Low-complexity Graph Sampling With Noise and Signal Reconstruction via Neumann Series

Fen Wang, Gene Cheung, Senior Member, IEEE, and Yongchao Wang, Member, IEEE

Abstract—Graph sampling addresses the problem of selecting a node subset in a graph to collect samples, so that a $K$-bandlimited signal can be reconstructed with high fidelity. Assuming an independent and identically distributed (i.i.d.) noise model, minimizing the expected mean square error (MMSE) leads to the known A-optimality criterion for graph sampling, which is expensive to compute and difficult to optimize. In this paper, we propose an augmented objective based on Neumann series that well approximates the A-optimality criterion and is amenable to greedy optimization. Specifically, we show that a shifted A-optimality criterion can be equivalently written as a function of an ideal low-pass (LP) graph filter, which in turn can be approximated efficiently via fast graph Fourier transform (FGFT). Minimizing the new objective, we select nodes greedily without large matrix inversions using a matrix inverse lemma. Further, for the dynamic subset sampling case where node availability varies across time, we propose an extended sampling strategy that replaces offline samples one-by-one in the selected set. For signal reconstruction, we propose an accompanied biased signal recovery strategy that reuses the approximated filter from sampling. Experiments show that our reconstruction is more robust to large noise than the least squares (LS) solution, and our sampling strategy far outperforms several existing schemes.

Index Terms—Graph signal processing, sampling, Neumann series theorem, matrix inversion, signal reconstruction.

I. INTRODUCTION

GRAPH SIGNAL PROCESSING (GSP) is the study of signals that reside on irregular data kernels described by graphs [1]–[3]. To analyze graph signals spectrally, a large amount of research strive to define frequencies on graphs, and subsequently design transforms and wavelets for spectral decomposition of signals [4]–[6]. Specifically, graph Fourier transform (GFT) of a graph signal $x \in \mathbb{R}^N$ is defined by its projection onto the eigenvector space of the graph Laplacian matrix [2] (or the adjacency matrix [7]).

A graph signal of bandwidth $K$ is a signal with non-zero GFT coefficients only for the $K$ eigenvectors associated with the $K$ smallest eigenvalues. Sampling of bandlimited graph signals is an important basic problem in GSP, since sensing in practice is often expensive, e.g., SAR image sampling for remote sensing [10] and label assignment in medical imaging [11]. Roughly speaking, sampling of graph signals in the literature can be defined from three different perspectives: selection sampling [12], aggregation sampling [13] and local measurement [14],[15]. In this paper, we focus on addressing the problem of selection sampling on a graph, such that the assumed bandlimited graph signal can be reconstructed with high fidelity.

Previous works in graph node selection can be broadly divided into two categories: noiseless sampling and noisy sampling. Given that samples are noiselessly observed, [16] showed that random sampling can result in perfect signal reconstruction with high probability, assuming that the sample size is larger than the signal’s bandwidth. However, the performance of random sampling is poor in the presence of noise. Assuming an independent and identically distributed (i.i.d.) additive noise model, minimizing the expected mean square error (MMSE) function leads to an A-optimality design criterion [17]. Minimizing the A-optimality value is known to be difficult: evaluating the criterion for a fixed sample set requires computation of eigenvectors and matrix inverse, and optimizing the sample set for a fixed budget is combinatorial.

In this paper, we propose a novel low-complexity deterministic graph sampling strategy via Neumann series expansion to minimize a variant of the A-optimality objective. One important contribution of our method is that it resolves the MMSE problem without full eigen-decomposition and large matrix inverses. Specifically, we first propose an augmented objective to approximate the A-optimality criterion, and prove the inverse of its information matrix is equal to a matrix series based on the Neumann series theorem [18]. We then rewire the objective as a function of an ideal low-pass (LP) graph filter with cutoff frequency $K$, which is efficiently approximated using fast Graph Fourier Transform (FGFT) [19]. To optimize the objective function, we select nodes greedily without any matrix inversions based on a matrix inverse lemma. Further, we prove that the greedy solution has near-optimal performance via supermodularity analysis [20].

The above proposed sampling strategy is designed for static global sampling, while in practice, the availability of nodes

1[8] proposed an alternative definition of generalized GFT that takes the irregularity of the graph into account. For direct graphs, [9] proposed the notion of graph direct variation. In this paper, we focus on the GFT derived from the Laplacian operator, but our sampling method is applicable to other symmetric graph operators.
often varies over time; e.g., one person goes offline in a social network, or one sensor runs out of battery in a sensor network [21], [22]. We additionally address the sampling problem for time-varying node subsets, called dynamic subset sampling, where the availability of each node varies as a function of time. Under this scenario, we develop an extended sampling strategy that replaces offline samples one-by-one in the selected set via Sherman-Morrison formula [23]. While there exist sampling works in dynamic scenarios, to the best of our knowledge, we are the first to formulate and tackle the sampling problem on dynamically available subsets in the literature.

Finally, we develop an accompanied signal reconstruction strategy to reuse the LP graph filter from sampling. This reconstruction strategy provides a biased estimator with low complexity, which is more robust to large noise than the conventional least squares (LS) solution [12]. Experiments demonstrate that our proposed sampling method outperforms several state-of-the-art sampling strategies in artificial and real-world graphs.

The outline of this paper is as follows. We review related works of graph sampling in Section II. In Section III, we overview GSP concepts and terminology used throughout this paper and discuss the A-optimality sampling problem. We propose the augmented A-optimality sampling criterion in Section IV and then a fast optimization algorithm is detailed in Section V. As an extension, we discuss the dynamic subset sampling problem in Section VI. Subsequently, an efficient and biased signal reconstruction method is introduced in Section VII. Finally, we present experimental results and conclusion in Sections VIII and IX, respectively.

Notations: In this paper, we use bold lowercase (x) and uppercase (A) letters to denote vectors and matrices, whose entries are x(i) and A_{ij} respectively. Notation (·)^T stands for transpose of a matrix or a vector; ∥x∥_2 is the l_2 norm of vector x. Tr(A) and ∥A∥_F compute the trace and Frobenius norm of A respectively. λ_{min}(A) and λ_{max}(A) are the smallest and largest eigenvalues of matrix A respectively. Calligraphic letters are used to denote sets such as S, whose cardinality is |S| and complementary set to V is S^c. The i-th element in S is S(i). x_S denotes a partial vector of x with entries indexed by set S. A_{S1,S2} is the sub-matrix of a matrix A with rows and columns indexed by S_1 and S_2, respectively. A_{S^c} is simplified to A_S. I is the identity matrix whose dimension depends on the context.

II. RELATED WORKS

We first review recent works on graph selection sampling in the GSP field. We next describe the other two alternative definitions of graph sampling. Then, we overview graph sampling in dynamic scenarios. Finally, we discuss some related works in sensor selection, which can be formulated similarly to graph selection sampling.

A. Selection Sampling of Graph Signals

In the absence of noise, authors in [24] proved a necessary and sufficient condition of a uniqueness set [25]—similar to Nyquist sampling theorem in time-series signal processing—and then proposed a sampling strategy with simple matrix-vector computations based on the notion of spectral proxies (SP). When the signal is corrupted by noise, [26] used a greedy procedure to optimize the MMSE problem directly, called minimum Frobenius norm (MFN). Alternatively, [27] used the E-optimality criterion for graph sampling, which can be interpreted as the worst case of MSE. Nonetheless, proposals in [26], [27] required eigen-decomposition of the Laplacian operator and costly computations in each greedy step, such as matrix inverse for MFN and singular value decomposition (SVD) for E-optimal sampling.

To reduce the complexity of evaluating the objective, an eigen-decomposition-free sampling strategy was recently proposed in [28], [29] by constraining the coverage of a localization operator, implemented using Chebyshev polynomial approximation.\(^2\)

In a separate development, authors in [30] proposed a sampling method based on Gershgorin disc alignment without explicit eigen-decomposition. However, those two works are not addressing the MMSE criterion directly. Orthogonally, our previous work, called matrix inversion approximation (MIA) [31], developed a sampling strategy to optimize an approximate MSE function based on truncated Neumann series, but the number of terms in the truncated sum must be sufficiently large for the approximation error to be small, limiting its practicality.

Because noisy graph sampling problem is NP-hard in general [20], all aforementioned criteria are optimized in a greedy manner. To lower complexity further, random node selection was proposed to sample signals based on a specified probability [32], [33]. However, for the same sampling budget, it cannot guarantee stable performance comparable to the deterministic counterparts [1], [29]. Unlike previous approaches [12], [26], [27], we propose a deterministic sampling method to minimize a DC-shifted A-optimality criterion directly but without eigen-decomposition or matrix inversion. Yet different from our work in [31], there is no explicit computation of the Neumann series sum in our augmented objective, and thus difficult truncation / approximation tradeoff is not necessary.

B. Other Definitions of Graph Sampling

Apart from graph subset sampling, local weighted sampling was proposed in paper [14] to aggregate samples via weighted summation of signals lived in different centerless local sets, which are partitioned from the original graph. Since graph local set partitions are required for each sampling budget, the complexity of local weighted sampling is hard to quantify, and its applicability for a general complex graph is limited. Another local sampling method was presented in [15], called random local aggregation, which first selects a random subset and then aggregates signals from those nodes and their local neighbors. From a different viewpoint, authors in [13] proposed to sample a signal at one node but with the graph signal shifted successively by some shift operator. The Vandermonde structure of the coefficient matrix is exploited to guarantee perfect recovery with few conditions and is further employed for graph support identification. However, the aggregation sampling still requires full eigen-decomposition of the Laplacian matrix. In this paper, we mainly focus on proposing an efficient selection sampling strategy without full eigen-decomposition.

