vertices for random graphs is set to $N = 500$; the number of vertices in the Minnesota Traffic graph is $N = 2,642$.

The performance of the proposed method is compared with the following approaches:

- GP-based methods: Entropy- and MI-based criteria (abbreviated as Entropy and MI) [23]–[25]
- Graph-based SSS methods with deterministic selection [14], [16], [21]
- Graph-based SSS using random sampling with nonuniform sampling probability distribution (abbreviated as RandSamp) [19]

For the GP-based methods, we need to estimate the covariance matrix. Based on (34), we simply set $\mathbf{K} = (\mathbf{L} + \delta\mathbf{I})^{-1}$, where $\delta = 0.01$. MaxCutoff needs the parameter k for estimating the cutoff frequency, which is set to $k = 14$. RandSamp also needs the parameter γ for the reconstruction, which is set to $\gamma = 1$. Although the optimal γ widely varies according to the graph used, $\gamma = 1$ is used for one of the experiments in [19].

As the signal model for a realistic situation, we use noisy bandlimited signals where

$$\overline{\boldsymbol{f}} = [\overline{\boldsymbol{f}}_{\mathrm{bl}}^T, \ \mathbf{0}_{N-|\mathcal{F}|}^T]^T + \boldsymbol{\epsilon}, \quad (49)$$

in which $\overline{\boldsymbol{f}}_{\mathrm{bl}} \in \mathbb{R}^{|\mathcal{F}|}$ is a random vector of length $|\mathcal{F}|$ whose elements conform with $\mathcal{N}(0, 0.2)$, and $\boldsymbol{\epsilon} \in \mathbb{R}^N$ is an iid noise vector following $\mathcal{N}(0, 5 \times 10^{-3})$. We set $|\mathcal{F}| = 100$ for all the experiments.

The proposed method uses the kernel $g(\lambda) = \exp(-s\lambda)$ with $s = \nu p_e p_s p_f / \lambda_{\max}$, where $\nu \in \mathbb{R}_+$ is a parameter, $p_e :=$
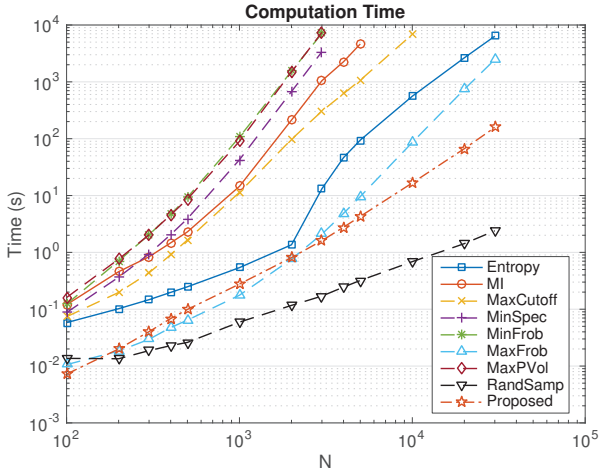
Fig. 2. Execution time comparison for random sensor graph. Note that both axes are represented on logarithmic scales.

$|\mathcal{E}|/N$ is the edge probability, $p_s := |\mathcal{S}|/N$ is the sampling ratio, and $p_f := |\mathcal{F}|/N$ is the (normalized) bandwidth. For all the experiments, ν was experimentally set to $\nu = 220$.[3] During the selection process, $g(\lambda)$ is approximated with Chebyshev polynomial approximation with the order $P = 12$. For the signal prediction, we fixed $k = 12$.

All the experiments were performed in MATLAB R2017a, running on a PC with an Intel Xeon E5 3 GHz CPU and 64 GB RAM. The MATLAB toolbox for submodular function optimization [30], [35] was used for implementations of the entropy and MI criteria.

B. Execution Time

First, we compare the execution time for choosing $|\mathcal{S}| = N/10$ vertices for various N. Figure 2 shows the execution time comparison plotted against N for the random sensor graph. The results are given by the average of 10 independent runs.

Among the deterministic approaches, Entropy, MaxFrob, and the proposed method are faster than the other methods. Specifically, for $N = 2000$ (thus $|\mathcal{S}| = 200$), the speedup factor of our method with respect to an alternative method, i.e.,

$$\frac{\text{Comp. time of alternative method}}{\text{Comp. time of proposed method}}$$

is summarized in Table VI. The proposed method is > 100 times faster than the methods with high prediction accuracies (presented in the following sections): MI, MaxCutoff, Min-Spec, MinFrob, and MaxPVol.

