Quantization-aware sampling set selection for bandlimited graph signals

Yoon Hak Kim[*]

*Correspondence:
yhk@chosun.ac.kr

Department of Electronic
Engineering, College of IT
Convergence Engineering,
Chosun University, 309
Pilmun-daero, Dong-gu,
61452 Gwangju, Republic
of Korea

Abstract

We consider a scenario in which nodes of a graph are sampled for bandlimited graph signals which are uniformly quantized with optimal rate and original signals are reconstructed from the quantized signal values residing on the nodes in the sampling set. We seek to construct the best sampling set in a greedy manner that minimizes the average reconstruction error. We manipulate the reconstruction error by using the QR factorization and derive an analytic result stating that the next minimizing node can be iteratively selected by finding the one that minimizes the geometric mean of the row vectors of the inverse upper triangular matrix $\mathbf{R}^{-1}$ in the QR factorization. We also compare the complexity of the proposed algorithm with different sampling methods and evaluate the performance of the proposed algorithm by experiments for various graphs, showing the superiority to the existing sampling methods when quantization is involved.

Keywords: Graph signal processing, Sampling set selection, Greedy algorithms, Quantization, QR factorization

1 Introduction

High-dimensional and irregularly structured data are generated from emerging applications such as neural, energy, transportation, social and sensor networks. Graph signal processing (GSP) techniques have been developed to efficiently deal with such data samples on the network nodes by regarding them as graph signals on the vertices of the graph and by describing similarity between data samples on two nodes by the weight of the edge linking the nodes [1, 2]. Among the GSP techniques, sampling a subset of vertices or nodes of the graph is of utmost importance to process such high-dimensional network data.

It should be noticed that the problem of finding an optimal set of nodes of networks in noisy situations is combinatorial, requiring a prohibitive cost of computation [3, 4]. Heuristic approaches such as convex relaxation [3] and cross-entropy optimization [4] were presented to achieve the reduced complexity which is still unfeasible especially for large-scale networks. To ensure good performance with feasible complexity, greedy approach has been adopted for many practical applications and various greedy sampling algorithms have been presented in [5–13]: Greedy selection methods were developed

such that the worst case of the reconstruction error is iteratively minimized [5–7] and the uncertainty principle for graph signals was derived and sampling strategies based on the uncertainty principle were suggested in [8]. Algorithms without eigendecomposition were presented to show fast and comparable performance [9, 10]. Greedy samplers were proposed to focus on reconstruction of the second-order statistics of graph signals from sampled graph signals since the second-order statistics (i.e., graph power spectrum) provides critical information in inference applications such as inpaining and prediction [11]. To evaluate the reconstruction performance of sampling methods, universal bounds were presented and greedy approaches minimizing the reconstruction error were shown to achieve a near-optimality by the concept of approximate supermodularity [12]. Recently, a QR factorization-based method that greedily selects one node at each iteration so as to directly minimize the reconstruction error was developed with a competitive complexity and reconstruction performance [13].

The sampling set can be also constructed by the greedy solutions to the sensor selection problem in which a subset of nodes in sensor networks is selected in a greedy manner such that the reconstruction error for parameter estimation or a well-defined proxy related to the reconstruction error is optimized [14, 15]: The log-determinant of the inverse estimation error covariance matrix is maximized to find the sampling set with a guaranteed near-optimality [14] and the frame potential as a proxy for the reconstruction error was introduced as a metric and a greedy removal algorithm was proposed to achieve near-optimality [15]. An efficient greedy technique was presented to find the least number of sensor nodes and their sensing locations by employing a new criterion with maximum projection onto the minimum eigenvalue of the dual observation matrix [16]. Quantization of graph signals at nodes before transmission is inevitable since nodes or vertices in physical networks operate with energy constraint and limited communication bandwidth. Hence, the effect of quantization of graphs signals on the selection process should be addressed while the previous work [17] derived an analytic solution to the problem of allocating optimal rate to each node in the sampling set where uniform quantization at nodes are assumed.

In this work, we consider a graph sampling problem formulated in previous work [5–15] for network applications such as social, energy, neural and sensor networks. We aim to devise an efficient sampling set selection algorithm for bandlimited graph signals which are uniformly quantized by using optimal rate. Specifically, we incorporate quantization into the selection process and seek to greedily select one node at each iteration that minimizes the average reconstruction error computed from quantized signal values with optimal rate allotted into the nodes in the sampling set. Note that the previous work [17] focused on optimal rate allocation to the sampling set after the set is constructed by selection methods. To the best of our knowledge, this is the first to take into account quantization to determine the best sampling set that minimizes the average reconstruction error. We first apply the QR factorization to formulate the reconstruction error and present a simple criterion to select the next minimizing node at iterations by using the analytic results proved in [13] and the optimal rate solution derived in [17]. Specifically, we show that the reconstruction error is iteratively minimized by selecting the next node minimizing the geometric mean of the row vectors of the inverse upper triangular matrix $\mathbf{R}^{-1}$ in the QR factorization. We also discuss the complexity of the proposed algorithm in

comparison with different novel sampling methods. Finally, we investigate through experiments the performance of the proposed algorithm, demonstrating a performance gain over previous selection methods for quantized signals on various graphs.

This paper is organized as follows. The sampling set selection problem with quantization is formulated in Sect. 2. The reconstruction error obtained from quantized graph signals is manipulated by using the QR factorization to produce a simplified metric, and an analytic result is derived to propose a simple selection criterion in Sect. 3. The complexity of the proposed algorithm is analyzed in comparison with different sampling methods in Sect. 4.1. The reconstruction performance is evaluated by extensive experiments in Sect. 4.2 and conclusions in Sect. 5.

2 Problem formulation

For an undirected and weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes indexed by the set $\mathcal{V} = \{1, \ldots, N\}$ and edges $\mathcal{E} = \{(i, j, w_{ij})\}$, where w_{ij} represents the edge weight from node i to node j, we consider a graph signal $\mathbf{f} = [f_1, \ldots, f_N]^\top$ which is a function defined on $\mathcal{V}$ with signal values f_i on the ith vertex. We employ variation operators (e.g., combinatorial graph Laplacian, normalized Laplacian) to describe the signal variation in a graph caused by the connectivity of the graph [2, 6]. Let $\mathbf{L}$, $N \times N$ matrix be a variation operator with eigenvalues $|\lambda_1| \leq \ldots \leq |\lambda_N|$ and corresponding orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$. Noting that an arbitrary graph signal $\mathbf{f}$ can be expressed as $\mathbf{f} = \mathbf{Uc}$ where $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_N]$ is the eigenvector matrix and $\mathbf{c} = [c_1, \ldots, c_N]^\top$ the graph Fourier transform (GFT) of $\mathbf{f}$, ω-bandlimited graph signal $\mathbf{f}$ (i.e., $c_i = 0, |\lambda_i| > \omega, \forall i > r$) can be given by

$$\mathbf{f} = \sum_{i=1}^{r} c_i \mathbf{u}_i = \mathbf{U}_{\mathcal{VR}} \mathbf{c}_{\mathcal{R}} \tag{1}$$

where $\mathbf{U}_{\mathcal{VR}}$ is an $N \times r$ matrix with rows of $\mathbf{U}$ indexed by $\mathcal{V}$ and columns of $\mathbf{U}$ indexed by $\mathcal{R} = \{1, \ldots, r\}$, and $\mathbf{c}_{\mathcal{R}}$ an $r \times 1$ column vector with its entries of $\mathbf{c}$ indexed by $\mathcal{R}$.