\(^2\)Our proposed method is a function of an ideal graph low-pass filter, which can be viewed as an ideal operator-based sampling defined in [29]. However, we do not implement this filter by a localized operator.
C. Graph Sampling in Dynamic Scenarios

In this paper, we further extend our sampling strategy to the dynamic subset sampling problem. There are emerging researches in GSP to cope with graph sampling in different dynamic scenarios. Assuming that the observed signal is slowly time-varying with dynamic additive noise, authors in [34] developed (possibly time-varying) sampling strategies for adaptive bandlimited graph signal reconstruction. Additionally, some sampling strategies were proposed in [35] for tracking the dynamic graph signal generated from a linear time-varying transform model. Based on graph-time bandlimited assumption, a structured sampling strategy was designed to select nodes on a product graph [36], [37]. If the underlying graph is dynamic with a random edge sampling model, authors in [38] proposed to sample nodes on its statistical expected graph. However, we note that dynamic subset sampling has not been studied in these works.

D. Sensor Placement Selection

Sensor placement selection is a fundamental problem in wireless communication to choose a subset of sensors for monitoring a physical phenomena, such that the unsensed spatial signal can be well recovered [39]–[41]. In sensor selection, the field signal is assumed to be represented by a low dimensional parameter vector with a measurement matrix generated by a Gaussian process [42]. However, from a spectral GSP standpoint, the physical signal is assumed to be bandlimited on the underlying graph. Mathematically, the error covariance matrices in graph sampling and sensor selection are similar but with different prior assumptions [43]. There are several proposed metrics to quantify error covariance matrix in both sensor selection and graph sampling fields, such as $A$-optimality [31], [44], $E$-optimality [27], [42], $D$-optimality [41], [45] and frame potential [36], [39].

While graph subset sampling and sensor selection share a similar low-dimensional assumption, a generic graph sampling method can be applied beyond wireless sensor selection [46], including semi-supervised learning [47], matrix completion [36] and covariance sensing [45]. In this paper, we propose a generic graph sampling method to tackle the $A$-optimality objective in GSP field in a computation-efficient manner.

III. PRELIMINARIES

Denote by $G = (V, E, W)$ a graph with $N$ nodes indexed by $V = \{1, \ldots, N\}$. $E$ specifies the set of connected node pairs $(i, j)$, and the entry $W_{ij}$ of an $N \times N$ adjacency matrix $W$ is the weight of an edge connecting nodes $i$ and $j$ ($W_{ij} = 0$ if nodes $i$ and $j$ are not connected). We additionally define a diagonal degree matrix $D$, where $D_{ii} = \sum_j W_{ij}$. In this paper we focus on connected, undirected graphs with no self-loops, and we adopt the symmetric combinatorial graph Laplacian matrix $L = D - W$ as the variation operator. Suppose that the eigen-decomposition of $L$ is $L = \sum_{i=1}^{\lambda} \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$, where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ contains a set of $N$ orthonormal eigenvectors as columns corresponding to non-decreasing eigenvalues $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$. Then the GFT of a graph signal $x \in \mathbb{R}^N$ is defined as its expansion on the eigenspace of $L$, i.e., $\tilde{x} = \mathbf{V}^\top x$, and the inverse GFT is $x = \mathbf{V} \tilde{x}$. A graph signal is called $K$-bandlimited if its GFT coefficients $\tilde{x}$ are non-zero only for the lowest $K$ frequencies.

A sampled $K$-bandlimited graph signal can now be written as $x_{S} = \mathbf{C} x_{K} \tilde{x}_{K}$. Sampling operators satisfying rank($\mathbf{C} \mathbf{V}_K$) $= K$ are called qualified sampling operators in [27], since any $K$-bandlimited graph signal $x$ can be perfectly recovered from its noiseless samples $x_{S}$. Specifically, if rank($\mathbf{C} \mathbf{V}_K$) $= K$, then, from a sampled graph signal, the LS solution can provide a perfect and unique reconstruction, i.e., $\tilde{x}_{LS} = \mathbf{C} \mathbf{V}_K \mathbf{V}_K^\top \mathbf{x}_{K}$, where $\mathbf{I}$ denotes the pseudo-inverse computation [12]. Further, rank($\mathbf{C} \mathbf{V}_K$) $= K$ was demonstrated empirically to be satisfied with high probability via random node selection when $M \geq K$ [16]. In the sequel, we focus on the sampling region $M \geq K$ and assume rank($\mathbf{C} \mathbf{V}_K$) $= K$ is satisfied.

When the samples $x_{S}$ are corrupted by additive noise $n_{S}$, the LS reconstruction will produce a minimum variance unbiased estimator $\tilde{x}_{LS} = \mathbf{C} \mathbf{V}_K (\mathbf{C} \mathbf{V}_K)^\top \mathbf{y}_{S} = \mathbf{C} \mathbf{V}_K (\mathbf{C} \mathbf{V}_K)^\top (x_{S} + n_{S}) = x + \mathbf{C} \mathbf{V}_K (\mathbf{C} \mathbf{V}_K)^\top n_{S}$ [49]. Assuming that noise $n_{S}$ is i.i.d. with zero mean and unit variance, the covariance matrix of the reconstruction error is $\mathbf{R}_{x_{LS}} = \mathbb{E}[(\tilde{x}_{LS} - \mathbb{E}[\tilde{x}_{LS}]) (\tilde{x}_{LS} - \mathbb{E}[\tilde{x}_{LS}])^\top] = \mathbf{C} \mathbf{V}_K (\mathbf{C} \mathbf{V}_K)^\top \mathbf{V}_K^\top \mathbf{V}_K$. By the theory of optimal experiments design [50], finding a sampling matrix $\mathbf{C}$ to minimize the trace of the covariance matrix leads to the known $A$-optimality criterion:

$$C^* = \arg \min_{C \in \mathbb{R}^{M \times N}} \text{Tr} \left( (\mathbf{C} \mathbf{V}_K)^\top \mathbf{C} \mathbf{V}_K^{-1} \right),$$

where $\mathbb{R}^{M \times N}$ is the set of all qualified $\mathbf{C}$ defined in (1).

The objective function in (2) is similar to that formulated in sensor selection field regarding $\mathbf{V}_K$ as the sensor measurement matrix [41]. However, in graph sampling, $\mathbf{V}_K$ is not given as a prior but explicitly computed from $\mathbf{L}$, whose complexity is $\mathcal{O}(N^3)$ in general. Minimizing the $A$-optimality objective (2) directly using $\mathbf{C}$ is difficult because finding an optimal $\mathbf{C}^*$ is NP-hard and evaluating the function value for a given $\mathbf{C}$ requires expensive matrix inverse computation.

IV. AUGMENTED A-OPTIMALITY GRAPH SIGNAL SAMPLING

In this section, we propose a new sampling objective to approximate the $A$-optimality criterion. We first review the Neumann series theorem in [18] and one result in [31], both of which will be used for proving a forthcoming theorem.

Theorem 1. (Neumann series theorem): If the eigenvalues $\lambda_i$ of a square matrix $\mathbf{A}$ have the property that $\rho(\mathbf{A}) = \max_i |\lambda_i| \leq 1$, then its Neumann series $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots$ converges to $(\mathbf{I} - \mathbf{A})^{-1}$, i.e., $(\mathbf{I} - \mathbf{A})^{-1} = \sum_{i=0}^{\infty} \mathbf{A}^i$.

In [20] defined the $K$-spectrally sparse graph signal by $x = \mathbf{V}_K \tilde{x}_K$, which is a generalization of our formulation since $K \in \mathcal{V}$ is not restricted to be the first $K$ elements. We use the more conventional definition here, but our proposed algorithm is applicable to the general case.
Proposition 1: Denote by $\Psi = (CV_K)^T CV_K$, $\Phi = I - \Psi$ and $\delta_1 \leq \cdots \leq \delta_K$ as the eigenvalues of $\Phi$. Then,

$$0 \leq \delta_i < 1,$$

(3)

if $CV_K$ is full column rank, i.e., rank($CV_K$) = $K$.

Proof: The proof is in Appendix A.

These two results help us prove Theorem 2 in the next subsection.

A. Augmented A-Optimality Criterion

We propose an augmented objective by adding a weighted identity matrix to (2):

$$C^* = \arg \min_{C \in \mathbb{R}^{M \times N}} \text{Tr} \left( [(CV_K)^T CV_K + \mu I]^{-1} \right),$$

(4)

where $\mu$ is a small weight (shift) parameter with $0 < \mu < 1$. Its design will be discussed in details later.

Like problem (2), problem (4) is also difficult to optimize directly. Before proposing its alternative simpler form, we first present the following proposition.

Proposition 2: Denote by $\epsilon_1 \leq \cdots \leq \epsilon_M$ the eigenvalues of $T_S$, then

$$0 \leq \epsilon_i < 1,$$

(5)

where $T = V_K V_K^T$ is an ideal LP graph filter with cutoff frequency $K$.

Proof: $\forall x \in \mathbb{R}^M$ and $\|x\|_2 = 1$,

$$x^T T_S x = x^T (CV_K V_K^T C)^T x = (C^T x)^T (V_K V_K^T x) = y^T (V_K V_K^T) y,$$

(6)

where $y = C^T x$. Hence $y_S = x$, $y_{S^c} = 0$ and $\|y\|_2 = 1$.

Since $T$ is symmetric and its eigen-decomposition can be written as $T = V_K V_K^T = \text{diag}(\{1, \cdots, 1, 0, \cdots, 0\}) V^T$, its eigenvalues are 0 or 1. Then, due to the Rayleigh quotient theorem [18], $\forall y \in \mathbb{R}^N$ and $\|y\|_2 = 1$, $0 \leq y^T (V_K V_K^T) y \leq 1$, which holds also for some specific $y$. Thus,

$$0 \leq x^T T_S x \leq 1,$$

(7)

which implies Proposition 2 via the Rayleigh quotient theorem.

Now, we proceed and prove the next theorem by employing the results from Theorem 1, Proposition 1 and Proposition 2.