Although Entropy and MaxFrob are very fast compared to the other alternative methods and their computation times are comparable to that of the proposed method, their performances on SSS and the signal value prediction are significantly worse than the other methods. RandSamp is significantly faster than

[3]The optimal ν differs for different graphs, but this ν works well for our experiments. The automatic parameter setting will be an interesting topic in the future.

the other methods, including the proposed method, because it is a "one-shot" algorithm. However, on average, the prediction performance of the proposed method outperforms that of RandSamp. These results are further discussed in the next subsections.

C. Comparison of Selected Vertices

We show the vertices selected using the alternative and proposed approaches. The original graph is a random sensor graph with $N = 500$, and we select $|\mathcal{S}| = 10$ vertices. The underlying graph and sampling results are shown in Fig. 3. Because RandSamp selects different vertices in every selection process, we show two sampling realizations from the same sampling distribution probability.

It can be observed that MI, MaxCutoff, MinSpec, MinFrob, MaxPVol, and the proposed method select evenly distributed sampling positions (and the proposed method is the fastest among the six). Entropy selects many vertices at the corners or boundaries of the graph because they have fewer connections with other vertices. Because MaxFrob does not take into account the position of the already-selected vertices in each iteration, it often selects a vertex very close to an already selected one, as can be seen in Fig. 3(g) (some of them are almost overlapped). It therefore leads to large reconstruction errors, as shown in the next subsection. Two sampling sets of RandSamp are quite different from each other, and sometimes vertices very close to each other are selected, as shown in Fig. 3(i).

D. Comparison of Reconstruction Errors

We compare the prediction errors between the proposed and alternative SSS methods. Owing to the long execution times, all the SSS methods have been compared only for random sensor and ER graphs. For the other graphs, we compare the performances of the relatively fast methods: Entropy, MaxFrob, RandSamp, and the proposed method. The results are the average of 100 runs.

The GP-based methods, MaxCutoff [16] and MinSpec [14] use the reconstruction shown in (10) where $\mathcal{F}$ is the set of Laplacian eigenvalues less than λ_F (for MinSpec [14]) or the estimated cutoff frequency $\Omega_k(\mathcal{S})$ (for the other methods). For MaxCutoff, $k = 14$ is used because it has been effective for the reconstruction of noisy bandlimited signals [16]. For GP-based methods, $k = 6$ is used because it presents better performance in our experiments. MinFrob, MaxFrob, and MaxPVol [21] use the reconstruction shown in (15), and the proposed method uses the reconstruction shown in (10). RandSamp reconstructs the signal with the same method as the original paper [19]: The quadratic equation with the Laplacian smoothness regularizer term. The regularizer function $g_{\text{reg}}(\lambda)$ we used is that proposed in [19], where $g_{\text{reg}}(\lambda) = \lambda^4$.

The average MSEs between the predicted and original signals are summarized in Fig. 4. It is clear that the proposed method presents the lowest MSEs for almost all cases, with the exception of the community graph. The MSEs of RandSamp greatly depend on the specific graphs: It is good for the ER, random regular, and BA graphs, whereas its MSEs are not
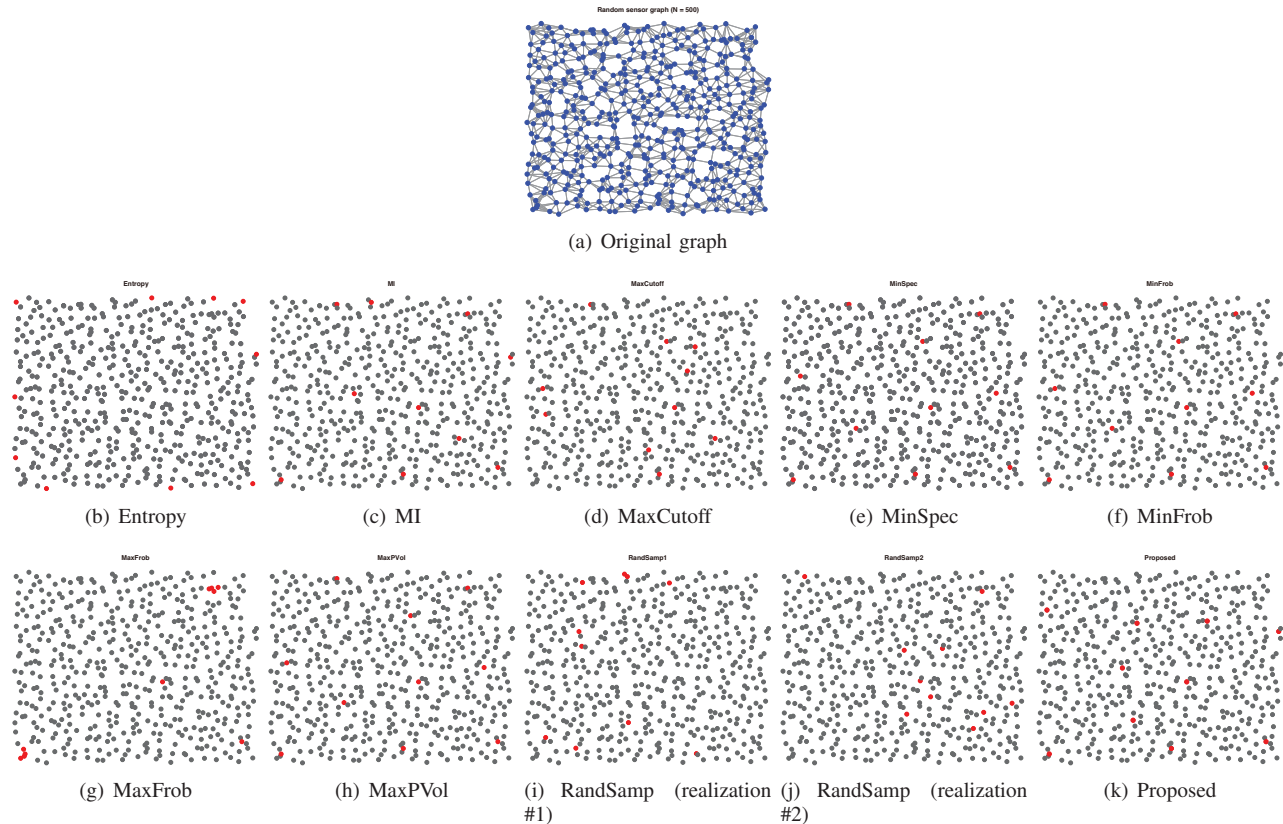
(a) Original graph

