EFFICIENT SENSOR POSITION SELECTION USING GRAPH SIGNAL SAMPLING THEORY

Akie Sakiyama[1]*, Yuichi Tanaka*[1]*, Toshihisa Tanaka*[1] *and Antonio Ortega*[1,2]

[1]Tokyo University of Agriculture and Technology, Tokyo, Japan
[2]University of Southern California, Los Angeles, USA

ABSTRACT

We consider the problem of selecting optimal sensor placements. The proposed approach is based on the sampling theorem of graph signals. We choose sensors that maximize the graph cut-off frequency, i.e., the most informative sensors for predicting the values on unselected sensors. We study the existing methods in the context of graph signal processing and clarify the relationship between these methods and the proposed approach. The effectiveness of our approach is verified through numerical experiments, showing advantages in prediction error and execution time.

***Index Terms*—** Sensor placement, graph signal processing, graph sampling theorem, Gaussian process, mutual information

1. INTRODUCTION

Choosing the best sensor placements is one of the fundamental tasks in sensor networks and is useful for monitoring spatial phenomena such as temperature sensing [1], indoor air quality, rainfall and smart grid system [2]. In these applications, we often have to handle large number of sensors which are distributed nonuniformly.

The sensor locations are usually selected so as to minimize the number of sensors, or optimize performance for a given number of sensors, in order to obtain the best prediction of variables on unsensed locations. One approach to select optimal sensor locations assumes that the sensors can capture the data within a fixed distance of their location [3]. This approach places sensors so as to cover overall area with the fewest sensors. The problem is formulated as an art-gallery model [3, 4]. Another research assumes that the spatial phenomena are modeled as a Gaussian process (GP) and places the sensors at the most informative locations [5, 6]. In [7, 8], a subset of sensors is determined to minimize entropy of the unobserved locations, i.e., maximizes uncertainty with respect to existing observations. In [9], a method that maximizes mutual information (MI) [10] between the unobserved and observed locations was proposed, where the sensors are placed to reduce the uncertainty of the unobserved locations. In this paper, we consider the sensor placement problem as an application of the emerging signal processing technology: graph signal processing [11, 12].

Traditional signal processing is a key to process complex-structured data but sometimes it has limitations since it usually considers (discrete) signals on regular grids. In graph signal processing, the relationship between data points is represented as edges in an underlying graph. Therefore, graph signal processing enables us to efficiently analyze complex, irregular, and high-dimensional data. It can be used for various applications such as traffic [13], learning [14, 15], image [16–18], brain networks [19, 20] and sensor networks [21].

Existing sensor selection methods, i.e., the entropy and MI criteria, have not been considered from a graph signal processing perspective. We show that the sensor placement problem can be interpreted as a graph sampling problem. Therefore, we propose a sensor selection approach based on the sampling theory for graph signals [22–28]. Sampling theory for graph signals addresses the problem of recovering the signals on unknown nodes from only a part of the samples. If the original signal is band-limited in the graph spectral domain and its bandwidth is less than the cut-off frequency, it can be perfectly recovered from the given samples on the subset of nodes. Our approach regards the sensors and sensor data as nodes and graph signals, respectively, and selects the sensors so as to maximize the cut-off frequency of the selected sensors. It allows us to reconstruct signals that has higher cut-off frequency than the entropy and MI criteria. We use greedy heuristics for sensor selection which are similar to the algorithms optimizing the entropy or MI based cost functions. Furthermore, we show that the conventional approaches can be viewed as graph vertex domain operations, but do not provide easy spectral interpretations, whereas our approach does. We show the advantages of the proposed method through numerical experiments.

The remaining of this paper is organized as follows. Preliminaries on graph signal processing are summarized in the rest of this section. Section 2 gives the problem setting on GP models and reviews existing approaches for selecting optimal sensor placements. Section 3 describes the proposed sensor selection based on graph signal sampling theory, and clarifies the existing approaches from the perspective of graph signal processing. Performance comparisons are provided in Section 4. Finally, Section 5 concludes the paper.