Theorem 2: The augmented A-optimal objective (4) is equivalent to

$$S^* = \arg \min_{S: |S| = M} \text{Tr} (T_S + \mu I)^{-1},$$

(8)

if the selected matrix $CV_K$ is full column rank.

Proof: Let $\tilde{\Psi} = (CV_K)^T CV_K + \mu I = \Psi + \mu I$, and $\tilde{\Phi} = I - \tilde{\Psi} = \Phi - \mu I$. We know from Proposition 1 that $0 \leq \delta_i < 1$. Hence, the eigenvalues of $\tilde{\Phi}$ are in $[-\mu, 1 - \mu]$. Since the shift parameter $0 < \mu < 1$, $\rho(\tilde{\Phi}) < 1$. By applying Theorem 1,

$$[(CV_K)^T CV_K + \mu I]^{-1} = \sum_{l=0}^{\infty} (I - \tilde{\Psi})^l.$$

(9)

Leveraging on a property of the trace operator, we write

$$\text{Tr}(I - \tilde{\Psi})^l = \text{Tr}[(1 - \mu)I - \tilde{\Psi}]^l$$

$$= \text{Tr} \left( (1 - \mu)I + \sum_{d=1}^{l} \left( \sum_{j} (\eta_j)(1 - \mu)^{(l-d)} \right) (-\tilde{\Psi})^d \right)$$

$$= (1 - \mu)^l K + \sum_{d=0}^{l} \left( \sum_{j} (\eta_j)(1 - \mu)^{(l-d)} \right) \text{Tr} \left( [-T_S]^d \right)$$

$$= (1 - \mu)^l (K - M) + \sum_{d=0}^{l} \left( \sum_{j} (\eta_j)(1 - \mu)^{(l-d)} \right) \text{Tr} \left( [-T_S]^d \right)$$

(10)

where (a) holds since $\text{Tr}(\Psi^d) = \text{Tr}(V_K^T C^T CV_K \cdots V_K^T C^T CV_K) = \text{Tr}(CV_K V_K^T C^T \cdots CV_K V_K^T C^T) = \text{Tr}(T_S^d)$.

Hence,

$$\text{Tr} \left[ \sum_{l=0}^{\infty} (I - \tilde{\Psi})^l \right] = \sum_{l=0}^{\infty} \text{Tr}(I - \tilde{\Psi})^l$$

$$= \sum_{l=0}^{\infty} (1 - \mu)^l (K - M) + \text{Tr} [(1 - \mu)I - T_S]^l$$

$$= (K - M) \left( \frac{1 - (1 - \mu)^\infty}{\mu} \right) + \text{Tr} \sum_{l=0}^{\infty} [(1 - \mu)I - T_S]^l$$

$$= K - M + \text{Tr} \sum_{l=0}^{\infty} [I - (T_S + \mu I)]^l,$$

(11)

where equality (a) holds since $0 < \mu < 1$.

From Proposition 2, the eigenvalues of $I - (T_S + \mu I)$ are in $[-\mu, 1 - \mu]$. Again, since the shift parameter $\mu$ is small and positive, $\rho[I - (T_S + \mu I)] = 1 - \mu < 1$. Furthermore, matrix $T_S + \mu I$ is always invertible when $0 < \mu < 1$. Using Theorem 1, we have

$$\sum_{l=0}^{\infty} [I - (T_S + \mu I)]^l = (T_S + \mu I)^{-1}.$$

(12)

Combining (9), (11) and (12),

$$\text{Tr} \left( [(CV_K)^T CV_K + \mu I]^{-1} \right) = \frac{K - M}{\mu} + \text{Tr} (T_S + \mu I)^{-1},$$

which implies Theorem 2 because $M$, $K$, and $\mu$ are all constant in a sampling problem.

Optimizing the new problem (8) requires computation of an ideal LP filter rather than first $K$ eigenvectors of $L$. We discuss an efficient approximation of the LP filter next.

B. Ideal Low-Pass Graph Filter Approximation

One classical method is to approximate an ideal LP filter $T$ in (8) by using a Chebyshev matrix polynomial of $L$, i.e., $T^{\text{poly}} = \sum_{l=1}^{N} (\sum_{j=0}^{q} \beta_{l,j} L^j \sum_{j=0}^{q} \beta_{l,j} L^j)^{l} [5]$. Another method is to apply the Jacobi eigenvalue algorithm: approximately diagonalize $L$ with an estimated eigenvector matrix $V = S_1 \cdots S_f$ [19]. Specifically, the goal is to solve the following optimization
where $\tilde{A}$ is constrained to be a diagonal matrix, and $S_i$ are constrained to be Givens rotation matrices.

A truncated Jacobi algorithm [51] was adopted to optimize (13) iteratively. With the optimized $\tilde{V}$ where $L = \tilde{V}A\tilde{V}^T$, an ideal LP filter $T$ can be implemented as $T^\text{FGFT} = V_K\tilde{V}_K^T$. We adopt $T^\text{FGFT}$ for the following reasons:

1) [19] claimed the superiority of $T^\text{FGFT}$ compared with $T^\text{Poly}$ when approximating ideal graph LP filters; specifically, $\frac{\|T^\text{FGFT}\|_p}{\|T^\text{Poly}\|_p} < \frac{\|T\|_p}{\|T\|_p}$ under the same approximation complexity [19]. In our work, $T$ is an ideal graph LP filter. Moreover, if the bandwidth of graph signals $K$ varies in one graph, we can reuse the same $\tilde{V}$ when realizing $T^\text{FGFT}$. 

2) A sufficient condition of Proposition 2 is that the eigenvalues of $T$ are in $\{0, 1\}$, based on which we can claim the eigenvalues of $I - (T_S + \mu I)$ are in $[-\mu, 1 - \mu]$ in Theorem 2. Then, (12) can hold with a small positive parameter $\mu$. Small $\mu$ is also required for (4) to well approximate the original A-optimality criterion (2). Because $\tilde{V}$ computed from Givens matrices is orthogonal, Proposition 2 also holds for $T^\text{FGFT}$. In contrast, the eigenvalue scope of $T^\text{Poly}_S$ is unpredictable before sampling, such that the value of $\mu$ is difficult to design for fulfilling (12).

Based on above discussions, we now write our final augmented objective as

$$S^* = \arg \min_{S:|S|=M} \text{Tr} \left( T^\text{FGFT}_S + \mu I \right)^{-1},$$

which requires just an approximated LP filter operator rather than eigen-decomposition of the graph Laplacian operator. For simplicity, we write $G = T^\text{FGFT} + \mu I$ in the sequel.

C. Selection of Shift Parameter $\mu$

We now discuss how to select an appropriate shift $\mu$. To well approximate the original criterion (2), the shift $\mu$ ought to be as small as possible, but a small $\mu$ would cause the matrix inverse computation in (14) to be inaccurate on a precision constrained platform. To ensure numerical stability, we can bound the condition number $\kappa(A)$ of $A$, where $\kappa(A) = \lambda_{\text{max}}(A)/\lambda_{\text{min}}(A)$ [52]. However, the condition number of $G_S$ cannot be computed before sampling since its explicit eigenvalues are dependent on the sample set $S$. Instead, we propose to bound its worst case for stability, which has no relation to the sample set. As discussed, Proposition 2 also holds for $T^\text{FGFT}_S^{-1}$, so $\mu \leq \lambda(G_S) \leq 1 + \mu$. We therefore bound the worst case of $\kappa(G_S)$ as follows:

$$\kappa(G_S) = \frac{\lambda_{\text{max}}(G_S)}{\lambda_{\text{min}}(G_S)} \leq \frac{1 + \mu}{\mu} \leq \kappa_0,$$

where $\kappa_0$ is the upper bound of an acceptable condition number. From the latter part, $\mu \geq \frac{1}{\kappa_0-1}$. Because $\mu$ should be as small as possible, the optimal $\mu$ is $\mu^* = \frac{1}{\kappa_0-1}$. We set $\kappa_0 = 100$ in our experiments and compare its performance with other design strategies via simulations.

V. Fast Greedy Sampling

It is difficult to compute an optimal solution of our proposed criterion (14) with much low complexity since the problem is combinatorial. Greedy algorithms have been commonly used in the graph sampling literature [12, 27] and proved to have bounded performance for some objective functions, such as logdet [41] and frame potential [36]. However, if (14) is minimized naively using a greedy scheme, the algorithm needs to perform one matrix inversion to evaluate each candidate sample set $S$. To alleviate this computation burden, we propose an accelerated greedy algorithm to avoid the matrix inverse operation. Furthermore, we analyze and demonstrate the suboptimality of the greedy solution in this section.

A. Accelerated Greedy Sampling

We first review the block matrix inversion formula, which we exploit when designing the subsequent accelerated greedy algorithm [53].

**Lemma 1:** The inverse of matrix $M$ can be computed using the inverse of sub-matrix $A$ and the inverse of its Schur complement, i.e.,

$$M^{-1} = \begin{bmatrix} A & U \\ V & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}UH^{-1}VA^{-1} & -A^{-1}UH^{-1} \\ -H^{-1}VA^{-1} & H^{-1} \end{bmatrix},$$

(16)

where $H = C - VA^{-1}U$ is the Schur complement $M/A$ of sub-matrix $A$ of matrix $M$. We observe that matrix $G$ is symmetric, and under some permutation, its sub-matrix $G_{S,i}(i)$ can be expressed as

$$G_{S,i}(i) = \begin{bmatrix} G_{S} & G_{S,i} \\ G_{i,S} & G_{ii} \end{bmatrix} = \begin{bmatrix} G_{S} & g_i \\ g_i^T & G_{ii} \end{bmatrix},$$

(17)

where $g_i$ denotes the partial vector of $i$-th column of $G$ indexed by $S$.