We consider a situation in which a given number of nodes are selected to obtain a sampling set $\mathcal{S}$ and the sampled signal $\mathbf{f}_{\mathcal{S}}$ with the entries f_i of $\mathbf{f}$ indexed by $\mathcal{S}$ and the signal value f_i at node i is uniformly quantized with a rate of R_i bits, given the total rate budget $R_T = \sum_i^{|S|} R_i$. Then, we seek to find the best sampling set that minimizes the average reconstruction error $\mathbb{E}[\| \mathbf{f} - \hat{\mathbf{f}}^Q \|^2]$ where $\hat{\mathbf{f}}^Q$ is the reconstructed signal from the quantized signal $\mathbf{f}_{\mathcal{S}}^Q$ which is obtained by uniformly quantizing the sampled signal $\mathbf{f}_{\mathcal{S}}$. It is assumed that $\mathbf{f}_{\mathcal{S}}$ is corrupted by an additive measurement noise $\mathbf{n} \in \mathbb{R}^{|\mathcal{S}|}$ consisting of independent and identically distributed (iid) entries with zero mean and variance σ^2:

$$\mathbf{f}_{\mathcal{S}}^Q = \mathbf{f}_{\mathcal{S}} + \mathbf{n} + \mathbf{e}, \tag{2}$$

where the quantization error vector $\mathbf{e} = [e_1, \ldots, e_{|\mathcal{S}|}]^\top$ is modeled as an additive noise with its entry e_i iid with zero mean and the variance $\frac{\Delta_i^2}{12}$ where Δ_i is the quantization step size given by $\Delta_i = \frac{\mathcal{I}}{2^{R_i}}$ with the quantization interval $\mathcal{I}$ and this additive noise model for quantization error is commonly employed for the analysis of quantization error since it behaves in a similar manner to that of additive white noise [18]. In this work, we produce

the uniqueness sampling set $\mathcal{S}$ defined as the set for ω-bandlimited signals if noise-free ω-bandlimited signals $\mathbf{f}$ can be perfectly reconstructed from the sampled signal $\mathbf{f}_{\mathcal{S}}$, implying that every ω-bandlimited signal has its unique samples in the uniqueness set $\mathcal{S}$ and it is shown that the uniqueness set $\mathcal{S}$ can be constructed by choosing r independent row vectors of $\mathbf{U}_{\mathcal{V}\mathcal{R}}$ (i.e., choosing the ith row vector corresponds to selecting the ith node) [6].

Letting $\mathbf{U}_{\mathcal{S}\mathcal{R}}$ be the $|\mathcal{S}| \times r$ matrix with $|\mathcal{S}|$ independent rows of $\mathbf{U}_{\mathcal{V}\mathcal{R}}$ indexed by $\mathcal{S}$, the quantized noisy signal $\mathbf{f}_{\mathcal{S}}^Q$ can be given from (1) and (2):

$$\mathbf{f}_{\mathcal{S}}^Q = \mathbf{U}_{\mathcal{S}\mathcal{R}}\mathbf{c}_{\mathcal{R}} + \mathbf{n} + \mathbf{e}. \tag{3}$$

To estimate $\mathbf{c}_{\mathcal{R}}$ in (3) from the quantized noisy signals $\mathbf{f}_{\mathcal{S}}^Q$, we let $\hat{\mathbf{c}}_{\mathcal{R}} = \mathbf{U}_{\mathcal{S}\mathcal{R}}^+ \mathbf{f}_{\mathcal{S}}^Q$ where $\mathbf{U}_{\mathcal{S}\mathcal{R}}^+$ is the pseudo-inverse of $\mathbf{U}_{\mathcal{S}\mathcal{R}}$. In this work, we use the notation of the pseudo-inverse for non-invertible matrices. Then, the reconstructed signal $\hat{\mathbf{f}}^Q$ is expressed by

$$\hat{\mathbf{f}}^Q = \mathbf{U}_{\mathcal{V}\mathcal{R}}\hat{\mathbf{c}}_{\mathcal{R}} = \hat{\mathbf{f}} + \mathbf{U}_{\mathcal{V}\mathcal{R}}\mathbf{U}_{\mathcal{S}\mathcal{R}}^+ \mathbf{e}$$

where $\hat{\mathbf{f}} = \mathbf{U}_{\mathcal{V}\mathcal{R}}\mathbf{U}_{\mathcal{S}\mathcal{R}}^+(\mathbf{f}_{\mathcal{S}} + \mathbf{n})$ is the reconstructed signal from the noisy sampled signal $\mathbf{f}_{\mathcal{S}}$. Now, we can obtain the average reconstruction error by using the analytic results in [17] as follows:

$$\mathbb{E}[\| \mathbf{f} - \hat{\mathbf{f}}^Q \|^2] \approx \mathbb{E}[\| \hat{\mathbf{f}} - \hat{\mathbf{f}}^Q \|^2] = tr[(\mathbf{U}_{\mathcal{S}\mathcal{R}}^\top \mathbf{E}_Q^{-1} \mathbf{U}_{\mathcal{S}\mathcal{R}})^{-1}], \tag{4}$$

where $\mathbf{E}_Q = \mathbb{E}[\mathbf{e}\mathbf{e}^\top]$ a diagonal covariance matrix with its ith diagonal element $\frac{\Delta_i^2}{12}$. It should be noticed that in computing the average reconstruction error in (4), it is assumed that $\mathbf{f} \approx \hat{\mathbf{f}}$, implying the noise-free sampled graph signal $\mathbf{f}_{\mathcal{S}}$ or the case of high signal-to-noise ratio (SNR). This assumption ensures no prior distributions of graph signals required in the metric to be minimized, leading to a simplified process. However, since graph signals are typically noise-corrupted in practical applications, the proposed algorithm is evaluated in comparison with different methods in noisy situations in Sect. 4.2.