(b) Entropy   (c) MI   (d) MaxCutoff   (e) MinSpec   (f) MinFrob

(g) MaxFrob   (h) MaxPVol   (i) RandSamp (realization #1)   (j) RandSamp (realization #2)   (k) Proposed

Fig. 3. Selected vertices for a random sensor graph ($N = 500$). Ten vertices are selected, which are colored in red.

(a) Random Sensor   (b) Erdős–Rényi   (c) Random Regular

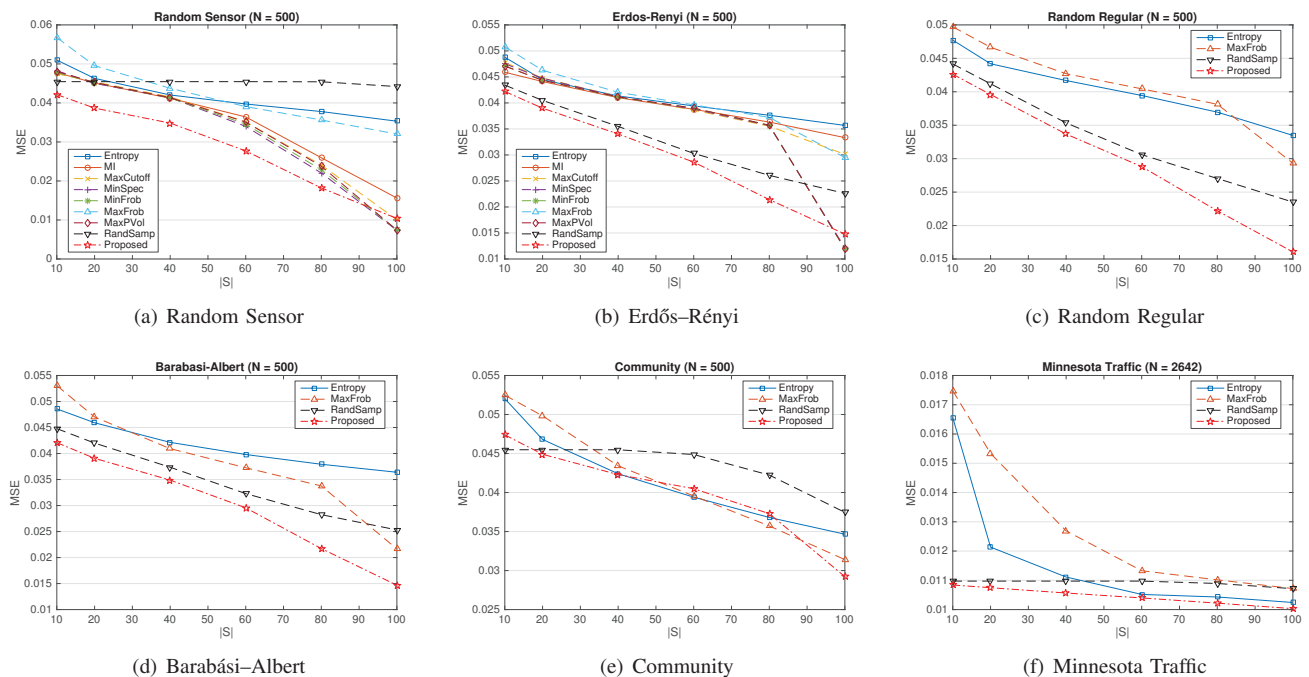(d) Barabási–Albert   (e) Community   (f) Minnesota Traffic