From Lemma 1, we can compute the inverse of $G_{S,i}(i)$ as

$$G^{-1}_{S,i}(i) = \begin{bmatrix} G^{-1}_{S} + h^{-1}G^{-1}_{S}g_ig_i^T G^{-1}_{S} & -h^{-1}G^{-1}_{S}g_i \\ -h^{-1}g_i^TG^{-1}_{S} & h^{-1} \end{bmatrix},$$

(18)

where $h = G_{ii} - g_i^TG^{-1}_{S}g_i$ is actually a scalar.

The above equation reveals that $G^{-1}_{S,i}(i)$ can be computed using $G^{-1}_{S}$ stored in the last iteration. Instead of computing $G^{-1}_{S,i}(i)$ from scratch, (18) performs only two matrix-vector products and then stacks the results. We outline pseudo-code of the fast greedy algorithm in Algorithm 1 and call it graph filter submatrix (GFS)-based sampling algorithm. We note that GFS requires neither computation of eigenvectors of $L$ nor large matrix inversion, and it produces exactly the same sampling set as naively optimizing problem (14) using a greedy scheme.

We next study and demonstrate the performance bound of this greedy solution via supermodularity analysis.
The set function $\alpha = 1 + T_{\text{Tr}} < 0 \text{ dB})$. Consequently, $K$ is not tight to bound the real value of $r$. The following Lemma can be established based on the analysis procedure in [20]:

**Lemma 2:** The set function $g(S) = \text{Tr}[(C \tilde{V} K)^{\top} C \tilde{V} K + \mu I]^{-1}$ is (i) monotone decreasing and (ii) $\alpha$-supermodular with

$$\alpha \geq \frac{\mu^3 (\mu + 2)}{(\mu + 1)^4}. \quad (23)$$

**Proof:** The detailed proof is shown in Appendix B.

Remark: In our paper, parameter $\mu$ mimics the ‘SNR’ in [20] via ‘SNR’ $= -10 \log_{10}(\mu)$. We illustrate the lower bound of $\alpha$ in (23) in terms of the value of $-10 \log_{10}(\mu)$ in Fig. 1. This figure matches the result demonstrated in [20]. In our paper, the value of $\mu$ is set to be $0 < \mu < 1 (-10 \log_{10}(\mu) > 0 \text{ dB})$. Consequently, the lower bound of $\alpha$ is not tight to bound the real value of $\alpha$ (close to 1 via exhaustive search), as shown in [20]. Although of interest, deriving a tighter bound of $\alpha$ is outside the scope of this paper. To show the near-optimality of the proposed GFS algorithm, we simulate some toy experiments on small graphs ($N = 20$) to compute the exact relative suboptimality value $r$ of the solution from Algorithm 1. The optimal solution is computed by exhaustive search. For comparison, we also illustrate the value of $r$ of random solution. In this simulation, the bandwidth of graph signals is $K = 5$, $\mu = 1/99$, $J = 12 \sqrt{N} \log N$, and the artificial graphs are community graphs (G1) and sensor graphs (G2). Experimental results are depicted in Fig. 2, which illustrates that the performance of the proposed GFS method at each sample size is close to the exhaustive optimal solution, and the resulted value $r$ is much smaller than that of random solution.

**C. Complexity Analysis**

We here analyze the complexity of the proposed GFS sampling strategy, assuming $M = K$. The overall complexity is divided into two parts: “Preparation” and “Selection.” The

**Algorithm 1: GFS Graph Signal Sampling Algorithm.**

**Input:** Graph operator $L$, bandwidth $K$, budget $M$ and parameter $\mu$, $S = \emptyset$.

1. Compute $\tilde{V} = S_1 \ldots S_J$ of $L$ via (13).
2. Compute $T^{\text{FGFT}} = \tilde{V} K \tilde{V}_{K}$ and $G = T^{\text{FGFT}} + \mu I$.
3. Select the first node by $u = \arg \max_{i \in V} G_{ii}$, update $G_{S}^{-1} = G^{-1}_{S}$ and $S \leftarrow S \cup \{u\}$
4. While $|S| < M$
5. $\forall i \in S^c$, compute

$$g_i = G_{S,(i)}$$

and $h = G_{ii} - g_i G_{S}^{-1} g_i$

$$G_{S,(i)}^{-1} = \left[ \begin{array}{ccc} G_{S}^{-1} + h^{-1} g_i g_i^{\top} G_{S}^{-1} & -h^{-1} G_{S}^{-1} g_i \\ -h^{-1} g_i^{\top} G_{S}^{-1} & h^{-1} \end{array} \right]$$

6. Select $u = \arg \min_{i \in S} \text{Tr}(G_{S,(i)}^{-1})$
7. Update $G_{S}^{-1} = G_{S,(i)}^{-1}$ and $S \leftarrow S \cup \{u\}$
8. **end While**
9. Return $S$ and $G_{S}^{-1}$

**B. Relative Suboptimality and Approximate Supermodularity**

First, the relative suboptimality value is defined as a criterion to evaluate different solutions.

**Definition 2:** Let $g^* = g(S^*)$ be the optimal value of problem

$$\min_{S \subseteq V, |S| = M} g(S), \quad (19)$$

and $\hat{S}$ is the solution of one algorithm with $|\hat{S}| = M$, then the relative suboptimality is defined as

$$r = \frac{g(\hat{S}) - g^*}{g(\emptyset) - g^*}. \quad (20)$$

If the objective function $g(S)$ is monotonic decreasing and supermodular, the value of $r$ of greedy solution is upper-bounded by $e^{-1}$ [54], thus guaranteeing the suboptimality of the greedy solution. MSE function is not supermodular in general, but authors in [20] proposed a definition of approximate supermodularity to study the MSE function. This property captures the following merit:

**Theorem 3:** Let $g^* = g(S^*)$ be the optimal value of problem (19) and $S_l$ is the set obtained via greedy algorithm with $l$ samples. If function $g$ is (i) monotonic decreasing and (ii) supermodular, then

$$\frac{g(S_l) - g^*}{g(\emptyset) - g^*} \leq \left(1 - \frac{\alpha}{M}\right)^l \leq e^{-\alpha l/M}. \quad (21)$$

In our proposal, the finalized objective function is $f(S) = \text{Tr}(T^{\text{FGFT}}_S + \mu I)^{-1}$. Proposition 2 also holds for $T^{\text{FGFT}}_S$, therefore

$$g(S) = \text{Tr}[(C \tilde{V} K)^{\top} C \tilde{V} K + \mu I]^{-1} = \frac{K - M}{\mu} + f(S), \quad (22)$$

which indicates that the samples from greedyly optimizing $f(S)$ are exactly the same as that of greedily optimizing $g(S)$.

Since $g(S)$ approximates the real MSE value in equation (2), it is more reasonable to evaluate the suboptimality of solution from Algorithm 1 with respect to the optimal solution of minimizing $g(S)$. The following Lemma can be established based on the analysis procedure in [20]:

![Fig. 1. Lower bound of $\alpha$ in equation (23) in terms of $-10 \log_{10}(\mu)$.](image1)

![Fig. 2. Relative suboptimality value defined in (20) in terms of sample size. G1: community graphs; G2: sensor graphs.](image2)
preparation step includes the complexity of computing the prior information, e.g., eigen-decomposition, and the selection step has the complexity of collecting samples. GFS needs to compute matrix G in the preparation step. Authors in [19] have shown that the complexity of computing V via parallel truncated Jacobi algorithm is O(NJ log N). During experiments, the number of Givens rotation matrices is set at J = O(N log N). Once we obtain S1,..., SJ, T_{PEGFT} can be implemented via T_{PEGFT} = S1...SjBS_j...S1^T, where B is a diagonal matrix with ones on the first K diagonal elements and zeros for the rest. Since every Givens matrix Si is sparse with 4 non-zero entries, T_{PEGFT} can be computed iteratively with complexity O(NJ). For the selection step, the complexity for matrix-vector product is O(M^2). Considering |S| = M and |S|^2 = N - M, the complexity of GFS in the sampling step is at most O(NM^3) as shown in Table I. Table I shows the complexity of different sampling algorithms, some of which are borrowed from [12]. In the preparation step, the E-optimal and MFN sampling methods require V_K, whose complexity is O((|E| M + RM)^2 T_1) via the locally optimal block preconditioned conjugated gradient (LOBPCG) [55]. T_1 denotes the average number of iterations for convergence of this method, and R is a constant for mixing the complexity of two parts in each iteration [55]. MIA requires the K-th eigenvalue of L and an approximate LP filter T_{Pol}', whose complexity is O(qN|E'|), where q is the length of a Chebyshev matrix polynomial [31]. In the selection step, the first eigen-pair of (L^\top L)^\top is required for SP sampling method, and its complexity is O(|E|M^2 T_2(k)), where k is the order of the proxies approximation, and T_2(k) is the number of iterations for convergence to get one eigen-pair [12]. E-optimal requires SVD, and MFN requires matrix inversion computation for selection, so their complexity in this step are both O(NM^4). The matrix multiplication in MIA has an asymptotic complexity of O(M^2 373), thus its complexity in the selection step is O(NLM^3 373), where L is a truncation parameter [31]. From Table I, we can see that the complexity of the proposed GFS method is lower than the MIA, MFN and E-optimal method in the selection step. Numerical analysis for complexity will be illustrated in the experimental section.

VI. DYNAMIC SUBSET SAMPLING

The proposed GFS sampling strategy is designed for static global sampling, while in practice, the availability of sensing nodes varies over time for some real-world applications [21]. In this section, we extend our proposed algorithm to dynamic subset sampling, which means that the available set for sampling is a time-varying subset, while the underlying graph remains the same. For concreteness, we first develop a mathematical model to generate a node subset A_t \subset V at time t that slowly evolves over time.