1.1. Preliminaries

A graph is represented as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ and $\mathcal{E}$ denote sets of nodes and edges, respectively. The graph signal is defined as $\boldsymbol{f} \in \mathbb{R}^N$. We will only consider a connected, finite, undirected graph with no multiple edges. The number of nodes is $N = |\mathcal{V}|$, unless otherwise specified. The (m, n)-th element of the adjacency matrix $\mathbf{A}$ is the weight of the edge between m and n if m and n are connected, and 0 otherwise. The degree matrix $\mathbf{D}$ is a diagonal matrix and its mth diagonal element is $D(m, m) = \sum_n A(m, n)$. The unnormalized graph Laplacian matrix (GLM) is defined as $\mathbf{L} := \mathbf{D} - \mathbf{A}$ and the symmetric normalized GLM is $\boldsymbol{\mathcal{L}} := \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$. The symmetric normalized GLM has the property that its eigenvalues are within the interval $[0, 2]$. The eigenvalues of $\mathbf{L}$ or $\boldsymbol{\mathcal{L}}$ are λ_i and ordered as: $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \ldots \leq \lambda_{N-1} = \lambda_{\max}$ without loss of generality. The eigenvector $\boldsymbol{u}_i$ corresponds to λ_i and satisfies $\boldsymbol{\mathcal{L}}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$. The eigenvectors $\mathbf{U} = [\boldsymbol{u}_0 \ldots \boldsymbol{u}_{N-1}]$ satisfy $\mathbf{U}\mathbf{U}^\dagger =$

This work was supported in part by MIC SCOPE Grant Number 142103007.

$\mathbf{I}_N$ where $\cdot^\dagger$ is the conjugate transpose of a matrix or a vector and $\mathbf{I}_N$ is the $N \times N$ identity matrix. The entire spectrum of $\mathcal{G}$ is defined by $\sigma(\mathcal{G}) := \{\lambda_0, \ldots, \lambda_{N-1}\}$. The graph Fourier transform is defined as follows [29, 30]: $\overline{f}(\lambda_i) := \langle \boldsymbol{u}_i, \boldsymbol{f} \rangle = \sum_{n=0}^{N-1} u_i^*(n) f(n)$, where $\cdot^*$ is the complex conjugate. For a vector $\boldsymbol{x} \in \mathbb{R}^N$ and a set $\mathcal{A}$, the restriction of $\boldsymbol{x}$ to its components indexed by $\mathcal{A}$ is denoted by $\boldsymbol{x}_\mathcal{A}$. For a matrix $\mathbf{X} \in \mathbb{R}^{N \times N}$, $\mathbf{X}_{\mathcal{AB}}$ denotes the restriction matrix of $\mathbf{X}$, which is obtained from extracting rows indexed by $\mathcal{A}$ and columns indexed by $\mathcal{B}$ from $\mathbf{X}$, and $\mathbf{X}_\mathcal{A} := \mathbf{X}_{\mathcal{AA}}$.

2. SENSOR POSITION SELECTION

The following problem is considered in this paper. The sets of measured locations and possible locations are denoted by $\mathcal{S} \subseteq \mathcal{V}$ and $\mathcal{V} = \{v_0, \ldots, v_{N-1}\}$, respectively. We consider the problem of finding $|\mathcal{S}| = F$ sensors from $|\mathcal{V}| = N$ possible locations. We assume that the stochastic graph signal $\boldsymbol{f}$ has the following Gaussian joint zero-mean distribution [31]:

$$p(\boldsymbol{f}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathbf{K}|} \exp\left(-\frac{1}{2} \boldsymbol{f}^T \mathbf{K}^{-1} \boldsymbol{f}\right), \tag{1}$$

where $\cdot^T$ is the transpose of a matrix or a vector, $\mathbf{K} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ is the covariance matrix of all locations $\mathcal{V}$, its (i,j)th element is $\mathcal{K}(i,j)$ with a symmetric positive-definite kernel function $\mathcal{K}(\cdot,\cdot)$, and $|\mathbf{K}|$ is the determinant of $\mathbf{K}$. The GP model is a well accepted model for the sensor placement problem [1, 5, 6, 9]. 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 whose conditional variance is $\sigma^2_{y|\mathcal{S}} = \mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}} \mathbf{K}_{\mathcal{SS}}^{-1} \mathbf{K}_{\mathcal{S}y}$. We introduce two state-of-the-art techniques based on the GP model: Entropy and MI criteria.