We continue to formulate the constrained optimization problem to find the best sampling set S^* that minimizes the average reconstruction error:

$$\mathcal{S}^* = \arg \min_{\mathcal{S},|\mathcal{S}|=r} tr[(\mathbf{U}_{\mathcal{S}\mathcal{R}}^\top \mathbf{E}_Q^{-1} \mathbf{U}_{\mathcal{S}\mathcal{R}})^{-1}], \quad \sum_{i=1}^{|\mathcal{S}|} R_i \leq R_T$$

In this work, since we construct $\mathcal{S}^*$ by greedily selecting one node at a time, we aim to minimize the intermediate reconstruction error at the ith iteration given by $tr[(\mathbf{U}_{\mathcal{S}_i\mathcal{R}}^\top (\mathbf{E}_Q^i)^{-1} \mathbf{U}_{\mathcal{S}_i\mathcal{R}})^+]$, where $(\mathbf{E}_Q^i)$ is the $i \times i$ diagonal covariance matrix with its entries, $\frac{\Delta_j^2}{12}, j = 1, \ldots, i$ and $\mathcal{S}_i$ the set of i nodes selected. Specifically, at the $(i+1)$th iteration, one node is selected out of the nodes in the complement of set $\mathcal{S}_i$ which consists of the remaining vertices in $\mathcal{V}$ denoted by $\mathcal{S}_i^C$ such that $\mathcal{S}_i \cup \mathcal{S}_i^C = \mathcal{V}$ and $\mathcal{S}_i \cap \mathcal{S}_i^C = \emptyset$: in other words, one independent row vector $(\mathbf{u}^{(i+1)})^\top$ is selected from the remaining rows in $\mathbf{U}_{\mathcal{S}_i^C\mathcal{R}}$ where $\mathbf{u}^{(i)}$ indicates the transpose of the row vector selected at the ith iteration. It should be noted that selection of the minimizing row at the $(i+1)$th iteration is

conducted after the optimal rate is allotted to each node in $\mathcal{S}_{i+1}$ with the intermediate total rate $R_{T_{i+1}}$ assumed to be uniformly incremented by using $R_{T_{i+1}} = R_{T_i} + \frac{R_T}{|\mathcal{S}|}$. More specifically,

$$
k^* = \arg \min_{\mathcal{S}_{i+1} = \mathcal{S}_i + \{k\}, k \in \mathcal{S}_i^C} tr[(\mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}}^\top (\mathbf{E}_Q^{i+1})^{-1} \mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}})^+]
$$

$$
\text{subject} \quad \text{to} \sum_{j=1}^{i+1} R_j \leq R_{T_{i+1}}
$$

(5)

$$
\mathcal{S}_{i+1} = \mathcal{S}_i + \{k^*\}, \quad i = 0, \ldots, |\mathcal{S}| - 1.
$$

(6)

where $\mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}}^\top = [\mathbf{u}^{(1)} \cdots \mathbf{u}^{(i+1)}]$ and it is assumed that $\mathbf{u}^{(i)}$'s are independent of each other. The selection process in (5) and (6) is iteratively conducted r times with $\mathcal{S}_i$ replaced by $\mathcal{S}_{i+1}$ at the next iteration. Note that the rate constraint includes both inequality and equality and all of the rate budget available at each iteration is fully used to minimize the reconstruction error.

3 Method: sampling set selection algorithm

We apply the QR factorization to $\mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}}^\top$ in (5) to generate $\mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}}^\top = \mathbf{Q}^{i+1}\mathbf{R}^{i+1}$ where $\mathbf{Q}^{i+1}$ is the $r \times (i+1)$ matrix with $(i+1)$ orthonormal columns $\mathbf{q}_1, \ldots, \mathbf{q}_{i+1}$ and $\mathbf{R}^{i+1}$ the $(i+1) \times (i+1)$ upper triangular matrix with columns $\mathbf{r}_1, \ldots, \mathbf{r}_{i+1}$. Note that in this work, the QR factorization is performed by the Householder transformation due to its lower complexity and more stability than the Gram–Schmidt orthogonalization [19]. Then, we have

$$
tr\left[(\mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}}^\top (\mathbf{E}_Q^{i+1})^{-1} \mathbf{U}_{\mathcal{S}_{i+1}\mathcal{R}})^+\right]
$$
$$
= tr\left[\left(\mathbf{Q}^{i+1}\mathbf{R}^{i+1}(\mathbf{E}_Q^{i+1})^{-1}(\mathbf{R}^{i+1})^\top (\mathbf{Q}^{i+1})^\top\right)^+\right]
$$
$$
= tr\left[(\mathbf{R}^{i+1}(\mathbf{E}_Q^{i+1})^{-1}(\mathbf{R}^{i+1})^\top)^{-1}\right]
$$

(7)

where (7) follows since $((\mathbf{Q}^{i+1})^\top)^+$ has orthonormal columns.

Now, we prove a theorem that presents a simple criterion by which each row from $\mathbf{U}_{\mathcal{V}\mathcal{R}}$ minimizing the intermediate reconstruction error is selected at iterations.

Theorem *Let $\mathbf{r}_{i+1}$ be the $(i+1)$th column vector of $\mathbf{R}^{i+1}$ and $(\mathbf{a}_j^{i+1})^\top$ the jth row vector of $(\mathbf{R}^{i+1})^{-1}$. Then, the selection process that iteratively minimizes the intermediate reconstruction error formulated in (5) can be conducted by searching the minimizing row $\mathbf{u}^{(i+1)*}, i = 0, \ldots, r-1$ at each iteration:*

$$
\mathbf{u}^{(i+1)*} = \arg \min_{\mathbf{u}^{(i+1)}, d \neq 0} \left[\prod_{j=1}^{i+1} \| \mathbf{a}_j^{i+1} \|^2\right]^{\frac{1}{i+1}}
$$

(8)

where $(\mathbf{u}^{(i+1)})^\top$ corresponds to one of the rows of $\mathbf{U}_{\mathcal{S}_i^C \mathcal{R}}$ and d the $(i+1)$th entry of $\mathbf{r}_{i+1}$.

Proof For each $\mathbf{u}^{(i+1)}$ selected, we perform the optimal rate allocation which can be found by solving the constrained optimization problem:

$$
\begin{aligned}
\Delta^* &= \arg\min_{\Delta_j, \forall j} tr\left[(\mathbf{R}^{i+1}(\mathbf{E}_Q^{i+1})^{-1}(\mathbf{R}^{i+1})^\top)^{-1} \right], \sum_j^{i+1} R_j \le R_{T_{i+1}} \\
&= \arg\min_{\Delta_j, \forall j} \sum_{j=1}^{i+1} \frac{\Delta_j^2}{12} \parallel \mathbf{a}_j^{i+1} \parallel^2, \quad \sum_j^{i+1} R_j \le R_{T_{i+1}}
\end{aligned}
\tag{9}
$$

where (9) follows from matrix multiplication and $\Delta^* = (\Delta_1^*, \ldots, \Delta_{i+1}^*)^\top$ is a column vector of optimal step sizes that minimize the reconstruction error and derived in [17] as follows:

$$
\Delta_j^* = \frac{\left[\prod_{m=1}^{i+1} \parallel \mathbf{a}_m^{i+1} \parallel \right]^{\frac{1}{i+1}}}{\parallel \mathbf{a}_j^{i+1} \parallel} \frac{\mathcal{I}}{2^{\frac{R_{T_{i+1}}}{i+1}}}, \quad j = 1, \ldots, i+1
\tag{10}
$$