At t = 0, an initial available subset A_0 is generated by randomly selecting P_0|N nodes from V, where P_0 is a fixed probability. Denote by z_j^t the state of node j at time t, where z_j^t = 1 if j \in A_t, and z_j^t = 0 otherwise. Define the state crossover probability as P(z_j^t+1 = 0 | z_j^t = 1) = P(z_j^t+1 = 1 | z_j^t = 0) = \epsilon. Given A_t, one can generate A_{t+1} from A_t using the state crossover probability. The crossover probability \epsilon tends to be small in practice.

Based on the introduced notations, the dynamic subset sampling problem can be formulated formally as follows

\[
\min_{S_{t+1}} \text{Tr} \left( T_{PEGFT}^T + \mu I \right)^{-1}
\]

s.t. \ S_t \subset A_t; \ S_t \not\subset A_{t+1};

S_{t+1} \subset A_{t+1}; |S_{t+1}| = |S_t| = M

(24)

where S_t is the sampling set at time t. Without losing generality, we assume that S_t \not\subset A_{t+1}, i.e., the sampling set is not available at the next coming time slot.

We next discuss our sampling strategy for this problem.

A. Sampling on Dynamic Subsets

For conciseness, denote by P = A_{t+1} \cap S_t the selected and available (SA) set and by U = S_t \setminus P the selected and unavailable (SU) node set. Then, our sampling strategy, based on the idea of node exchange (NE), can be described as follows.

1) We replace a node j \in U with an available and unselected (AU) node k \in Q where Q = A_{t+1} \setminus P. The optimal node k^* is greedily chosen if exchanging j with k^* minimizes the objective function in all the (j, k) exchange pairs. This procedure is repeated until all unavailable nodes in SU set are replaced by available nodes, which will form a new sampling set S. This step is sketched in the first FOR-loop in Algorithm 2.

2) To further improve performance, node j \in S will be replaced by node k \in H if the (j, k) exchange can induce a lower objective, where H = A_{t+1} \setminus S. This procedure will return the final sampling set S_{t+1} until it achieves the iteration number constraint, or it could not find one more node-pair exchange to make the objective decrease in this constraint. This step is shown in the second outer FOR-loop in Algorithm 2.

For each node-pair exchange evaluation, G_{S^1}^\top in (24) has to be computed. In the following, we propose one method to relieve this computation burden.

B. One-Pair Node Exchange

The complexity of computing G_{S^1}^\top for each node-pair exchange can be reduced based on Sherman-Morrison formula [23], reviewed below.

Lemma 3: Suppose A \in \mathbb{R}^{N \times N} is an invertible square matrix and u, v \in \mathbb{R}^N are column vectors. If A + uv^\top is invertible,
Algorithm 2: GFS-NE Dynamic Subset Sampling Algorithm.

Apply the GFS algorithm on \(A_0\) and then output \(S_0\) and \(G_{S_0}^{-1}\). Given \(K_0, S_t\) and \(G_{S_t}^{-1}\),

1: Initialization: \(S = S_t\) and \(G_S^{-1} = G_{S_t}^{-1}\)
2: \(r = 1, \ldots, |U|\)
3: \(j = \mathcal{U}(r)\) and \(i = d|S(d) = j|\)
4: For all \(k \in G, S = S \cup \{k\} \setminus \{j\}\), compute \(G_S^{-1}\) based on equations (29) and (30).
5: Select \(k^* = \arg \min_{k \in G} \text{Tr}(G_S^{-1})\), and \(S = S \cup \{k^*\} \setminus \{j\}\)
6: \(\xi = \xi + 1, S = S\), \(G_S^{-1} = G_S^{-1}\), and \(H = A_{t+1} \setminus S\)
7: \(\xi \geq K_0, S_{t+1} = S\) and \(G_{S_{t+1}}^{-1} = G_{S_{t+1}}^{-1}\); end if
8: \(i = 1; \) go to step 1 in this second part
9: end if
10: end for
11: end for
12: Return \(S_{t+1} = S\) and \(G_{S_{t+1}}^{-1} = G_S^{-1}\)

From Lemma 3, the inverse of \(G_S\) can be computed via:

\[
F^{-1} = G_S^{-1} - \frac{G_S^{-1}e_i p^T G_S^{-1}}{1 + p^T G_S^{-1} e_i},
\]

(29)

\[
G_S^{-1} = F^{-1} - \frac{F^{-1} e_i q^T F^{-1}}{1 + e_i q^T},
\]

(30)

composed of matrix-vector products only.

Using rank-1 updates, the complexity of computing matrix inverse after each node-pair exchange will be reduced from \(M^3\) to \(M^2\). Initially, we compute the set \(S_0\) using our proposed GFS algorithm described in Section V. During time evolution, the qualified subset will be collected by the proposed strategy which is called the GFS-NE sampling method. The details of pseudo code are outlined in Algorithm 2. For the first part of Algorithm 2, its complexity is \(O(|U|N M^2)\), where \(|U| < M\). For the second part, the complexity is at most \(O(K_0 N M^3)\), where \(K_0\) is the upper bound of exchange iterations to constrain complexity. In the experimental section, we will compare this proposed dynamic subset sampling algorithm to existing greedy sampling algorithms which select nodes from scratch at each time step.

VII. ACCOMPANIED GRAPH SIGNAL RECONSTRUCTION

Extending our derivation in sampling, we propose a biased signal reconstruction method with lower variance and complexity than the conventional LS estimator.

A. Graph Filter Sub-Matrix-Based Reconstruction

From Section III, the LS estimator is \(x_{LS} = V_K \left[\left(\mathbf{C}V_K\right)\mathbf{K}^{-1}\left(\mathbf{C}V_K\right)^\top\right] y_S\), where \(y_S\) is an observed noise-corrupted signal. The LS reconstruction requires computation of the first \(K\) eigenvectors of the graph operator \(L\) and a matrix inverse operation, which has the lowest MSE value among unbiased recovery solutions. We next derive a biased signal reconstruction method.

**Proposition 3:** Adding a shift \(0 < \beta < 1\) into the LS solution, we obtain a biased estimator \(\hat{x} = V_K \left[\left(\mathbf{C}V_K\right)\mathbf{K}^{-1}\left(\mathbf{C}V_K + \beta I\right)^{-1}\left(\mathbf{C}V_K\right)^\top\right] y_S\), which can be approximated by

\[
\hat{x} = \mathbf{T}^{\text{FGFT}} H_S^{-1} y_S,
\]

(31)

where \(\mathbf{T}^{\text{FGFT}}\) was obtained during GFS sampling using \(\mathbf{T}^{\text{FGFT}} = V_K V_K^\top\), and \(H = \mathbf{T}^{\text{FGFT}} + \beta I\).

**Proof:** Similar derivation in (9) and (10) can show that the shifted LS solution is equal to

\[
\hat{x} = V_K \left[\sum_{l=0}^{\infty} \sum_{d=0}^{l} \left(\beta\right)^{l-d} \left(-\Psi\right)^d\right] \left(\mathbf{C}V_K\right)^\top y_S
\]

\[
= T \left[\sum_{l=0}^{\infty} \sum_{d=0}^{l} \left(\beta\right)^{l-d} \left(-P\right)^d\right] C^\top y_S
\]

\[
= T \sum_{l=0}^{\infty} \left[I - (P + \beta I)^l\right] C^\top y_S,
\]

(32)

where \(P = C^\top \mathbf{C}V_K V_K^\top\) and (a) holds since \(V_K \Psi^d V_K = V_K (V_K C^\top \mathbf{C}V_K) \cdots (V_K C^\top \mathbf{C}V_K) V_K = V_K V_K^\top P^d\).
From the definition of C, we know that \( C^T C = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \) under appropriate permutation, which leads to \( P = \begin{bmatrix} T_0 & T \end{bmatrix} \). Hence, \( I - (P + \beta I) = \begin{bmatrix} 1 - (T_S + \beta) & T_{33}^\top \\ 0 & (1 - \beta)I \end{bmatrix} \), resulting in

\[
[I - (P + \beta I)]^l = \begin{bmatrix} I - (T_S + \beta I) & 0 \\ 0 & (1 - \beta)^lI \end{bmatrix}
\]

where \(^l\cdot\) denotes a nonzero matrix whose dimension is \( M \times (N - M) \).

Then, (32) can be written as

\[
\hat{x} = T \sum_{l=0}^{\infty} \begin{bmatrix} I - (T_S + \beta I)^l & 0 \\ 0 & (1 - \beta)^lI \end{bmatrix} y_S
\]

\[
= T_{VS} \sum_{l=0}^{\infty} [I - (T_S + \beta I)^l] y_S
\]

\[
= T_{VS} (T_S + \beta I)^{-1} y_S,
\]

where equality \((a)\) holds from a similar derivation of (12) with a different shift. Finally, we approximate \( T \) using \( \text{FGFT} \).

Remark: If we choose \( \beta = \mu \), then \( H = G \). All involved matrices in (31) have been obtained in sampling. However, the value for \( \mu \) has been designed in the sampling procedure based on the condition number constraint. If \( \beta \neq \mu \), we can customize a \( \beta \) based on the bias-variance tradeoff (to be discussed), at the cost of computing the inverse \( H_S \).

We next prove the robustness of the proposed estimator \( \hat{x} \) and propose one strategy to compute the value of \( \beta \).

### B. Robustness Analysis

Assume that the signal has the same energy for different SNRs, and noise \( n \) is i.i.d. with zero mean and variance \( \omega^2 \) which varies with SNR. We have the following proposition.

**Proposition 4:** Given \( \text{FGFT} \) sufficiently approximates the ideal LP filter, the expected MSE of the proposed GFS reconstruction and the LS reconstruction are respectively:

\[
E \| \hat{x} - x \|^2_2 = \sum_{i=1}^{K} \left( 1 + \frac{\sigma_i}{\beta} \right)^{-2} (u_i^T x_K)^2 + \omega^2 \sum_{i=1}^{K} \frac{\sigma_i}{(\sigma_i + \beta)^2},
\]

and

\[
E \| \hat{x}_{LS} - x \|^2_2 = \omega^2 \sum_{i=1}^{K} \frac{1}{\sigma_i},
\]

where \( (\sigma_i, u_i) \) is the \( i \)-th eigen-pair of the matrix \( \Psi = (C V_K)^T C V_K \).