2.1. Entropy Criterion

For the entropy criterion, the sensors are selected so that the uncertainty of a measurement with respect to previous measurements is maximized [7, 8]. The objective function is represented as

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = k}{\arg\min}\ H(\boldsymbol{f}_{\mathcal{S}^c} | \boldsymbol{f}_\mathcal{S}) = \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = k}{\arg\max}\ H(\boldsymbol{f}_\mathcal{S}), \tag{2}$$

where $\mathcal{S}^c = \mathcal{V} \setminus \mathcal{S}$. Since the problem in (2) is NP-complete, a greedy algorithm is proposed in [7, 8]. We firstly set $\mathcal{S} = \emptyset$ and add a sensor, which ensures the maximum increase of the uncertainty of the observed sensors, to $\mathcal{S}$ from the set of unselected sensors $\mathcal{S}^c$ one by one. The entropy of random variable $f(y)$, where y is the sensor of interest, conditioned on variable $\boldsymbol{f}_\mathcal{S}$ is a monotonic function of its variance: $H(f(y)|\boldsymbol{f}_\mathcal{S}) = \frac{1}{2}\log(2\pi e(\mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}}\mathbf{K}_{\mathcal{SS}}^{-1}\mathbf{K}_{\mathcal{S}y}))$. Hence, the node that satisfies the following equation is selected at each step:

$$y^* \leftarrow \underset{y \in \mathcal{S}^c}{\arg\max}\, \mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}} \mathbf{K}_{\mathcal{SS}}^{-1} \mathbf{K}_{\mathcal{S}y}. \tag{3}$$

2.2. Mutual Information Criterion

The MI criterion 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 [9]:

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{V}: |\mathcal{S}| = k}{\arg\max}\ H(\boldsymbol{f}_{\mathcal{S}^c}) - H(\boldsymbol{f}_{\mathcal{S}^c} | \boldsymbol{f}_\mathcal{S}) := \mathrm{MI}(\mathcal{S}). \tag{4}$$

Since this problem is also hard to optimize, a greedy method [9] is used for optimization that adds sensor y^* to $\mathcal{S}$ to maximize the MI at each step:

$$y^* \leftarrow \underset{y \in \mathcal{S}^c}{\arg\max}\ \frac{\mathcal{K}(y,y) - \mathbf{K}_{y\mathcal{S}} \mathbf{K}_{\mathcal{SS}}^{-1} \mathbf{K}_{\mathcal{S}y}}{\mathcal{K}(y,y) - \mathbf{K}_{y\bar{\mathcal{S}}} \mathbf{K}_{\bar{\mathcal{S}}\bar{\mathcal{S}}}^{-1} \mathbf{K}_{\bar{\mathcal{S}}y}}, \tag{5}$$

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

3. SENSOR POSITION SELECTION USING SAMPLING THEOREM FOR GRAPH SIGNALS

Optimal sensor selection can be viewed as the problem of selecting the most informative nodes to reconstruct graph signals, where the sensors and observed signals are viewed as the nodes and graph signals, respectively. The connection of nodes can be determined from the covariance matrix[1]. We assume that the random signals have following distributions:

$$\begin{aligned} p(\boldsymbol{f}) &\propto \exp\left(-\sum_i \sum_j A(i,j)(f(i) - f(j))^2 - \delta \sum_i f(i)^2\right) \\ &= \exp\left(-\boldsymbol{f}^T (\mathbf{L} + \delta \mathbf{I}) \boldsymbol{f}\right). \end{aligned} \tag{6}$$