Fig. 4. MSE comparison of estimated signals (average of 100 tested signals).

improved for the random sensor, community, and Minnesota Traffic graphs, even when we select a large number of vertices. This could be due to the reconstruction parameter γ. Graph sampling theory-based approaches perform well when enough samples, i.e., close to the cutoff frequency, are selected. In contrast, they have larger MSEs than the proposed method for small $|\mathcal{S}|$.

VI. CONCLUSION

We proposed a SSS method based on the localization operator for graph signals. The proposed method has strong connections with the alternative GP-based sensor selection and the graph frequency-based SSS. The proposed SSS does not need the eigendecomposition of the graph Laplacian, whereas it still considers the graph frequency information as well as vertex information. It is significantly faster than the existing approaches, and its performance is better than those that are proven through numerical experiments.

APPENDICES

A. Derivation of Alternative Approaches: Sensor Position Selection

1) Entropy: In (4), the sensors are selected such that the uncertainty of a measurement with respect to previous measurements is maximized [23], [24], [26]:

$$
\begin{aligned}
\mathcal{S}^* &= \operatorname*{arg\,min}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} H(\boldsymbol{f}_{\mathcal{S}^c}|\boldsymbol{f}_{\mathcal{S}}) \\
&= \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} H(\boldsymbol{f}_{\mathcal{S}}) = \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \log \det [\mathbf{K}_{\mathcal{S}}],
\end{aligned}
\tag{50}
$$

where $H(\cdot)$ is (conditional) entropy.

Because the problem in (4) is NP-complete, a greedy algorithm has been proposed in [23], [24]. We first set $\mathcal{S} = \emptyset$ and add a sensor, which ensures the maximum increase in the uncertainty of the observed sensors, to $\mathcal{S}$ from the set of unselected sensors $\mathcal{S}^c$ one by one.

The entropy of the random variable $f(y)$, where y is the sensor of interest, conditioned on the variable $\boldsymbol{f}_{\mathcal{S}}$ is a monotonic function of its variance:

$$
H(f(y)|\boldsymbol{f}_{\mathcal{S}}) = \frac{1}{2} \log(2\pi e (\mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}}\mathbf{K}_{\mathcal{S}}^{-1}\mathbf{K}_{\mathcal{S}y})). \tag{51}
$$

Hence, the vertex that satisfies (5) is selected at each step.

2) MI: (6) maximizes the MI between the selected locations $\mathcal{S}$ and unselected locations $\mathcal{S}^c$, i.e., it selects the locations that more significantly reduce the uncertainty of the rest of the space [25], [26]:

$$
\begin{aligned}
\mathcal{S}^* &= \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} H(\boldsymbol{f}_{\mathcal{S}^c}) - H(\boldsymbol{f}_{\mathcal{S}^c}|\boldsymbol{f}_{\mathcal{S}}) \\
&= \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \log \det [\mathbf{K}_{\mathcal{S}}] + \log \det [\mathbf{K}_{\mathcal{S}^c}].
\end{aligned}
\tag{52}
$$

From (51) and (52), a greedy method [25] that adds sensor y^* satisfying (7) is used for the optimization.

B. Derivation of Alternative Approaches: Graph Sampling Based on Graph Fourier Basis

1) Based on Cutoff Frequency (MaxCutoff): [16] introduces a measure of quality for the sampling sets, namely the cutoff frequency, and selects the sampled vertices to maximize the cutoff frequency. It guarantees unique reconstruction and does not need the calculation of the graph Fourier basis.

The cutoff frequency associated with the subset $\mathcal{S}$ is a bound on the maximum frequency of a signal that can be perfectly recovered from the samples on the subset $\mathcal{S}$. Let us denote by $PW_\omega(\mathcal{G}) \in \mathbb{R}^N$ the *Paley–Wiener space* which is the space of all ω-bandlimited signals, by $L_2(\mathcal{S}^c)$ the space of signals with zero values on $\mathcal{S}$, i.e., if $\phi \in L_2(\mathcal{S}^c)$ then $\phi = [\boldsymbol{f}_{\mathcal{S}^c}^T \ \mathbf{0}^T]^T$, and by $\omega(\phi)$ the minimum eigenvalue of ϕ that have non-zero graph Fourier coefficients. [10] states the sampling theorem for graph signals as follows.