**Proof:** The proof is detailed in Appendix C.

When the noise variance \( \omega^2 \) is large enough to dominate MSE, the GFS recovery has lower MSE than the LS solution since \( \frac{\sigma_i}{(\sigma_i + \beta)^2} < \frac{1}{\sigma_i}, \forall i \). Hence, the proposed method is more robust to large noise than the LS reconstruction.

The RHS of formula (35) is consisted of bias and variance of estimator \( \hat{x} \) respectively. For the bias part, \( u_i \) and \( x_K \) are constant in reconstruction, thus smaller \( \beta \) will bring smaller bias. For the variance part, smaller \( \beta \) will bring larger variance. Hence, given the noise variance \( \omega^2 \), the optimal \( \beta \) can be designed to balance the bias-variance tradeoff to achieve the lowest MSE.

### VIII. Experimental Results

Our proposed sampling and reconstruction strategies are evaluated via experimental simulations. All experiments were performed in MATLAB R2017b, running on a PC with Intel Core i3 3.7 GHz CPU and 16 GB RAM. The simulated artificial graphs are described as follows:

- **G1** Community graphs with 1000 nodes and 31 communities;
- **G2** Sensor graphs with 1000 nodes;
- **G3** Hyper-cube graphs with 1002 nodes in 3-dimension.

We also perform experiments on the Minnesota network, which is a real-world graph with 2642 nodes. Artificial graph signals are assumed to be bandlimited with bandwidth \( K = 50 \). They are constructed by generating appropriate GFT coefficients: the non-zero GFT coefficients are randomly generated from the distribution \( N(1, 0.5^2) \), and the coefficients after \( K = 50 \) are zeros. The generated graph signals are corrupted by additional white Gaussian noise (AWGN) with different signal-to-noise ratios (SNRs). In experiments, we use the MATLAB’s built-in code to finish the eigen-decomposition work, rather than the LOBPCG method mentioned in the complexity analysis section.

### A. Static Global Sampling

1) Reconstruction MSE Evaluation: We compare the reconstruction MSE of our proposed GFS sampling algorithm to competing schemes that employ other criteria: SP in [12], E-optimal in [27], MIA in [31] and MFN in [26]. The truncation degree in MIA is \( L = 10 \). The SGWT toolbox [56] is adopted to approximate the ideal LP filter in the MIA method, where \( p = 25 \) and \( a = 30 \) [57]. For SP, the approximation order of cut-off frequency is \( k = 10 \). As described in Section IV, for GFS sampling, the shift parameter \( \mu \) is set to be \( 1/(\kappa_0 - 1) \), where we set \( \kappa_0 = 100 \) as the condition number constraint. The number of Given rotations matrices is \( J = 6N \log N \) for G1—G3 and \( J = 30N \log N \) for the Minnesota graph.

\[\text{All of these graphs are generated using GSP open source in [56].}\]
Fig. 3. Reconstruction MSE of different sampling algorithms in terms of sample size in different graph types, where the original graph signals are all recovered from samples by the LS reconstruction.

As shown in Fig. 3, our proposed GFS achieves lower MSE than three other competing sampling strategies, and closely approximates MFN’s performance in community, sensor and hyper-cube graphs at different SNRs. As discussed, MFN greedily minimizes MSE one node at a time directly, thus it needs to compute the first $K$ eigenvectors once and perform matrix inverse for each metric evaluation, as detailed in Section V. In contrast, our proposed strategy can obtain comparable performance with significantly lower complexity. Authors in paper [20] claimed that the solution from MFN is approximate near-optimal to the optimal solution. Therefore, the solution of our GFS method experimentally approximates the optimal solution of the original MSE problem.

Fig. 4 depicts the reconstruction MSE of different sampling methods in terms of the value of SNR on a given sample size. The averaged MSE values for all graph sampling methods decrease as the SNR increases. However, if we observe the detailed performance tendency in high SNR, the proposed sampling method still outperforms other three competing methods and well approximates the MFN’s performance.
For comparison, we simulate the proposed method, i.e., \( \mu = \beta \), at 0 dB. As illustrated in Fig. 8, at every sample size, the GFS reconstruction outperforms the LS reconstruction for different graphs when the shift parameter \( \beta \) is designed by (38). Fig. 9 demonstrates the performance comparison of different reconstruction methods in terms of SNR and it illustrates that the proposed reconstruction method is superior to the LS reconstruction when noise is large, i.e., SNR is small, which indicates the robustness of our proposed GFS reconstruction method.

Suppose that \( \beta = \mu = 1/(\kappa_0 - 1) \), the proposed GFS reconstruction sometimes achieves better performance than the LS solution. In this case, the GFS reconstruction does not compute any matrix inverse, thus enjoys lower complexity in recovery, as remarked in Section VII. Based on the experimental results in Fig. 8 and Table II, we can design the shift \( \mu \) in sampling and \( \beta \) in reconstruction both from equation (38).

Fig. 5 shows the reconstruction MSE of different sampling methods for the real-world Minnesota network. Our proposed GFS sampling also outperforms SP and E-optimal sampling, where the performance of MIA is too poor to be shown. For visualization, we further present one specific demonstration in Fig. 7. Fig. 7(a) and (b)–(f) show the original graph signal and the interpolated signal of different sampling methods, respectively. Fig. 7(b)–(f) demonstrate that the proposed GFS achieves the lowest MSE value among competing strategies, and the reconstructed signal is visually smoother with respect to the graph topology compared with signal recovered by SP and MIA.

2) Numerical Complexity Comparison: We evaluate the complexity of different sampling method listed in Table I via numerical simulations. Specifically, we compute the numerical complexity order value in Table I assuming \( M = 0.05N \). In these experiments, we set \( R = 10, T_1 = T_2(k) = 100 \) and \( \xi = O(N) \). Fig. 6 shows the simulated numerical complexity results which indicate the proposed GFS sampling has lower complexity than MFN, E-optimal and MIA, especially for large graphs.

3) Appropriateness of the Shift Parameter \( \mu \): The augmented A-optimality function will well approximate the original criterion (2) if \( \mu \) is sufficiently small. However, we customize a \( \mu \) based on condition number in the GFS sampling which would bring some approximation error. Table II shows the performance of the designed \( \mu = 1/(\kappa_0 - 1) \) approaches that of a extremely small \( \mu \). This indicates the proposed \( \mu \) for stable computation won’t bring a large performance gap between the original A-optimality value and the augmented one. In Section VII, we propose a new shift parameter \( \beta \) based on equation (38).

Table II: Reconstruction MSE of the Proposed GFS Algorithm in Different Shift \( \mu \) at 0 dB

| Graph | Sample size | 100 | 110 | 120 | 130 | 140 |
|-------|-------------|-----|-----|-----|-----|-----|
| G1    | 10^{-5}     | 16.10 | 14.55 | 13.43 | 12.44 | 11.63 |
|       | (38)        | 16.10 | 14.62 | 13.47 | 12.50 | 11.64 |
| G2    | 10^{-5}     | 20.77 | 18.68 | 17.09 | 15.77 | 14.63 |
|       | (38)        | 21.31 | 19.36 | 17.59 | 16.33 | 15.08 |
| G3    | 10^{-5}     | 23.09 | 20.79 | 18.99 | 17.48 | 16.19 |
|       | (38)        | 23.11 | 20.87 | 19.04 | 17.53 | 16.32 |

B. Static Graph Signal Reconstruction

Simulations on evaluating the proposed GFS reconstruction algorithm are performed subsequently, where the samples were all collected by the GFS sampling algorithm. Fig. 8 shows the reconstruction MSE of different recovery strategies in terms of sample size in G1–G3 at 0 dB. As illustrated in Fig. 8, at every sample size, the GFS reconstruction outperforms the LS reconstruction for different graphs when the shift parameter \( \beta \) is designed by (38). Fig. 9 demonstrates the performance comparison of different reconstruction methods in terms of SNR and it illustrates that the proposed reconstruction method is superior to the LS reconstruction when noise is large, i.e., SNR is small, which indicates the robustness of our proposed GFS reconstruction method.

To make the GFS reconstruction not perform extremely bad on it. We propose one strategy to recognize bad subsets, see appendix D for details, on which we do not collect samples.
Fig. 7. One specific demonstration of interpolated graph signal of different sampling algorithms at SNR = 0 dB on Minnesota graph with $M = 90$.

Fig. 8. Experimental results of different recovery strategies with respect to sample size in different graph types at SNR = 0 dB. The original graph signals are all sampled via GFS algorithm.

Fig. 9. Experimental results of different recovery strategies with respect to SNR in different graph types. The original graph signals are all sampled via GFS algorithm.
from scratch. However, it has to compute the initial objective value of the SA set to proceed sampling, and its performance will be necessarily be no better than the “start-over” approach. Hence our method will also outperform competing schemes with unavailable node replacement.

IX. CONCLUSION

Graph sampling with noise remains a challenging problem: MMSE leads to the known A-optimality criterion for independent noise, which is expensive to evaluate and difficult to optimize. In this paper, we propose an augmented objective based on Neumann series expansion to approximate the A-optimality criterion, which can be expressed as a function of an ideal LP graph filter, efficiently approximated via fast graph Fourier transform. Using the augmented objective, we select nodes greedily without any matrix inverse computation based on a matrix inversion lemma. Further, we extend our sampling scheme to the dynamic network case, where the availability of nodes is time-varying. For signal recovery, we design an accompanied signal reconstruction strategy to obtain a biased but robust estimator. Experimental results validate the superiority of the proposed sampling strategy compared with existing schemes and demonstrate the effectiveness of our biased recovery algorithm over unbiased LS reconstruction.