From (1) and (6), the graph Laplacian matrix can be obtained from the inverse covariance matrix:

$$\mathbf{L} = \mathbf{K}^{-1} - \delta \mathbf{I}. \tag{7}$$

The parameter δ prevents the precision matrix from being singular. The precision matrix has the same set of eigenvectors $\{\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{N-1}\}$ as the unnormalized graph Laplacian matrix $\mathbf{L}$ with corresponding eigenvalues $\{\sigma_i = \frac{1}{\lambda_i + \delta}\}_{i=0,\ldots,N-1}$. Since $\boldsymbol{u}_0 = \mathbf{1}_N$ and $\lambda_0 = 0$, $\sigma_0 = 1/\delta$ is the variance of the DC component of $\boldsymbol{f}$.

3.1. Selection Algorithm

Our approach uses the sampling theorem of graph signals. It considers the problem of reconstructing the ω-bandlimited graph signals from their sub-sampled versions [24–27]. The ω-bandlimited graph signal has zero graph Fourier coefficients corresponding to the eigenvalues greater than ω: $\overline{f}(\lambda_i) = 0$ for $\lambda_i > \omega$. The space of all ω-bandlimited signals is called *Paley-Wiener space* and denoted as $PW_\omega(\mathcal{G}) \in \mathbb{R}^N$. The cut-off 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 $L_2(\mathcal{S}^c)$ the space of signals having 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. The sampling theorem for graph signals is stated as follows.

Theorem 1 (Graph Sampling Theorem [22, 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*

$$\omega < \omega_c(\mathcal{S}) := \inf_{\phi \in L_2(\mathcal{S}^c)} \omega(\phi), \tag{8}$$

and $\omega_c(\mathcal{S})$ is the exact cut-off frequency.

[1] We can also use a graph Laplacian estimated directly from prior information, if signals do not have the distribution (1).

The sensors are chosen so as to maximize the cut-off frequency in the graph spectral domain under the fixed number F of sensors:

$$\mathcal{S}^* = \arg\max_{\mathcal{S}} \omega_c(\mathcal{S}) \quad \text{subject to} \quad |\mathcal{S}| = F. \tag{9}$$

We use three techniques for optimizing the cost function (9): eigenvalue-based [22], singular value decomposition-based [26], and standard basis-based [27] algorithms (hereafter, they are denoted as EV, SVD and SB, respectively). All methods use a greedy heuristic that adds the sensor that maximizes the cut-off frequency at each iteration. The set of selected sensors $\mathcal{S}$ is firstly initiated to the empty set. The selection approach of additional node is different in each method:

a) EV: We add the node $y^* \leftarrow \arg\max_y [(\phi_k^*(y))^2]$ where ϕ_k^* is the eigenvector of $(\boldsymbol{\mathcal{L}}^k)_{\mathcal{S}^c}$ associated with the minimum value $\lambda_{\min}^*$ and $k \in \mathbb{Z}^+$ is a parameter. Each k provides a trade-off between performance and complexity.

b) SVD: We add the node $y^* \leftarrow \arg\max_y \sigma_{\min}(\mathbf{U}_{F(\mathcal{S}\cup y)})$ where $\sigma_{\min}(\mathbf{U}_{F(\mathcal{S}\cup y)})$ is the smallest singular value of $\mathbf{U}_{F(\mathcal{S}\cup y)}$.

c) SB: We initiate $[\boldsymbol{b}_0 \dots \boldsymbol{b}_{N-1}] = [\boldsymbol{e}_0 \dots \boldsymbol{e}_{N-1}]$ where $[\boldsymbol{e}_0 \dots \boldsymbol{e}_{N-1}]$ is standard basis of $\mathbb{R}^N$. Then, we add the node $y^* \leftarrow \arg\max_y |\alpha_y|$, where $\boldsymbol{u}_j = \sum_i \alpha_i \boldsymbol{b}_i$, $\boldsymbol{u}_j$ is the jth eigenvector of $\boldsymbol{\mathcal{L}}$ and j is the number of iteration, and update $\boldsymbol{b}_y = \boldsymbol{u}_y$ at each step.