Theorem 2 (Graph Sampling Theorem [10, Theorem 2]). *The signal on a graph can be perfectly reconstructed from signal values $\boldsymbol{f}_{\mathcal{S}}$ on $\mathcal{S}$ if and only if $\boldsymbol{f} \in PW_\omega(\mathcal{G})$, where*

$$
\omega < \omega_c(\mathcal{S}) := \inf_{\phi \in L_2(\mathcal{S}^c)} \omega(\phi), \tag{53}
$$

and $\omega_c(\mathcal{S})$ is the exact cutoff frequency.

To avoid the calculation of the true cutoff frequency $\omega_c(\mathcal{S})$, which needs the computation of the graph Fourier basis, we can use the estimated cutoff frequency $\Omega_k(\mathcal{S})$ for the sampling set $\mathcal{S}$:

$$
\Omega_k(\mathcal{S}) = \inf_{\phi \in L_2(\mathcal{S}^c)} \left(\frac{\phi^T \mathbf{L}^k \phi}{\phi^T \phi} \right), \tag{54}
$$

where $k \in \mathbb{Z}^+$ is a parameter that provides a trade-off between performance and complexity. A large k leads the estimated cutoff frequency to be close to the actual bandwidth. (8) selects vertices so as to maximize the estimated cutoff frequency in (54):

$$
\begin{aligned}
\mathcal{S}^* &= \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \Omega_k(\mathcal{S}) \\
&= \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \min_{\psi} \frac{\psi^T (\mathbf{L}^k)_{\mathcal{S}^c} \psi}{\psi^T \psi} \\
&= \operatorname*{arg\,max}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \mu_{\min}((\mathbf{L}^k)_{\mathcal{S}^c}).
\end{aligned}
\tag{55}
$$

2) Based on Error Minimization: The objective functions shown in (11) and (13) are determined so as to minimize the reconstruction error. It assumes that the measured signal is corrupted by noise and/or not bandlimited; that is, the measured signal is defined as $\boldsymbol{o} = \boldsymbol{f}_{\mathcal{S}} + \boldsymbol{n}_{\mathcal{S}}$. This method also uses the reconstruction method in (10); then, the reconstruction error becomes $\boldsymbol{e} = \boldsymbol{f} - \mathbf{U}_{\mathcal{V}\mathcal{F}}(\mathbf{U}_{\mathcal{S}\mathcal{F}})^+(\boldsymbol{f}_{\mathcal{S}} + \boldsymbol{n}_{\mathcal{S}}) = \mathbf{U}_{\mathcal{V}\mathcal{F}}(\mathbf{U}_{\mathcal{S}\mathcal{F}})^+ \boldsymbol{n}_{\mathcal{S}}$.

- *MinSpec:* (11) is obtained by minimizing the ℓ_2 norm of the error:

$$
\begin{aligned}
\mathcal{S}^* &= \operatorname*{arg\,min}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \|\boldsymbol{e}\|_2 \\
&= \operatorname*{arg\,min}_{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F} \|\mathbf{U}_{\mathcal{V}\mathcal{F}}(\mathbf{U}_{\mathcal{S}\mathcal{F}})^+ \boldsymbol{n}_{\mathcal{S}}\|_2.
\end{aligned}
\tag{56}
$$

Its objective function can be further written as follows.

$$\|\mathbf{U}_{\mathcal{VF}}(\mathbf{U}_{\mathcal{SF}})^+\boldsymbol{n}_{\mathcal{S}}\|_2 \le \|\mathbf{U}_{\mathcal{VF}}\|_2\|(\mathbf{U}_{\mathcal{SF}})^+\|_2\|\boldsymbol{n}_{\mathcal{S}}\|_2$$
$$= \|(\mathbf{U}_{\mathcal{SF}})^+\|_2$$
$$= \sigma_{\min}(\mathbf{U}_{\mathcal{SF}}).$$
$$(57)$$

This leads to (11).

- *MinTrac:* (13) minimizes the mean squared errors, i.e., minimizes the trace of the error covariance matrix $\mathbf{E}$:

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \mathrm{tr}[\mathbf{E}]$$
$$= \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \mathrm{tr}[\boldsymbol{e}\boldsymbol{e}^H]$$
$$= \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \mathbf{U}_{\mathcal{VF}}(\mathbf{U}_{\mathcal{SF}}^H\mathbf{U}_{\mathcal{SF}})^{-1}\mathbf{U}_{\mathcal{VF}}$$
$$= \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \mathrm{tr}[(\mathbf{U}_{\mathcal{SF}}^H\mathbf{U}_{\mathcal{SF}})^{-1}].$$
$$(58)$$