APPENDIX A

PROOF OF PROPOSITION 1

From the definition of $C$, $C^\top C = [I_\delta \ 0]$ under appropriate permutation. Then, $\forall x \in \mathbb{R}^K$ and $\|x\|_2 = 1$,

$$x^\top \Psi x = (V_K x)^\top (C^\top C) (V_K x) = b^\top \begin{bmatrix} I_\delta & 0 \\ 0 & 0 \end{bmatrix} b = b_\delta^\top b_\delta, \quad (39)$$

where $b = V_K x$.

Since $b^\top b = (V_K x)^\top (V_K x) = 1$, $0 \leq x^\top \Psi x \leq 1$. Because rank($CV_K$) = $K$, $\Psi$ is positive definite, which results in $0 < x^\top \Psi x \leq 1$ and $0 \leq x^\top \Phi x < 1$. Due to the Rayleigh quotient theorem, $0 \leq \delta_i < 1$.

APPENDIX B

PROOF OF LEMMA 2

If we view matrix $\tilde{V}_K$ as $\tilde{V}_K = [t_1 \ t_2 \ldots \ t_N]^\top$ with $t_i \in \mathbb{R}^K$, then $(CV_K)^\top CV_K = \sum_{i \in S} t_i t_i^\top$. Define $Z(S) = \sum_{i \in S} t_i t_i^\top + \mu I$, we know that $Z(S \cup \{j\}) = Z(S) + t_j t_j^\top$ for any $j \notin S$ and $g(S) = \text{Tr}[Z(S)^{-1}]$.

i) Monotonic decreasing

$$g(S \cup \{j\}) - g(S) = \text{Tr} \left[ (Z(S) + t_j t_j^\top)^{-1} - Z(S)^{-1} \right]$$

for any $j \notin S$.

From Lemma 3, we know that

$$g(S \cup \{j\}) - g(S) = -\text{Tr} \left[ \frac{Z(S)^{-1} t_j t_j^\top Z(S)^{-1}}{1 + t_j^\top Z(S)^{-1} t_j} \right]$$

$$= -\frac{\|Z(S)^{-1} t_j\|^2}{1 + t_j^\top Z(S)^{-1} t_j}. \quad (40)$$

It is easy to prove the eigenvalues of $Z(S)$ are in $(\mu + \mu, 1 + \mu]$ for any $S$ from the similar derivation in Proposition 1 [31]. Therefore, $t_j^\top Z(S)^{-1} t_j \geq \lambda_{\text{max}}^{-1} Z(S) \|t_j\|^2 \geq \frac{t_j^\top Z(S)^{-1} t_j}{1 + \mu} > 0$ from Rayleigh quotient theorem. Using this result and combining it with the above two equations, we know that $g(S \cup \{j\}) - g(S) \leq 0$ for any $j \notin S$ which implies the set function $g$ is monotonic decreasing.

ii) $\alpha$-supermodular

We first present the definition of $\alpha$-supermodularity introduced in paper [20].

Definition 3: A set function $g : 2^V \to \mathbb{R}$ is $\alpha$-supermodular if for all sets $A \subseteq B \subseteq V$ and all $j \notin B$, the following equation holds for some $\alpha \geq 0$

$$g(A \cup \{j\}) - g(A) \leq \alpha \left[ g(B \cup \{j\}) - g(B) \right]. \quad (41)$$

$\alpha$-supermodularity is only of interest when $\alpha$ takes the largest value [20]:

$$\alpha = \min_{\delta_j \in \mathbb{R}^+} \frac{g(A \cup \{j\}) - g(A)}{g(B \cup \{j\}) - g(B)}$$

$$= \min_{\delta_j \in \mathbb{R}^+} \frac{1 + t_j^\top Z(B)^{-1} t_j}{1 + t_j^\top Z(A)^{-1} t_j}, \quad (42)$$
Since \( \lambda_{\max}[Z(S)] \leq \sum_j t_j \leq \lambda_{\min}[Z(S)] \), the lower bound of \( \alpha \) is
\[
\alpha \geq \frac{1 + \lambda_{\max}[Z(E)] \|t_j\|_2^2}{1 + \lambda_{\min}[Z(A)] \|t_j\|_2^2} \geq \frac{\lambda_{\max}[Z(A)] \|t_j\|_2^2}{\lambda_{\min}[Z(E)] \|t_j\|_2^2}.
\]
(43)

As we claimed, the eigenvalues of \( Z(S) \) are in \( (\mu, 1 + \mu) \), so
\[
\alpha \geq \left( \frac{(1 + \mu)^{-1} \|t_j\|_2^2}{1 + \mu^{-1} \|t_j\|_2^2} \right) \frac{(1 + \mu - \mu^{-1})}{\mu^{-2}} = \alpha'.
\]
(44)

It is easy to drive that
\[
\frac{\partial \alpha'}{\partial \|t_j\|_2^2} = \frac{(1 + \mu)^{-2} - (1 + \mu)^{-1} - \mu^{-1}}{(1 + \mu^{-1} \|t_j\|_2^2)^2} < 0,
\]
which demonstrates the function \( \alpha' \) is decreasing in terms of \( \|t_j\|_2^2 \). Since \( t_j \) is a row of \( V_K \), i.e., \( \|t_j\|_2 \leq 1 \), we have
\[
\alpha \geq \frac{(1 + \mu)^{-1} (1 + \mu - \mu^{-1})}{1 + \mu^{-1} \|t_j\|_2^2} \geq \frac{\mu^3 (\mu + 2)}{(\mu + 1)^4} > 0.
\]
(46)

**APPENDIX C**

**PROOF OF PROPOSITION 4**

Assume the eigen-decomposition of \( \Psi \) is \( \Psi = U \Lambda U^\top \) where \( \Lambda = \text{diag}\{\sigma_i\} \) and \( UU^\top = I \). We have derived in Appendix B in the paper [31] that
\[
E \|\hat{x}_{1:S} - x\|_2^2 = \omega^2 \sum_{i=1}^K \frac{1}{1 - \delta_i} = \omega^2 \sum_{i=1}^K \frac{1}{\sigma_i}.
\]
(47)

Rewrite the GFS reconstruction \( \hat{x} = V_K [H \Psi K + \beta I]^{-1} [H \Psi K + \beta I]^{-1} y_S \) by \( \hat{x} = H y_S = H C(x + n) \), then the bias of this estimator is
\[
\text{Bias}(\hat{x}; \beta) = \|E[\hat{x}] - x\|_2 = \|HC - I\|_2.
\]
(48)

Recall that the original graph signal was assumed \( K \)-bandlimited, so
\[
\text{Bias}^2(\hat{x}; \mu) = \|H C V_K - V_K \hat{x}\|_2^2.
\]
(49)

in which
\[
(H C V_K - V_K)^\top (H C V_K - V_K)
= [V_K (\Psi + \beta I)^{-1} \Psi - V_K] [V_K (\Psi + \beta I)^{-1} \Psi - V_K]
= \Psi (\Psi + \beta I)^{-2} \Psi (\Psi + \beta I)^{-1} - (\Psi (\Psi + \beta I)^{-1} + I)
= \text{Udiag} \left( \frac{\sigma^2}{\sigma_i + \beta} \right) \text{U}^\top
= \text{Udiag} \left( \frac{\sigma_i}{\sigma_i + \beta} \right)^{-2} \text{U}^\top
= \sum_{i=1}^K \left( \frac{\sigma_i}{\sigma_i + \beta} \right)^{-2} u_i u_i^\top.
\]

Therefore,
\[
\text{Bias}^2(\hat{x}; \beta) = \sum_{i=1}^K \left( 1 + \frac{\sigma_i}{\beta} \right)^{-2} (u_i^\top x_K)^2.
\]
(50)

The variance part of this solution can be derived from its covariance matrix: \( \text{cov}(\hat{x}) = E(\hat{x}E[\hat{x}]) = E[H C n n^\top C^\top I^\top] = E[H C n n^\top C^\top I^\top] \). Because \( E[n n^\top] = \omega^2 I \) and \( C^\top I^\top = I \), the variance of \( \hat{x} \) is
\[
\text{Var}(\hat{x}; \beta) = \text{Tr}[\text{cov}(\hat{x})] = \omega^2 \text{Tr}[HH^\top].
\]
(52)

From the definition of \( H \) and the property of trace operator,
\[
\text{Tr}[HH^\top] = \text{Tr} \left[ V_K (\Psi + \beta I)^{-1} \Psi (\Psi + \beta I)^{-1} V_K \right]
= \text{Tr} \left[ U \Lambda U^\top U (\Lambda + \beta I)^{-1} U \right]
= \text{Tr} \left[ \text{Udiag} \left( \frac{\sigma_i}{\sigma_i + \beta} \right) \right] \text{U}^\top
= \sum_{i=1}^K \frac{\sigma_i}{\sigma_i + \beta}.
\]
(51)

Therefore, the variance of \( \hat{x} \) is
\[
\text{Var}(\hat{x}; \beta) = \omega^2 \sum_{i=1}^K \frac{\sigma_i}{\sigma_i + \beta}.
\]
(52)

**APPENDIX D**

**RECOGNIZE THE BAD INITIAL AVAILABLE SUBSET**

Space \( PW_{\eta}(G) \) contains all bandlimited graph signals with bandwidth less than \( \eta \) and space \( L_2(S^c) \) covers graph signals fulfilling \( f(S) = 0 \) [25]. Authors in [12] claimed that for set \( S \) to be a uniqueness set of space \( PW_{\eta}(G) \), \( \eta \) needs to be less than the minimum bandwidth of all signals in space \( L_2(S^c) \), which is defined as the cutoff frequency of \( L_2(S^c) \). Thus, a set \( S \) to be good means its cutoff frequency should be as large as possible. They defined the cutoff frequency of space \( L_2(S^c) \) as
\[
\gamma_c(S) = \min_{f \in L_2(S^c)} \gamma(f),
\]
(53)

where \( \gamma(f) \) is the frequency of a specific graph signal \( f \).