In EV, we allow relaxations for the calculation of $\omega_c(\mathcal{S})$, since finding the exact cut-off frequency requires high computational costs. The estimated cut-off frequency $\Omega_k(\mathcal{S})$ for $\mathcal{S}$ can be obtained as

$$\Omega_k(\mathcal{S}) = (\lambda_{\min}^*)^{1/k} = \inf_{\phi \in L_2(\mathcal{S}^c)} \left(\frac{\phi^T \boldsymbol{\mathcal{L}}^k \phi}{\phi^T \phi} \right). \tag{10}$$

where $\lambda_{\min}^*$ is the minimum eigenvalue of $(\boldsymbol{\mathcal{L}}^k)_{\mathcal{S}^c}$. A large k leads to the estimated cut-off frequency to be close to the actual bandwidth. SVD and SB need to calculate the eigenvectors associated with the minimum F eigenvalues, which require additional complexity. However, the cut-off frequency of these methods is equivalent to $\omega_c(\mathcal{S}) = \lambda_{F-1}$, and therefore, these methods can result in good performance without the approximation of the cut-off frequency of (10) based on parameter k.

3.2. Relationships with Existing Methods

From the block matrix inversion formula, the inversion of the covariance matrix can be represented as [25]:

$$\begin{aligned} \mathbf{K}^{-1} &= \begin{bmatrix} \mathbf{K}_{\mathcal{S}^c} & \mathbf{K}_{\mathcal{S}^c\mathcal{S}} \\ \mathbf{K}_{\mathcal{S}\mathcal{S}^c} & \mathbf{K}_{\mathcal{S}} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{K}_{\mathcal{S}^c|\mathcal{S}}^{-1} & -(\mathbf{K}_{\mathcal{S}^c})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}\mathbf{K}_{\mathcal{S}|\mathcal{S}^c}^{-1} \\ -(\mathbf{K}_{\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}^T\mathbf{K}_{\mathcal{S}^c|\mathcal{S}}^{-1} & \mathbf{K}_{\mathcal{S}|\mathcal{S}^c}^{-1} \end{bmatrix}, \end{aligned} \tag{11}$$

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 $\mathbf{K}_{\mathcal{S}|\mathcal{S}^c} = \mathbf{K}_{\mathcal{S}} - \mathbf{K}_{\mathcal{S}\mathcal{S}^c}(\mathbf{K}_{\mathcal{S}^c})^{-1}\mathbf{K}_{\mathcal{S}\mathcal{S}^c}^T$. By using (7) and (11), the graph Laplacian and the covariance matrix have a following relationship:

$$\mathbf{L}_{\mathcal{S}^c} + \delta\mathbf{I} = (\mathbf{K}_{\mathcal{S}^c} - \mathbf{K}_{\mathcal{S}^c\mathcal{S}}(\mathbf{K}_{\mathcal{S}})^{-1}\mathbf{K}_{\mathcal{S}^c\mathcal{S}}^T)^{-1}. \tag{12}$$

To clarify the characteristic of conventional approaches from a graph signal processing perspective, we consider a toy example

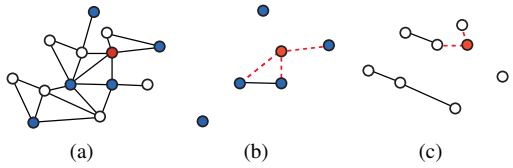

Fig. 1. (a) Original graph. The blue nodes and red node indicate $\mathcal{S}$ and y, respectively. (b) $\mathbf{L}^y$. $L^y(y,y)$ is the total weight of red dashed edges. (c) $\bar{\mathbf{L}}^y$. $\bar{L}^y(y,y)$ is the total weight of red dashed edges.