3) Based on Localized Basis: [21] uses the basis localized both in the vertex and graph frequency domains for graph sampling and signal recovery. The basis $\boldsymbol{\psi}_i, i = 0, 1, \dots, N-1$ perfectly localized in the graph frequency domain and highly localized in the vertex domain, is designed by solving the following problem:

$$\boldsymbol{\psi}_i = \underset{\boldsymbol{\psi}_i}{\arg\max} \ \|\mathbf{D}_{\mathrm{ver}}\boldsymbol{\psi}_i\|_2$$
$$\text{s. t. } \|\boldsymbol{\psi}_i\|_2 = 1, \mathbf{D}_{\mathrm{sp}}\boldsymbol{\psi}_i = \boldsymbol{\psi}_i,$$
$$\langle\boldsymbol{\psi}_i, \boldsymbol{\psi}_j\rangle = 0, j = 1, \dots, i-1.$$
$$(59)$$

Its optimal solution coincides with the eigenvectors of $\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}}$, i.e., $\boldsymbol{\psi}_i = \boldsymbol{v}_i(\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})$.

The reconstruction shown in (15) indicates that the sampled signal $\mathbf{D}_{\mathrm{ver}}\boldsymbol{f}$ is interpolated by $\boldsymbol{\psi}_i$ for recovering the original signal:

$$\widehat{\boldsymbol{f}} = \sum_{i=0}^{|\mathcal{F}|-1} \frac{1}{\sigma_i^2(\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})}\langle\mathbf{D}_{\mathrm{ver}}\boldsymbol{f}, \boldsymbol{\psi}_i\rangle\boldsymbol{\psi}_i$$
$$= \boldsymbol{\Psi}_{\mathcal{VF}}\boldsymbol{\Sigma}_{\mathcal{FF}}^{-1}\boldsymbol{\Psi}_{\mathcal{VF}}^H\mathbf{D}_{\mathrm{ver}}\boldsymbol{f}$$
$$= (\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+\mathbf{D}_{\mathrm{ver}}\boldsymbol{f}.$$
$$(60)$$

(16)–(19) reduce the error caused by noise which can be written as $\boldsymbol{e} = \boldsymbol{f} - \widetilde{\boldsymbol{o}} = \boldsymbol{f} - (\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+\mathbf{D}_{\mathrm{ver}}(\boldsymbol{f} + \boldsymbol{n}) = (\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+\mathbf{D}_{\mathrm{ver}}\boldsymbol{n}$ where $\widetilde{\boldsymbol{o}}$ is the sampled signal with zero interpolation: $\widetilde{\boldsymbol{o}}_{\mathcal{S}} = \boldsymbol{o}_{\mathcal{S}}$ and $\widetilde{\boldsymbol{o}}_{\mathcal{S}^c} = \mathbf{0}_{|\mathcal{S}^c|}$.

- *MinFrob:* (16) is obtained by minimizing the Frobenius norm of the error $(\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+$:

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \|(\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+\mathbf{D}_{\mathrm{ver}}\boldsymbol{n}\|_F.$$
$$(61)$$

Its objective function has the following relationship:

$$\|(\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}})^+\mathbf{D}_{\mathrm{ver}}\boldsymbol{n}\|_F$$
$$\le \|(\mathrm{diag}(\mathbf{1}_{\mathcal{F}})\mathbf{U}^H\mathbf{D}_{\mathrm{ver}})^+\|_F\|\mathbf{D}_{\mathrm{ver}}\boldsymbol{n}\|_F$$
$$= \sum_{i=0}^{|\mathcal{F}|-1} \frac{1}{\lambda_i(\mathbf{U}_{\mathcal{SF}}^H)}.$$
$$(62)$$

It leads to (16).

- *MaxFrob:* (18) is the approximation of (i) and maximizes the Frobenius norm of $\mathbf{D}_{\mathrm{sp}}\mathbf{D}_{\mathrm{ver}}\mathbf{D}_{\mathrm{sp}}$. It does not need any

eigendecomposition of the variation operator for solving the problem.

- *MaxPVol:* (19) maximizes the volume of the parallelepiped formed with the columns of $\mathbf{U}_{\mathcal{SF}}^H$ which can be computed by the determinant of $\mathbf{U}_{\mathcal{SF}}^H\mathbf{U}_{\mathcal{SF}}$.