Now, we use the optimal solution Δ^* in (10) to compute the reconstruction error with the optimal rate:

$$
\begin{aligned}
&\sum_{j=1}^{i+1} \frac{(\Delta_j^*)^2}{12} \parallel \mathbf{a}_j^{i+1} \parallel^2 \\
&= \frac{2^{-2R_{T_{i+1}}/(i+1)}}{12}(i+1)\mathcal{I}^2 \left[\prod_{j=1}^{i+1} \parallel \mathbf{a}_j^{i+1} \parallel^2 \right]^{\frac{1}{i+1}} \\
&\propto \left[\prod_{j=1}^{i+1} \parallel \mathbf{a}_j^{i+1} \parallel^2 \right]^{\frac{1}{i+1}}
\end{aligned}
$$

To be concise, the selection of the next node that minimizes the intermediate reconstruction error with optimal rate allocation can be simply determined by finding one of $\mathbf{u}^{(i+1)}$'s that minimizes the geometric mean of the row vectors of $(\mathbf{R}^{i+1})^{-1}$ regardless of the total rate budget $R_{T_{i+1}}$. Furthermore, with the analytic results derived in [13], $(\mathbf{R}^{i+1})^{-1}$ is just constructed from $(\mathbf{R}^i)^{-1}$ at the previous iteration and $\mathbf{r}_{i+1}^\top \equiv [\mathbf{b}^\top \quad d]$ by appending the $(i+1)$th column vector to the last column position:

$$
(\mathbf{R}^{i+1})^{-1} = \begin{bmatrix} (\mathbf{R}^i)^{-1} & -\frac{(\mathbf{R}^i)^{-1}\mathbf{b}}{d} \\ \mathbf{0}_{1 \times i} & d^{-1} \end{bmatrix} \equiv [\tilde{\mathbf{r}}_1 \quad \cdots \quad \tilde{\mathbf{r}}_{i+1}]
\tag{11}
$$

where $\tilde{\mathbf{r}}_i$ denotes the ith column vector of $(\mathbf{R}^{i+1})^{-1}$. To guarantee the existence of the inverse of $\mathbf{R}^{i+1}$ (equivalently, the independence of $\mathbf{u}^{(i)}$'s), it should be assumed that $d \ne 0$ (see [13] for the detail). $\square$

Note that at the $(i+1)$th iteration where $(\mathbf{R}^i)^{-1}$ with its row vectors $(\mathbf{a}_j^i)^\top$ is given from the previous iteration, the geometric mean in (8) can be computed with $\parallel \mathbf{a}_j^{i+1} \parallel^2, j = 1, \ldots, i+1$ updated from $\parallel \mathbf{a}_j^i \parallel^2$ and $\tilde{\mathbf{r}}_{i+1}$ by a simple addition:

$$\parallel \mathbf{a}_j^{i+1} \parallel^2 = \parallel \mathbf{a}_j^i \parallel^2 + \tilde{\mathbf{r}}_{i+1}(j)^2, \quad j = 1, \ldots, i \tag{12a}$$

$$\tilde{\mathbf{r}}_{i+1}(i+1)^2, \quad j = i+1, \tag{12b}$$

where $\tilde{\mathbf{r}}_{i+1}(j)$ is the jth entry of the $(i+1)$th column vector of $(\mathbf{R}^{i+1})^{-1}$. Initially, we have $\mathbf{Q}^1 = \frac{\mathbf{u}^{(1)}}{\parallel\mathbf{u}^{(1)}\parallel}, \mathbf{R}^1 = \parallel \mathbf{u}^{(1)} \parallel = d$. Thus, the metric (8) is reduced to

$$\mathbf{u}^{(1)*} = \arg \min_{\mathbf{u}^{(1)}, d \neq 0} \frac{1}{d^2} \tag{13}$$

Starting with $\mathbf{u}^{(1)*}, \mathbf{Q}^1, \mathbf{R}^1$ and $\mathbf{R}^{-1} = \frac{1}{\parallel\mathbf{u}^{(1)*}\parallel}$, the criterion (8) are repeatedly evaluated from $i = 1$ to $|\mathcal{S}| - 1$ until the best sampling set S with the cardinality of r is constructed. The proposed sampling set selection algorithm is summarized as follows:

Algorithm A greedy sampling set selection algorithm

---
1. Initialize $\mathcal{S}_0 = \emptyset$.
2. select $\mathbf{u}^{(1)*}$ by using (13).
3. $\mathcal{S}_1 = \{k^*\}$ if $\mathbf{u}^{(1)*}$ is the transpose of the k^*-th row in $\mathbf{U}_{\mathcal{VR}}$.
4. For $i = 1, \ldots, r - 1$ do
   1) For each $\mathbf{u}^{(i+1)}$ selected from $\mathbf{U}_{\mathcal{S}_i^C \mathcal{R}}$
      - given $\mathbf{Q}^i$ and $\mathbf{R}^i$, compute $\mathbf{r}_{i+1}^\top \equiv [\mathbf{b}^\top \quad d]$, $\tilde{\mathbf{r}}_{i+1}$ by (11) and $\parallel \mathbf{a}_j^{i+1} \parallel^2, j = 1, ..., i+1$ by (12a) and (12b).
   2) select $\mathbf{u}^{(i+1)*}$ by using (8).
   3) $\mathcal{S}_{i+1} = \mathcal{S}_i + \{k^*\}$ if $\mathbf{u}^{(i+1)*}$ is the transpose of the k^*-th row in $\mathbf{U}_{\mathcal{S}_i^C \mathcal{R}}$.
   4) for $\mathbf{u}^{(i+1)*}$, construct the Householder matrix for QR factorization to generate $\mathbf{Q}^{i+1}$ and $\mathbf{R}^{i+1}$ and compute $(\mathbf{R}^{i+1})^{-1}$ by (11).
5. end for
6. Return $\mathcal{S} = \mathcal{S}_r$

---

4 Results and discussion

4.1 Complexity of proposed algorithm

With the eigenvector matrix of the variation operator computed as a preparation step, the proposed algorithm seeks to select the next minimizing node at the $(i + 1)$th iteration by using $(\mathbf{R}^i)^{-1}$ with its row vectors $(\mathbf{a}_j^i)^\top$. The selection process is conducted by two main tasks: for each $\mathbf{u}^{(i+1)}$, the generation of the $(i + 1)$th column vector $\mathbf{r}_{i+1}$ of $\mathbf{R}^{i+1}$ and the computation of $\tilde{\mathbf{r}}_{i+1}$ of $(\mathbf{R}^{i+1})^{-1}$ by (11) and the geometric mean of $\parallel \mathbf{a}_j^{i+1} \parallel^2$ by (12a) and (12b). This task is repeatedly executed for each of $(N - i)$ remaining rows, producing the operation count $C_{GM} \approx (N - i)(2ri + 2i^2)$ flops. After finding the next row $\mathbf{u}^{(i+1)*}$ that takes the minimum of the $(N - i)$ geometric means, the second task is performed to generate the Householder matrix for $\mathbf{u}^{(i+1)*}$ which needs $C_H \approx 2ri^2 - 4r^2i + 2r^3$ flops. Thus, at the $(i + 1)$th iteration, the operation count spent for the two main tasks is $C_{i+1} \approx 2Nri - 4r^2i$. Furthermore, this selection process is executed $(|\mathcal{S}| - 1)$ times, resulting in the total operation count of the proposed algorithm given by $O(Nr|\mathcal{S}|^2)$.