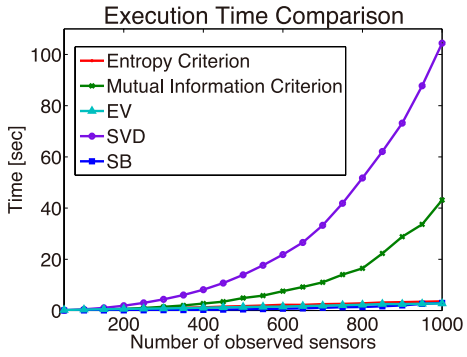

Fig. 2. Execution time comparison (Average of 10 executions).

shown in Fig. 1. We use a synthesized simple graph in this figure for the sake of clarity. From (12), we can rewrite the entropy criterion in (3) as:

$$y^* \leftarrow \arg\max_{y \in \mathcal{S}^c} \frac{1}{L^y(y,y) + \delta^y}, \tag{13}$$

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

The MI criterion in (5) can also be rewritten as

$$y^* \leftarrow \arg\max_{y \in \mathcal{S}^c} \frac{\bar{L}^y(y,y) + \bar{\delta}^y}{L^y(y,y) + \delta^y}, \tag{14}$$

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

Note that the proposed method and the MI criterion select the same sensor at the first iteration, when we use (7) for constructing the graph Laplacian. This is because all proposed algorithms select the node having the maximum element in the eigenvector corresponding to the minimum eigenvalue $u_0(m) = \sqrt{D(m,m)}$ of the normalized graph Laplacian matrix, i.e., the sensor having the maximum degree in the graph is selected as the first selected sensor. On the other hand, the entropy criterion selects a random sensor at the first iteration, because its cost function in (13) is constant at all locations. It is interesting to note that the conventional entropy and

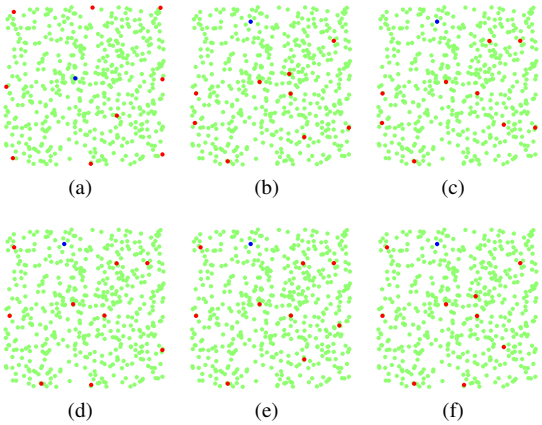

Fig. 3. 10 selected locations from 500 locations. The red nodes are selected sensors. The blue node indicates the first selected sensor. (a) Entropy ($\Omega(\mathcal{S}) = 0.023$). (b) MI ($\Omega(\mathcal{S}) = 0.026$). (c) EV with $k = 1$ ($\Omega(\mathcal{S}) = 0.026$). (d) EV with $k = 8$ ($\Omega(\mathcal{S}) = 0.030$). (e) SVD ($\omega_c(\mathcal{S}) = 0.027$). (f) SB ($\omega_c(\mathcal{S}) = 0.027$).

MI criterion select sensors according to the edge information in the *graph vertex domain*, whereas the proposed method selects sensors while considering the frequency in the *graph spectral domain*.

4. EXPERIMENTAL RESULTS

The proposed sensor selection is compared with the existing entropy criterion [7, 8] and the MI criterion [9] through numerical experiments. Although the kernel in (1) can be arbitrarily chosen, all experiments use the following stationary kernel[2]: $\mathcal{K}(i,j) = \exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|/\theta\right)$, where $\boldsymbol{x}_i \in \mathbb{R}^2$ is the coordinate of the ith node v_i and θ is a parameter. $\theta = 1$ is used for all experiments. We do not have a model of covariance because it is assumed that we do not have other observed data, so we only use distance to create a graph. All experiments were performed in Matlab R2013a, running on a PC with Intel Xeon E5 3 GHz CPU and 64 GB RAM. Matlab toolbox for submodular function optimization [32, 33] is used for implementations of the entropy and MI criteria. The proposed methods use a graph Laplacian which is constructed by (7) and removed self loops and negative edges.