A strategy was proposed in [12] to approximate the cutoff frequency \( \gamma_c(S) \) by
\[
\Omega_k(S) = \min_{f \in L_2(S^c)} \left( \frac{\|L_k f\|}{\|f\|} \right)^{1/k} = \min_{f \in L_2(S^c)} \left( \frac{\|L_k f\|}{\|f\|} \right)^{1/2k} = \lambda_{1:1,1}^{1/2k},
\]
(54)

where \( \lambda_{1:1} \) is the smallest eigenvalue of the reduced matrix \( (L_k)^{1/2} L_k (L_k)^{1/2} S^c \).

The approximate cutoff frequency can be used for judging if a set is good for initialization because a reasonable set won’t have small \( \Omega_k(S) \). We take this as a criterion to detour a bad initial subset. During experiments, the size of the available set is
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and Associate Editor, Dr. Pierre Borgnat, of this paper for their constructive comments. Moreover, the authors appreciate the patient assistance from Dr. Yan Wang for conducting the simulations.

REFERENCES

[1] A. Ortega, P. Frossard, J. Kovacevic, J. M. Moura, and P. Vandergheynst, “Graph signal processing: Overview, challenges, and applications,” Proc. IEEE, vol. 106, no. 5, pp. 808–828, 2018.

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

[3] G. Cheung, E. Magli, Y. Tanaka, and M. K. Ng, “Graph spectral image processing.” Proc. IEEE, vol. 106, no. 5, pp. 907–930, 2018.

[4] A. Gavili and X.-P. Zhang, “On the shift operator, graph frequency, and optimal filtering in graph signal processing,” IEEE Trans. Signal Process., vol. 63, no. 22, pp. 6503–6518, 2015.

[5] D. K. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” Appl. Comput. Harmon. Anal., vol. 30, no. 2, pp. 129–150, 2011.

[6] D. B. Tay and A. Ortega, “Bipartite graph filter banks: Polyphase analysis and generalization,” IEEE Trans. Signal Process., vol. 65, no. 18, pp. 4833–4846, 2017.

[7] A. Sandryhaila and J. M. Moura, “Big data analysis with signal processing on graphs,” IEEE Signal Process. Mag., vol. 31, no. 5, pp. 80–90, 2014.

[8] B. Girault, A. Ortega, and S. S. Narayanan, “Irregularity-aware graph fourier transforms,” IEEE Trans. Signal Process., vol. 66, no. 21, pp. 5746–5761, 2018.

[9] S. Sandellitti, S. Barbarossa, and P. Di Lorenzo, “On the graph fourier transform for directed graphs,” IEEE J. Sel. Topics Signal Process., vol. 11, no. 6, pp. 796–811, 2017.

[10] H. Shen et al., “Missing information reconstruction of remote sensing data: A technical review,” IEEE Geosci. Remote Sens. Mag., vol. 3, no. 3, pp. 61–85, 2015.

[11] S. A. Danziger et al., “Predicting positive p53 cancer rescue regions using most informative positive (MIP) active learning,” PLOS Comput. Biol., vol. 5, no. 9, Sep. 4, 2009.

[12] A. Anis, A. Gaddde, 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, 2016.

[13] 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.

[14] W. X. Wang, J. Chen, and Y. Gu, “Local measurement and reconstruction for noisy bandlimited graph signals,” Signal Process., vol. 129, pp. 119–129, 2016.

[15] D. Valsesia, G. Fracastoro, and E. Magli, “Sampling of graph signals via randomized local aggregations,” IEEE Trans. Signal Inf. Process. Over Netw., 2018.

[16] H. Shomorony and A. S. Avestimehr, “Sampling large data on graphs,” in Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP), 2014, pp. 933–936.

[17] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[18] R. A. Horn, R. A. Horn, and C. R. Johnson, Matrix Analysis. Cambridge, U.K.: Cambridge Univ. Press, 1990.
A. Hashemi, M. Ghasemi, H. Vikalo, and U. Topcu, “Randomized greedy sensor selection: Leveraging weak submodularity,” arXiv preprint arXiv:1807.08627, 2018.

S. P. Chepuri and G. Leus, “Graph sampling for covariance estimation,” IEEE Trans. Signal Inf. Process. Over Netw., vol. 3, no. 3, pp. 451–466, 2017.

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. (ICASSP), Mar. 2016, pp. 6225–6229.

P.-Y. Chen and D. Wei, “On the supermodularity of active graph-based semi-supervised learning with stieltjes matrix regularization,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2018, pp. 2801–2805.

S. Chen, A. Sandryhaila, and J. Kovačević, “Sampling theory for graph signals,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2015, pp. 3392–3396.

S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

F. Pukelsheim, Optimal Design of Experiments. Philadelphia, PA, USA: SIAM, 1993, vol. 50.

G. H. Golub and H. A. Van der Vorst, “Eigenvalue computation in the 20th century,” in Numerical Analysis: Historical Developments in the 20th Century. Amsterdam, The Netherlands, Elsevier, 2001, pp. 209–239.

J. W. Demmel, Applied Numerical Linear Algebra. Philadelphia, PA, USA: SIAM, 1997, vol. 56.

D. S. Bernstein, Matrix Mathematics. Princeton, NJ, USA: Princeton Univ. Press, 2005.

S. Fujishige, Submodular Functions and Optimization. Amsterdam, The Netherlands: Elsevier, 2005, vol. 58.

A. V. Knyazev, “Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method,” SIAM J. Scientific Comput., vol. 23, no. 2, pp. 517–541, 2001.

N. Perraudin et al., “Gspbox: A toolbox for signal processing on graphs,” arXiv preprint arXiv:1408.5781, 2014.

A. Gaddde, A. Anis, and A. Ortega, “Active semi-supervised learning using sampling theory for graph signals,” in Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2014, pp. 492–501.

Gene Cheung (M’00–SM’07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively.

He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He was an associate professor in National Institute of Informatics in Tokyo, Japan, from 2009 till 2018. He is now an associate professor in York University, Toronto, Canada.

His research interests include 3D imaging and graph signal processing. He has served as associate editor for multiple journals, including IEEE TRANSACTIONS ON MULTIMEDIA (2007–2011), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–2017) and IEEE TRANSACTIONS ON IMAGE PROCESSING (2015–present). He served as a member of the Multimedia Signal Processing Technical Committee (MMSP-TC) in IEEE Signal Processing Society (2012–2014), and a member of the Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP-TC) (2015–2017, 2018–2020). He is a co-author of several paper awards, including the best student paper award in ICIP 2013, ICIP 2017 and IVMSP 2016, best paper runner-up award in ICME 2012, and IEEE Signal Processing Society (SPS) Japan best paper award 2016. He is a recipient of the Canadian NSERC Discovery Accelerator Supplement (DAS) 2019.

Gene Cheung (M’00–SM’07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively.

He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He was an associate professor in National Institute of Informatics in Tokyo, Japan, from 2009 till 2018. He is now an associate professor in York University, Toronto, Canada.

His research interests include 3D imaging and graph signal processing. He has served as associate editor for multiple journals, including IEEE TRANSACTIONS ON MULTIMEDIA (2007–2011), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–2017) and IEEE TRANSACTIONS ON IMAGE PROCESSING (2015–present). He served as a member of the Multimedia Signal Processing Technical Committee (MMSP-TC) in IEEE Signal Processing Society (2012–2014), and a member of the Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP-TC) (2015–2017, 2018–2020). He is a co-author of several paper awards, including the best student paper award in ICIP 2013, ICIP 2017 and IVMSP 2016, best paper runner-up award in ICME 2012, and IEEE Signal Processing Society (SPS) Japan best paper award 2016. He is a recipient of the Canadian NSERC Discovery Accelerator Supplement (DAS) 2019.

Gene Cheung (M’00–SM’07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively.

He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He was an associate professor in National Institute of Informatics in Tokyo, Japan, from 2009 till 2018. He is now an associate professor in York University, Toronto, Canada.

His research interests include 3D imaging and graph signal processing. He has served as associate editor for multiple journals, including IEEE TRANSACTIONS ON MULTIMEDIA (2007–2011), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–2017) and IEEE TRANSACTIONS ON IMAGE PROCESSING (2015–present). He served as a member of the Multimedia Signal Processing Technical Committee (MMSP-TC) in IEEE Signal Processing Society (2012–2014), and a member of the Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP-TC) (2015–2017, 2018–2020). He is a co-author of several paper awards, including the best student paper award in ICIP 2013, ICIP 2017 and IVMSP 2016, best paper runner-up award in ICME 2012, and IEEE Signal Processing Society (SPS) Japan best paper award 2016. He is a recipient of the Canadian NSERC Discovery Accelerator Supplement (DAS) 2019.

Gene Cheung (M’00–SM’07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively.

He was a senior researcher in Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He was an associate professor in National Institute of Informatics in Tokyo, Japan, from 2009 till 2018. He is now an associate professor in York University, Toronto, Canada.

His research interests include 3D imaging and graph signal processing. He has served as associate editor for multiple journals, including IEEE TRANSACTIONS ON MULTIMEDIA (2007–2011), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016–2017) and IEEE TRANSACTIONS ON IMAGE PROCESSING (2015–present). He served as a member of the Multimedia Signal Processing Technical Committee (MMSP-TC) in IEEE Signal Processing Society (2012–2014), and a member of the Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP-TC) (2015–2017, 2018–2020). He is a co-author of several paper awards, including the best student paper award in ICIP 2013, ICIP 2017 and IVMSP 2016, best paper runner-up award in ICME 2012, and IEEE Signal Processing Society (SPS) Japan best paper award 2016. He is a recipient of the Canadian NSERC Discovery Accelerator Supplement (DAS) 2019.