Since our algorithm is developed based on the analytic results in [13], it would be noteworthy to discuss how quantization changes the sampling set selection process by comparing with the selection algorithm without quantization proposed in [13], denoted by the QR factorization-based method (QRM). Specifically, the reconstruction error without quantization at the $(i + 1)$th iteration is given by $tr\left[(\mathbf{R}^{i+1}(\mathbf{R}^{i+1})^\top)^{-1}\right]$, the cost function for QRM and can be further manipulated as follows:
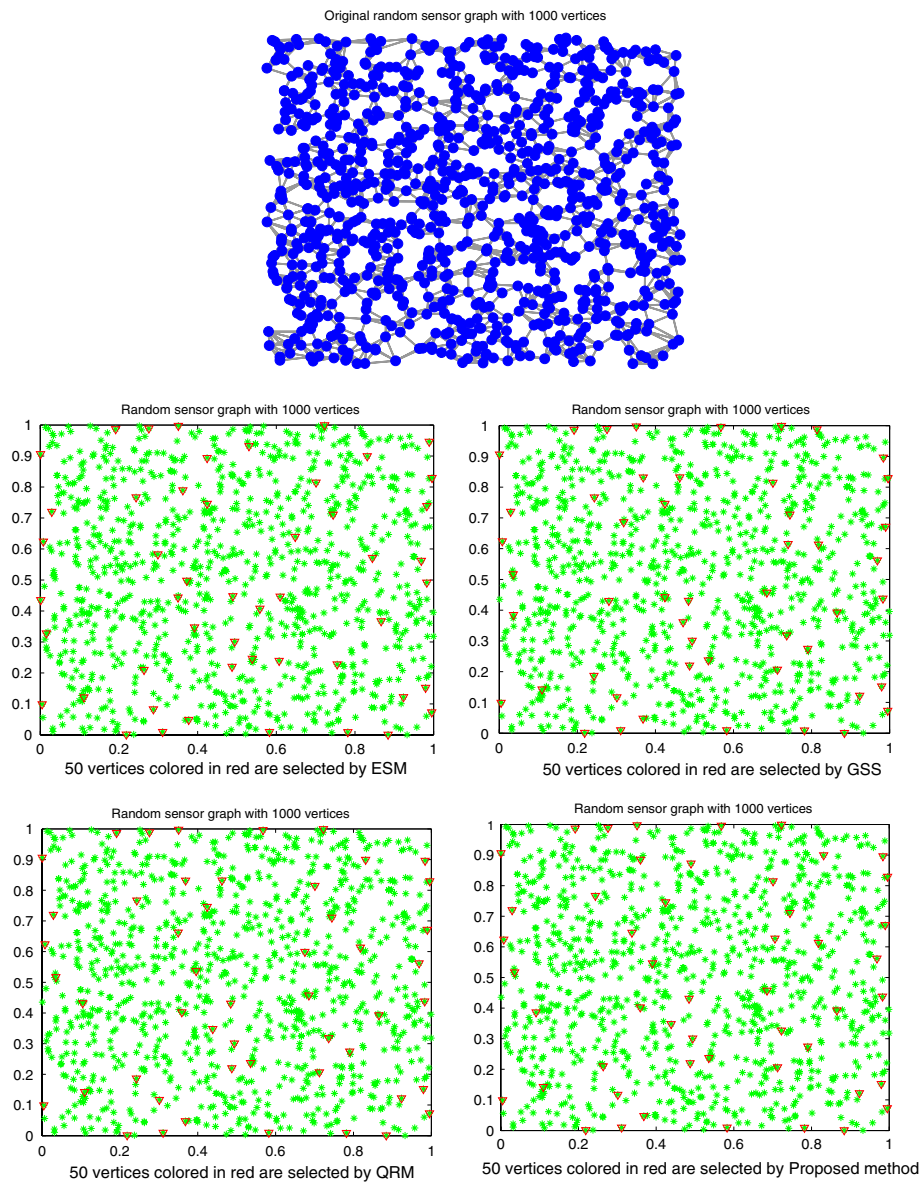
Fig. 1 Fifty vertices in red selected from RSG by using four different sampling methods

$$tr\left[(\mathbf{R}^{i+1}(\mathbf{R}^{i+1})^{\top})^{-1}\right] = \sum_{j=1}^{i+1} \parallel \mathbf{a}_j^{i+1} \parallel^2 = \sum_{j=1}^{i+1} \parallel \tilde{\mathbf{r}}_j \parallel^2$$

$$= \sum_{j=1}^{i} \parallel \tilde{\mathbf{r}}_j \parallel^2 + \parallel \tilde{\mathbf{r}}_{i+1} \parallel^2 \propto \parallel \tilde{\mathbf{r}}_{i+1} \parallel^2 \qquad (14)$$