We compare the execution time for choosing various number of locations from randomly generated locations. $|\mathcal{S}| = |\mathcal{V}|/10$ sensors are selected with various number of possible locations $\mathcal{V}$. We experimentally use $k = 8$ for EV. Figure 2 shows the execution time comparison plotted against $|\mathcal{V}|$. From this figure, we can see that the MI criterion and SVD are very slow. The EV and SB are faster than the entropy criterion.

Next, we compare the selected sensor positions. In this experiment, 10 sensor locations are selected from 500 randomly generated locations by using each method. The results of sensor selections are shown with their (estimated) cut-off frequency in Fig. 3. It can be seen that both the proposed methods and the MI criterion select similar locations. As previously mentioned, they selected the same locations at the first iteration. The selected sensors by using the entropy criterion are placed close to the boundaries of the space.

Finally, we predict graph signal values on the unobserved locations and compare the reconstruction errors. Tested signals are randomly generated according to the GP model[3] by using GPML toolbox [34] and are corrupted by the additive white Gaussian noise with $\sigma = 0.02$. We use $k = 6$ for EV. We select the set of nodes $\mathcal{S}$ from 500 randomly generated locations, set the signals on $\mathcal{S}^c$ to zero and reconstruct the original signals only from the signals on $\mathcal{S}$. The estimated signal is represented as $\boldsymbol{f} = \mathbf{U}_{\mathcal{V}\mathcal{R}}\mathbf{U}_{\mathcal{S}\mathcal{R}}^{-1}\boldsymbol{f}_{\mathcal{S}}$ where $\mathcal{R}$ is the set of Laplacian eigenvalues less than the estimated cut-off frequency $\Omega_k(\mathcal{S})$ with $k = 6$ (for existing approaches and EV) or λ_{F-1} (for SVD and SB).

Table 1. Performance Comparison (Average of 500 Tested Signals): SNR [dB]

$\lvert\mathcal{S}\rvert$	10	20	40	60	80	100
Entropy	-2.77	-2.17	0.28	2.02	2.13	5.75
MI	3.44	5.88	7.57	8.75	9.38	10.25
EV	3.52	**6.46**	7.74	8.53	**9.44**	**10.37**
SVD	**3.83**	5.98	7.55	**8.88**	9.14	10.29
SB	3.61	6.34	**7.86**	8.21	8.67	10.31

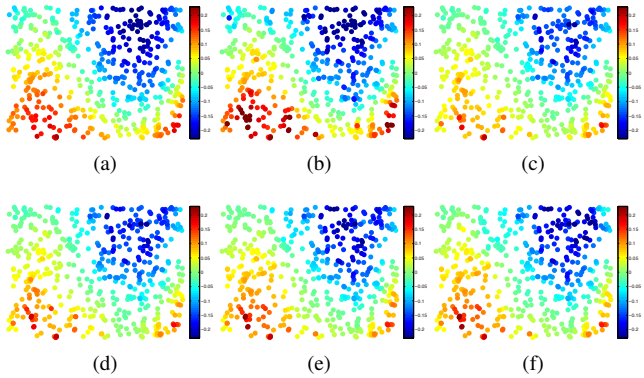

Fig. 4. Reconstructed samples from 50 observations using the exact method. (a) Original signal. (b) Entropy (6.64 dB). (c) MI (9.77 dB). (d) EV (10.82 dB). (e) SVD (11.59 dB). (f) SB (11.52 dB).

The average SNRs after 500 independent runs between the predicted signal and the original signal are shown in Table 1. The original signal and signals reconstructed from 50 samples are shown in Fig. 4. From the experiments, it can be seen that the proposed methods show better performance than the entropy and MI criteria regardless of the number of the observed locations.

5. CONCLUSION

The optimal sensor selection method based on the graph sampling theorem has been proposed for predicting the signal values on unobserved locations. We show that conventional methods using the GP models can be viewed as operations in graph vertex domain. All the proposed methods achieved better performance than the existing approaches. Moreover, the EV and SB are much faster than the entropy and MI criteria. As a future work, we will further investigate the theoretical issues and better reconstruction algorithms.

[2]This is one of kernels for the GP model used in [9].