C. Derivations of Objective Functions in Section III-B

The spectral norm of the error covariance matrix can be bounded as

$$\|\mathbf{E}\|_2 = \|\mathbf{T}^{k/2}((\mathbf{T}^{k/2})_{\mathcal{SV}}^H(\mathbf{T}^{k/2})_{\mathcal{SV}})^+\mathbf{T}^{k/2}\|_2$$
$$\le \|\mathbf{T}^{k/2}\|_2\|((\mathbf{T}^{k/2})_{\mathcal{SV}})^+((\mathbf{T}^{k/2})_{\mathcal{SV}})^+)^H\|_2\|\mathbf{T}^{k/2}\|_2.$$
$$(63)$$

Therefore, the objective function for minimizing the error covariance matrix can be represented as follows:

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \|\mathbf{E}\|_2$$
$$= \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \|((\mathbf{T}^{k/2})_{\mathcal{SV}})^+\|_2$$
$$(64)$$

Its objective function has the following relationship:

$$\|((\mathbf{T}^{k/2})_{\mathcal{SV}})^+\|_2 = \le \|((\mathbf{T}^{k/2})_{\mathcal{SV}})^+\|_F \qquad (65)$$
$$= \sqrt{\mathrm{tr}\left[((\mathbf{T}^{k/2})_{\mathcal{SV}}^+)^H(\mathbf{T}^{k/2})_{\mathcal{SV}}^+\right]}$$
$$= \sqrt{\mathrm{tr}\left[(\mathbf{T}^{-k})_{\mathcal{S}}\right]} \qquad (66)$$

Similarly, starting from

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\max} \ \|(\mathbf{T}^{k/2})_{\mathcal{SV}}\|_2, \qquad (67)$$

we have the following objective function:

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\max} \ \sqrt{\mathrm{tr}\left[(\mathbf{T}^k)_{\mathcal{S}}\right]}. \qquad (68)$$

The above objective functions minimize the spectral norm (64), Frobenius norm (65), or trace (66) of the error covariance matrix. Furthermore, because $(g(\mathbf{\Lambda}_{\mathcal{F}^c})/\beta)^k \approx \mathbf{0}_{|\mathcal{F}^c|}$ for large k, the determinant of the error covariance matrix becomes

$$\det[\mathbf{E}]$$
$$= \det[\mathbf{T}^{k/2}(\mathbf{U}g(\mathbf{\Lambda})^{k/2}\mathbf{U}_{\mathcal{SV}}^H\mathbf{U}_{\mathcal{SV}}g(\mathbf{\Lambda})^{k/2}\mathbf{U}^H)^+\mathbf{T}^{k/2}]$$
$$= \det[\mathbf{T}^{k/2}\mathbf{U}(g(\mathbf{\Lambda}_{\mathcal{F}})^{k/2}\mathbf{U}_{\mathcal{SF}}^H\mathbf{U}_{\mathcal{SF}}g(\mathbf{\Lambda}_{\mathcal{F}})^{k/2})^+\mathbf{U}^H\mathbf{T}^{k/2}]$$
$$= \det[\mathbf{T}^{k/2}\mathbf{U}]\det[(g(\mathbf{\Lambda}_{\mathcal{F}})^{k/2}\mathbf{U}_{\mathcal{SF}}^H\mathbf{U}_{\mathcal{SF}}g(\mathbf{\Lambda}_{\mathcal{F}})^{k/2})^+]$$
$$\times \det[\mathbf{U}^H\mathbf{T}^{k/2}]$$
$$= \det[\mathbf{T}^{k/2}\mathbf{U}]\det[(\mathbf{U}_{\mathcal{SF}}g(\mathbf{\Lambda}_{\mathcal{F}})^k\mathbf{U}_{\mathcal{SF}}^H)^+]\det[\mathbf{U}^H\mathbf{T}^{k/2}]$$
$$= \det[\mathbf{T}^{k/2}\mathbf{U}]\det[(\mathbf{T}_{\mathcal{S}})^{-1}]\det[\mathbf{U}^H\mathbf{T}^{k/2}].$$
$$(69)$$

The objective function to minimize the determinant of the error covariance matrix can be represented as

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\min} \ \det[\mathbf{E}] = \det[(\mathbf{T}_{\mathcal{S}})^{-1}]. \qquad (70)$$

D. Proof of Proposition 1

MaxCutoff also can be regarded as the SSS for minimizing the error caused by the reconstruction with $\mathbf{T} = (\mathbf{L} + \delta\mathbf{I})^{-1}$. Because $A_T\sigma_{\min}(\mathbf{T}_\mathcal{S}) \leq \sigma_{\min}((\mathbf{T}^{-1})_{\mathcal{S}^c}) \leq B_T\sigma_{\min}(\mathbf{T}_\mathcal{S})$ [36], where A_T and B_T are some constant values determined from $\mathbf{T}$, (64) can be lower-bounded as

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}:|\mathcal{S}|=F}{\arg\max} \sigma_{\min}((\mathbf{T}^k)_\mathcal{S}). \tag{71}$$

Its objective function has the following relationship:

$$\begin{aligned} \sigma_{\min}((\mathbf{T}^k)_\mathcal{S}) &\geq \sigma_{\min}((\mathbf{T}^{-k})_{\mathcal{S}^c}) \\ &= \sigma_{\min}(((\mathbf{L} + \delta\mathbf{I})^k)_{\mathcal{S}^c}). \end{aligned} \tag{72}$$

When δ goes to zero, (72) becomes

$$\lim_{\delta \to 0} \sigma_{\min}(((\mathbf{L} + \delta\mathbf{I})^k)_{\mathcal{S}^c}) = \sigma_{\min}((\mathbf{L}^k)_{\mathcal{S}^c}). \tag{73}$$

As a result, maximizing the objective function of MaxCutoff shown in (8) leads to the minimization of the spectral norm of the error covariance matrix, where the error is caused by our reconstruction method in (21).

REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, Oct. 2013.

[2] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 1644–1656, Apr. 2013.

[3] ——, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, 2014.

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

[5] M. Crovella and E. Kolaczyk, "Graph wavelets for spatial traffic analysis," in *Proc. INFOCOM'03*, vol. 3, 2003, pp. 1848–1857.

[6] H. Higashi, T. M. Rutkowski, T. Tanaka, and Y. Tanaka, "Multilinear discriminant analysis with subspace constraints for single-trial classification of event-related potentials," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1295–1305, Oct. 2016.

[7] W. Huang, T. A. W. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proc. IEEE*, vol. 106, no. 5, pp. 868–885, May 2018.

[8] G. Shen and A. Ortega, "Transform-based distributed data gathering," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3802–3815, 2010.

[9] N. Leonardi and D. Van De Ville, "Tight wavelet frames on multislice graphs," *IEEE Trans. Signal Process.*, vol. 16, no. 13, pp. 3357–3367, Jul. 2013.

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

[11] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Proc. ICASSP'13*, 2013, pp. 5445–5449.

[12] A. Gadde, A. Anis, and A. Ortega, "Active semi-supervised learning using sampling theory for graph signals," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2014, pp. 492–501.

[13] A. Gadde and A. Ortega, "A probabilistic interpretation of sampling theory of graph signals," in *Proc. IEEE Conf. Acoust. Speech, Signal Process.*, 2015, pp. 3257–3261.

[14] 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, Dec. 2015.

[15] H. Shomorony and A. S. Avestimehr, "Sampling large data on graphs," in *Proc. GlobalSIP'14*, 2014, pp. 933–936.

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

[17] 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, 2016.

[18] Y. Tanaka, "Spectral domain sampling of graph signals," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3752–3767, Jul. 2018.

[19] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "Random sampling of bandlimited signals on graphs," *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, Mar. 2018.

[20] N. Perraudin, B. Ricaud, D. I. Shuman, and P. Vandergheynst, "Global and local uncertainty principles for signals on graphs," *APSIPA Transactions on Signal and Information Processing*, vol. 7, p. e3, 2018.

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

[22] A. Jayawant and A. Ortega, "A distance-based formulation for sampling signals on graphs," in *Proc. Int. Conf. Acoust. Speech, Signal Process.*, 2018, pp. 6318–6322.

[23] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 2015.

[24] M. C. Shewry and H. P. Wynn, "Maximum entropy sampling," *Journal of applied statistics*, vol. 14, no. 2, pp. 165–170, 1987.

[25] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[26] D. Sharma, A. Kapoor, and A. Deshpande, "On greedy maximization of entropy," in *Proc. Int. Conf. Machine Learn.*, 2015, pp. 1330–1338.

[27] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Efficient sensor position selection using graph signal sampling theory," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2016, pp. 6225–6229.

[28] ——, "Accelerated sensor position selection using graph localization operator," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2017, pp. 5890–5894.

[29] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *Proc. Thirtieth international conference on Very large data bases*, 2004, pp. 588–599.

[30] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proc. AAAI*, vol. 7, 2007, pp. 1650–1654.

[31] M. S. Uddin, A. Kuh, A. Kavcic, and T. Tanaka, "Nested performance bounds and approximate solutions for the sensor placement problem," *APSIPA Trans. Signal and Information Processing*, vol. 3, no. e4, 2014.

[32] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, 1997.

[33] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, Mar. 2011. [Online]. Available: http://wiki.epfl.ch/sgwt

[34] N. D. Lawrence, "A unifying probabilistic perspective for spectral dimensionality reduction: Insights and new models," *Journal of Machine Learning Research*, vol. 13, no. May, pp. 1609–1638, 2012.

[35] A. Krause, "SFO: A toolbox for submodular function optimization," *Journal of Machine Learning Research*, vol. 11, no. Mar., pp. 1141–1144, 2010.

[36] W. Govaerts and J. D. Pryce, "A singular value inequality for block matrices," *Linear Algebra and its Applications*, vol. 125, pp. 141–148, 1989.