where (14) follows since the first term $\sum_{j=1}^{i} \parallel \tilde{\mathbf{r}}_j \parallel^2$ is irrelevant in finding the next minimizing row $\mathbf{u}^{(i+1)*}$. Hence, the sampling set selection without quantization involves the arithmetic mean of $\parallel \mathbf{a}_j^{i+1} \parallel^2$ which in turn requires a computation of only $\parallel \tilde{\mathbf{r}}_{i+1} \parallel^2$. In
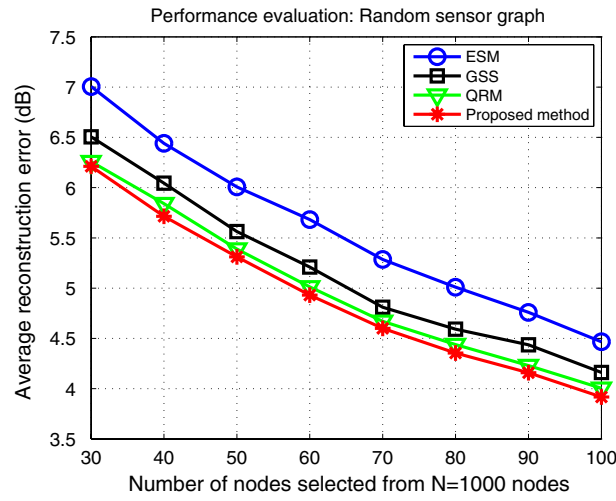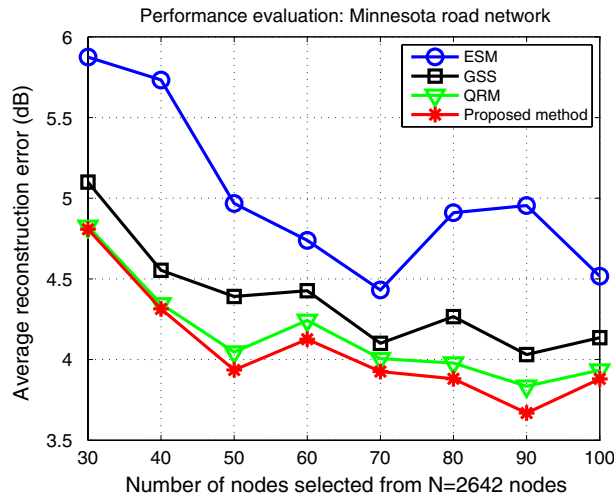
Fig. 2 Performance evaluation for noisy non-bandlimited graph signals on RSG with $\sigma = 0.1$ by varying the sample size, $|\mathcal{S}|$
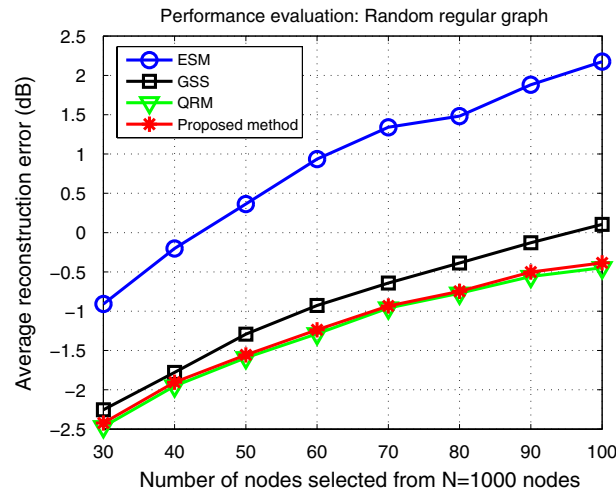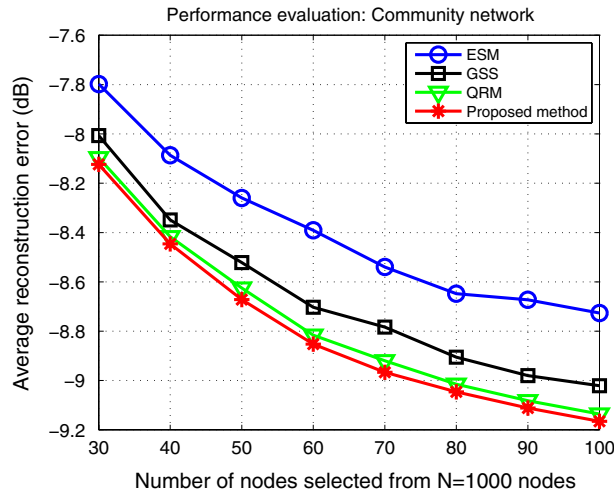

Fig. 3 Performance evaluation for noisy non-bandlimited graph signals on RRG with $\sigma = 0.1$ by varying the sample size, $|\mathcal{S}|$

contrast, the quantization-aware sampling process makes us take into account the geometric mean which needs each update of row vectors of $(\mathbf{R}^{i+1})^{-1}$. It should be noted that with the extra complexity $O(N|\mathcal{S}|^2)$ related to computation of the geometric means at iterations, the proposed algorithm offers the complexity of the same order as that of QRM.

For the performance evaluation, we compare with different sampling methods such as efficient sampling method (ESM) [6], greedy sensor selection (GSS) [14] and QRM [13]. ESM constructs sampling sets by simply performing a column-wise Gaussian elimination on $\mathbf{U}_{\mathcal{VR}}$, leading to a low weight selection process. GSS iteratively selects the next node that maximizes the log-determinant of the inverse estimation error covariance
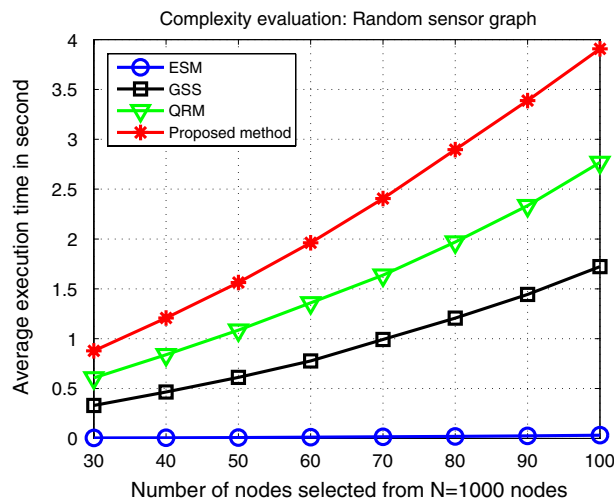
Fig. 4 Performance evaluation for noisy non-bandlimited graph signals on MRNG with $\sigma = 0.1$ by varying the sample size, $|\mathcal{S}|$
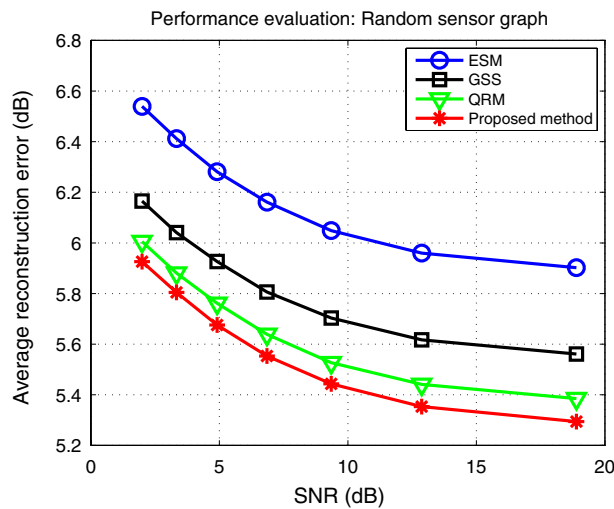

Fig. 5 Performance evaluation for noisy non-bandlimited graph signals on CNG with $\sigma = 0.1$ by varying the sample size, $|\mathcal{S}|$

matrix. Assuming zero-mean graph signals with the covariance matrix $\mathbf{\Sigma} = \mathbf{U}_{\mathcal{VR}} \mathbf{\Lambda} \mathbf{U}_{\mathcal{VR}}^{\top}$ corrupted by an iid additive measurement noise with unit variance, the inverse estimation error covariance matrix is given by $\mathbf{\Sigma}_e^{-1} = (\mathbf{\Lambda}^{-1} + \mathbf{U}_{\mathcal{SR}}^{\top} \mathbf{U}_{\mathcal{SR}})$ [12, 14]. In evaluating the metric for GSS, the high SNR is assumed (i.e., $\mathbf{\Lambda}^{-1} = \frac{1}{\sigma_f^2} \mathbf{I}$, $\frac{1}{\sigma_f^2} \ll 1$) since the metric of the proposed algorithm is formulated under the assumption of high SNR. It is noteworthy to examine the relation of the above-mentioned sampling methods with the sampling strategies based on uncertainty principle, denoted by MaxVol and MinPinv in [8] where MaxVol seeks to maximize the determinant of $\mathbf{U}_{\mathcal{SR}}^{\top} \mathbf{U}_{\mathcal{SR}}$ and MinPinv aims at minimizing the reconstruction error without quantization given by $tr[(\mathbf{U}_{\mathcal{SR}}^{\top} \mathbf{U}_{\mathcal{SR}})^{-1}]$. Clearly, GSS and QRM offer simplified greedy methods for MaxVol and MinPinv, respectively. In Table 1, the optimality criteria and the complexity are given for