[3]We use a Gaussian kernel $\mathcal{K}(i,j) = \exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/\widetilde{\theta}^2\right)$ in this experiment where $\widetilde{\theta}$ is determined from a training signal $f(m) = \sin(x_m(1))\sin(x_m(2))$ in which $\boldsymbol{x}$ is the set of randomly generated two-dimensional coordinates in $[0,1] \times [0,1]$.

6. REFERENCES

[1] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication cost," in *Proc. ACM/IEEE IPSN*, 2006, pp. 2–10.

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

[3] H. H. Gonzalez-Banos and J. Latombe, "A randomized art-gallery algorithm for sensor placement," in *Proc. SoCG'01*, 2001, pp. 232–240.

[4] D. S. Hochbaum and W. Maas, "Approximation schemes for covering and packing problems in image processing and VLSI," *Journal of the ACM*, vol. 31, no. 1, pp. 130–136, 1985.

[5] A. Singh, A. Krause, C. Guestrin, W. J. Kaiser, and M. A. Batalin, "Efficient planning of informative paths for multiple robots," in *Proc. IJCAI*, vol. 7, 2007, pp. 2204–2211.

[6] A. Nordio, A. Tarable, F. Dabbene, and R. Tempo, "Sensor selection and precoding strategies for wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4411–4421, 2015.

[7] N. A. C. Cressie, *Statistics for Spatial Data*. Wiley, 1991.

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

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

[10] W. F. Caselton and J. V. Zidek, "Optimal monitoring network designs," *Statistics and Probability Letters*, vol. 2, no. 4, pp. 223–227, 1984.

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

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

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

[14] M. Gavish, B. Nadler, and R. R. Coifman, "Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning," in *Proc. ICML'10*, 2010, pp. 367–374.

[15] D. Thanou, D. I. Shuman, and P. Frossard, "Learning parametric dictionaries for signals on graphs," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3849–3862, Aug. 2014.

[16] A. Sakiyama and Y. Tanaka, "Edge-aware image graph expansion methods for oversampled graph Laplacian matrix," in *Proc. ICIP'14*, 2014, pp. 2958–2962.

[17] F. Zhang and E. R. Hancock, "Graph spectral image smoothing using the heat kernel," *Pattern Recognition*, vol. 41, no. 11, pp. 3328–3342, 2008.

[18] A. Kheradmand and P. Milanfar, "A general framework for kernel similarity-based image denoising," in *Proc. GlobalSIP'13*, 2013, pp. 415–418.

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

[20] H. Higashi, T. Tanaka, and Y. Tanaka, "Smoothing of spatial filter by graph Fourier transform for EEG signals," in *Proc. APSIPA ASC*, 2014, pp. 1–8.

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

[22] A. Anis, A. Gadde, and A. Ortega, "Towards a sampling theorem for signals on arbitrary graphs," in *Proc. ICASSP'14*, 2014, pp. 3892–3896.

[23] 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.

[24] A. Gadde, A. Anis, and A. Ortega, "Active semi-supervised learning using sampling theory for graph signals," in *Proc. KDD'14*, 2014, pp. 492–501.

[25] A. Gadde and A. Ortega, "A probabilistic interpretation of sampling theory of graph signals," in *Proc. ICASSP'15*, 2015, pp. 3257–3261.

[26] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, to appear, 2015.

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

[28] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, 2016.

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

[30] 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, 2011. [Online]. Available: http://wiki.epfl.ch/sgwt

[31] A. Deshpande, C. Guestrin, S. Madden, J. Hellerstein, and W. Hong, "Model-driven data acquisition in sensor networks," in *Proc. VLDB*, 2004, pp. 588–599.

[32] A. Krause, "SFO: A toolbox for submodular function optimization," *Journal of Machine Learning Research*, 2010. [Online]. Available: http://las.ethz.ch/sfo/

[33] A. Krause and C. Guestrin, "Near-optimal observation selection using submodular functions," in *Proc. AAAI'07*, vol. 7, 2007, pp. 1650–1654.

[34] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006. [Online]. Available: http://www.gaussianprocess.org/gpml/code/matlab/doc/