**Fig. 6** Complexity evaluation in terms of average execution time in second for RSG by varying the sample size, $|\mathcal{S}|$

Fig. 7 Performance evaluation for noisy non-bandlimited graph signals on RSG with $|\mathcal{S}| = 50$ by varying SNR

comparison. As expected, ESM yields a fast selection process at the cost of performance loss since it aims to minimize the indirect metric, i.e., the worse case of the reconstruction error. It can be also seen that the proposed algorithm provides a competitive complexity in comparison with GSS and QRM, especially for $|\mathcal{S}| \leq r$. The reconstruction performance of the various methods is evaluated in the experiments in Sect. 4.2

4.2 Experimental results

We investigate the performance of various sampling set selection methods for four different graphs given below:
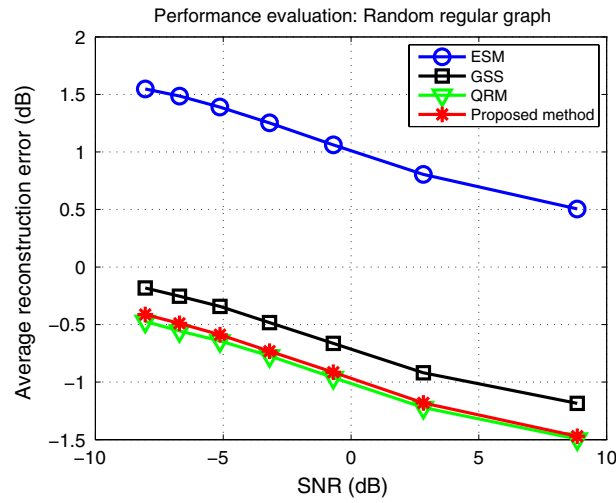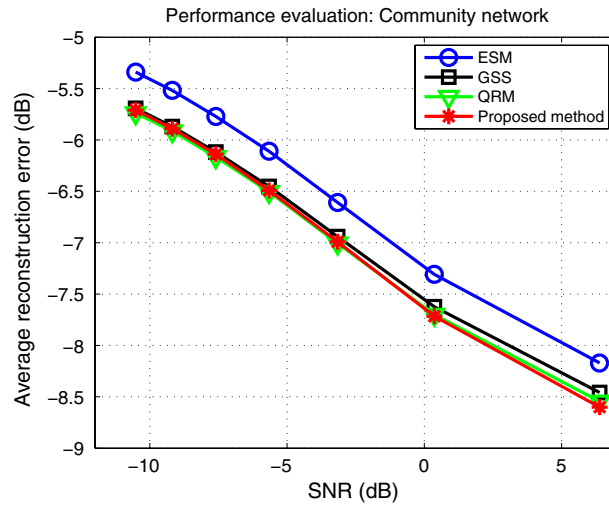
Fig. 8 Performance evaluation for noisy non-bandlimited graph signals on RRG with $|\mathcal{S}| = 50$ by varying SNR
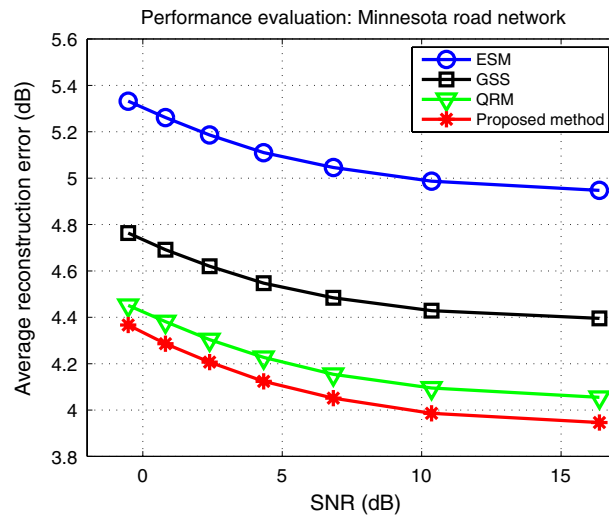
Fig. 9 Performance evaluation for noisy non-bandlimited graph signals on MRNG with $|\mathcal{S}| = 50$ by varying SNR

- Random sensor graph (RSG) with $N = 1000$ vertices
- Random regular graph (RRG) with $N = 1000$ vertices and each vertex connected to six vertices
- Minnesota road network graph (MRNG) with $N = 2642$ vertices
- Community network graph (CNG) with 16 communities of random sizes

For each graph, we generate 50 graph realizations and adopt the combinatorial Laplacian matrix $\mathbf{L}$ as a variation operator to compute $\mathbf{U}_{\mathcal{V}\mathcal{R}}$ and $\mathbf{c}_{\mathcal{R}}$ in (1). We construct uniqueness sampling sets $\mathcal{S}$ with size $|\mathcal{S}| = r$ by greedily selecting one node at each iteration based on four different techniques: ESM, GSS, QRM and the proposed method. In applying GSS, we let $\sigma_f^2 = 10^3$. For performance evaluation, we examine

Fig. 10 Performance evaluation for noisy non-bandlimited graph signals on CNG with $|\mathcal{S}| = 50$ by varying SNR

Table 1 Complexity comparison for various sampling methods

Method	Optimality criteria	Operation count
ESM [6]	$\| \mathbf{U}_{\mathcal{SR}}^{+} \|_2^2$	$O(Nr|\mathcal{S}|)$
GSS [14]	$\log \det (\Lambda^{-1} + \mathbf{U}_{\mathcal{SR}}^{\top} \mathbf{U}_{\mathcal{SR}})$	$O(Nr^2|\mathcal{S}|)$
QRM [13]	$tr[(\mathbf{U}_{\mathcal{SR}}^{\top} \mathbf{U}_{\mathcal{SR}})^{-1}]$	$O(Nr|\mathcal{S}|^2)$
Proposed method	$tr[(\mathbf{U}_{\mathcal{SR}}^{\top} \mathbf{E}_Q^{-1} \mathbf{U}_{\mathcal{SR}})^{-1}]$	$O(Nr|\mathcal{S}|^2)$

the case of noisy bandlimited and non-bandlimited signals. We consider a random graph signal $\mathbf{f}$ assumed to follow the Gaussian joint distribution:

$$p(\mathbf{f}) \propto \exp(-\mathbf{f}^{\top} \mathbf{K}^{-1} \mathbf{f}) = \exp(-\mathbf{f}^{\top} (\mathbf{L} + \delta \mathbf{I}) \mathbf{f})$$

where the covariance matrix $\mathbf{K} = (\mathbf{L} + \delta \mathbf{I})^{-1}$ with δ set to $= 0.01$ in the experiments to guarantee a proper covariance matrix. We generate the noisy signals $\mathbf{f}_S$ residing on the nodes in the sampling sets S by using an iid additive Gaussian noise drawn from $\mathcal{N}(0, \sigma^2)$ and make uniform quantization of those signal values with the approximated optimal rate $\hat{R}_i$ obtained by adjusting the optimal rate $R_i^* = \log_2 \frac{I}{\Delta_i^*}$ to its neighboring integer value such that $\sum \hat{R}_i = R_T = |\mathcal{S}|$. We compute the average reconstruction error in dB given by $10 \log_{10}(\mathbb{E} \frac{\|\mathbf{f} - \hat{\mathbf{f}}^Q\|^2}{N})$ where 1000 signal samples at each node are averaged for each of 50 graph realizations. We also provide visual demonstration of RSG and the 50 vertices in red selected from the graph by using the four methods in Fig. 1. In the experiments, we generate graphs and the attributes (i.e., $\mathbf{L}, w_{ij}, \mathbf{U}, \lambda_i$) with their default values by using the graph signal processing toolbox (GSPBox) for MATLAB [20].

4.2.1 Performance evaluation with respect to sample size

We first construct the uniqueness sampling sets $\mathcal{S}$ with $|\mathcal{S}| = r = 30$ to 100 by applying the four methods and test noisy non-bandlimited graph signals with $\sigma = 0.1$. As demonstrated in Figs. 2, 3, 4 to 5, the proposed algorithm outperforms ESM, GSS and QRM for most of the graphs. Interestingly, the proposed method and QRM produce the similar reconstruction performance for RRG since the optimal rate for RRG would be almost uniform due to its regular connectivity, implying no additional gain achieved by the optimal rate allocation. We also evaluate the complexity of the selection methods by measuring the execution time in second in the same condition. In Fig. 6, the average execution time for RSG is demonstrated with respect to the size of the sampling set $|S| = r$. Note that the proposed method spends more time than QRM since it requires the extra computation related to the geometric mean.

4.2.2 Performance evaluation with respect to noise level

We also test noisy non-bandlimited graph signals on the sampling sets $\mathcal{S}$ with $|\mathcal{S}| = 50$ by varying the SNR in dB given by $10 \log_{10} \frac{\mathbb{E}\|\mathbf{f}\|^2}{\sigma^2}$ and plot the experimental results in Figs. 7, 8, 9 to 10. As compared with the different sampling methods, the proposed method performs well for quantized graph signals in the presence of various noise levels. Notably, the experiments for noisy bandlimited signals provide better reconstruction accuracy than the case of noisy non-bandlimited signals and the performance curves for the sampling methods show a similar trend to those for noisy non-bandlimited signals.

5 Conclusions

We addressed the sampling set selection problem for quantized graph signals and presented an analytic result by which an iterative greedy selection can be simply conducted so as to minimize the average reconstruction error. Furthermore, we discussed the complexity of the proposed method, resulting in a simple extra computation needed compared with the recent algorithm in [13]. We demonstrated by experiments that our method offers a competitive performance with a reasonable complexity when graph signals are quantized.

Abbreviations

GSP	Graph signal processing
GFT	Graph Fourier transform
iid	Independent and identically distributed
SNR	Signal-to-noise ratio
ESM	Efficient sampling method
GSS	Greedy sensor selection
QRM	QR factorization-based selection method
RSG	Random sensor graph
RRG	Random regular graph
MRNG	Minnesota road network graph
CNG	Community network graph

Author contributions
The author read and approved the final manuscript.

Availability of data and materials

All experiments are described in detail within a reproducible signal processing framework. Please contact the author for data requests.

Declarations

Competing interests

The authors declare that they have no competing interests.

References

1. D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, 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. **30**(3), 83–98 (2013)
2. A. Ortega, P. Frossard, J. Kovačević, J.M.F. Moura, P. Vandergheynst, Graph signal processing: overview, challenges and applications. Proc. IEEE **106**(5), 808–828 (2018)
3. S. Joshi, S. Boyd, Sensor selection via convex optimization. IEEE Trans. Signal Process. **57**(2), 451–462 (2009)
4. M. Nareem, S. Xue, D. Lee, Cross-entropy optimization for sensor selection problems. In: International Symposium on Commuication and Information Technology (ISCIT), pp. 396–401 (2009)
5. A. Anis, A. Gadde, A. Ortega, Towards a sampling theorem for signals on arbitrary graphs. In: IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP), Florence, Italy, pp. 3864–3868 (2014)
6. A. Anis, A. Gadde, A. Ortega, Efficient sampling set selection for bandlimited graph signals using graph spectral proxies. IEEE Trans. Signal Process. **64**(14), 3775–3789 (2016)
7. S. Chen, R. Varma, A. Sandryhaila, J. Kovačević, Discrete signal processing on graphs: sampling theory. IEEE Trans. Signal Process. **63**(24), 6510–6523 (2015)
8. S.B. M. Tsitsvero, P.D. Lorenzo, Signals on graph: uncertainty principle and sampling. IEEE Trans. Signal Process. 64(18), 4845–4860 (2016)
9. A. Sakiyama, Y. Tanaka, T. Tanaka, A. Ortega, Eigendecompostion-free sampling set selection for graph signals. IEEE Trans. Signal Process. **67**(10), 2679–2692 (2019)
10. F. Wang, G. Cheung, Y. Wang, Low-complexity graph sampling with noise and signal reconstruction via Neumann series. IEEE Trans. Signal Process. **67**(21), 5511–5526 (2019)
11. S.P. Chepuri, G. Leus, Graph sampling for covariace estimation. IEEE Trans. Signal Inform. Process. Networks **3**(3), 451–466 (2017)
12. L.F.O. Chamon, A. Ribeiro, Greedy sampling of graph signals. IEEE Trans. Signal Process. **66**(1), 34–47 (2018)
13. Y.H. Kim, QR facrorization-based sampling set selection for bandlimited graph signals. Signal Process. **179**, 1–10 (2021)
14. M. Shamaiah, S. Banerjee, H. Vikalo, Greedy sensor selection: leveraging submodularity. In: 49th IEEE International Conference on Decision and Contron (CDC), pp. 2572–2577 (2010)
15. J. Ranieri, A. Chebira, M. Vetterli, Near-optimal sensor placement for linear inverse problems. IEEE Trans. Signal Process. **62**(5), 1135–1146 (2014)
16. C. Jiang, Y.C. Soh, H. Li, Sensor placement by maximal projection on minimum eigenspace for linear inverse problems. IEEE Trans. Signal Process. **64**(21), 5595–5610 (2016)
17. Y.H. Kim, A. Ortega, Toward optimal rate allocation to sampling sets for bandlimited graph signals. IEEE Signal Process. Lett. **26**(9), 1364–1368 (2019)
18. K. Sayood, *Introductin to Data Compression* (Morgan Kaufmann publishers, San Francisco, 2000)
19. G. Strang, *Linear Algebra and Its Applications*, 3rd edn. (Harcourt Brace Jovanovich College Publishers, Florida, 1988)
20. N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, D.K. Hammond, Gspbox: A toolbox for signal processing on graphs. ArXiv e-prints (2014)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.