JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 24, No. 12: 1682~1687, Dec. 2020

대역폭 제한 그래프신호를 위한 저 복잡도 샘플링 집합 선택 알고리즘

김윤학[*]

Low-complexity Sampling Set Selection for Bandlimited Graph Signals

Yoon Hak Kim[*]

[*]Associate Professor, Department of Electronic Engineering, Chosun University, Gwangju 61452 Korea

요 약

대역폭 제한 그래프신호의 신호복원을 위해서 최대의 정보를 제공하기 위한 그래프 상의 노드를 선택하는 샘플링 집합 선택 알고리즘에 대해 연구한다. 저 복잡도 선택알고리즘을 구현하기 위해 직접적인 비용함수인 신호 복원오차를 최소화 하는 대신, 신호 복원오차의 최대값을 최소화하는 방법에 대해 집중한다, 이를 위해, 추가적인 복잡도 개선을 위해 유용한 근사화공식을 적용하여 성능손실을 최소화하면서 복잡도를 개선한 저 복잡도 탐욕알고리즘을 제안한다. 다양한 그래프신호에 대한 폭넓은 실험을 통해, 기존 저 복잡도 방식과 신호복원성능 및 복잡도를 평가 비교하여 기존방식대비 신호복원 및 복잡도면에서 모두 성능 개선이 있음을 보였으며, 이는 실시간 응용분야에서 실용적인 해결방식으로써 경쟁력 있는 대안을 제시한다.

ABSTRACT

We study the problem of sampling a subset of nodes of graphs for bandlimited graph signals such that the signal values on the sampled nodes provide the most information in order to reconstruct the original graph signal. Instead of directly minimizing the reconstruction error, we focus on minimizing the upper bound of the reconstruction error to reduce the complexity of the selection process. We further simplify the upper bound by applying useful approximations to propose a low-weight greedy selection process that is iteratively conducted to find a suboptimal sampling set. Through the extensive experiments for various graphs, we inspect the performance of the proposed algorithm by comparing with different sampling set selection methods and show that the proposed technique runs fast while preserving a competitive reconstruction performance, yielding a practical solution to real-time applications.

키워드 : 그래프신호처리, 대역폭 제한 그래프신호, 샘플링집합 선택, 탐욕알고리즘, 신호복원

Keywords : Graph signal processing, Bandlimited graph signals, Sampling set selection, Greedy algorithm, Signal reconstruction

Received 28 September 2020, **Revised** 5 October 2020, **Accepted** 19 October 2020

***Corresponding Author** Yoon Hak Kim (Email:yhk@chosun.ac.kr, Tel:+82-62-230-7129)
Associate Professor, Department of Electronic Engineering, Chosun University, Gwangju 61452 Korea

Open Access **http://doi.org/10.6109/jkiice.2020.24.12.1682** print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. Introduction

Graphs signal processing (GSP) has recently received enormous research interests from signal processing society since it provides efficient tools to handle high dimensional data generated from irregularly structured network nodes in emerging applications (e.g., social, transportation, neural, energy and sensor networks). Specifically, the data samples on nodes are represented by graph signals defined on nodes of graphs and signal variation is measured by using variation operators which can be created based on connectivity of nodes of graphs [1, 2]. One crucial task in GSP is to find an optimal subset of data samples (equivalently, subset of nodes of graphs) with which the original graph signal is reconstructed to accomplish certain application objectives [3-8]. Most of the subset selection or sampling set selection methods focused on minimizing the reconstruction error or related metrics [3-6]. Greedy selection strategy has been adopted in [3-5] to minimize the worst case of the reconstruction error. A non-uniform sampling was presented so as to select more samples with higher local sparsity level [6]. To expedite selection process for practical applications, eigen decomposition-free algorithms have been developed to show a fast running time with reasonable reconstruction performance [7, 8].

In this work, we seek to find a suboptimal sampling set that can be constructed with substantially reduced complexity while preserving a competitive signal recovery. To this end, we derive a simple metric by applying useful approximations for the upper bound of the reconstruction error, leading to a low-weight selection process. We evaluate the proposed algorithm in comparison with different sampling methods by conducting experiments for various graphs and show that the proposed method operates fast while maintaining a competitive reconstruction performance.

This paper is organized as follows. The problem formulation is given in Section II and the low-weight algorithm along with the simplified metric is presented in Section III. Experimental evaluation of the proposed algorithm is provided in Section IV and the conclusion in Section V.

Ⅱ. Problem Formulation

For a graph *G(V,E)* with a set of *N* nodes *V*={1,...,*N*} and edges *E*={(i,j,w_{ij})} with weight w_{ij} of the edge between node i and j, we define an $N\times 1$ graph signal $\boldsymbol{f}=[f_1\cdots \mathrm{f_N}]^{\mathrm{T}}$ with f_i representing a data sample on the i-th vertex. Variation operators (e.g., combinatorial graph Laplacian, normalized Laplacian) can be introduced to describe the connectivity of nodes of graphs [2, 4]. We assume an $N\times N$ variation operator $\boldsymbol{L}$ with orthonormal eigenvectors $\boldsymbol{u}_1,...\boldsymbol{u}_N$ and corresponding eigenvalues $|\lambda_1|\leq\cdots\leq|\lambda_N|$. Then the graph signal $\boldsymbol{f}$ can be represented by using the eigenvector matrix $\boldsymbol{U}=[\boldsymbol{u}_1\cdots \boldsymbol{u}_N]$:

$$\boldsymbol{f}=\boldsymbol{U}\tilde{\boldsymbol{f}}=\sum_{i=1}^{N}\tilde{f}(\lambda_i)\boldsymbol{u}_i \quad (1)$$

where $\tilde{\boldsymbol{f}}=\boldsymbol{U}^{-1}\boldsymbol{f}=\boldsymbol{U}^T\boldsymbol{f}$ is the graph Fourier transform (GFT) of $\boldsymbol{f}$ with entries $\tilde{f}(\lambda_i)$. In this work, we consider the ω-bandlimited graph signal with its GFT entries $\tilde{f}(\lambda_i)=0,\lambda_i\geq\omega,\forall_{i>r}$. Specifically,

$$\boldsymbol{f}=\sum_{i=1}^{r}\tilde{f}(\lambda_i)\boldsymbol{u}_i=\boldsymbol{U}_{VR}\tilde{\boldsymbol{f}}_R \quad (2)$$

where $\tilde{\boldsymbol{f}}_R$ is an $r\times 1$ column vector with the entries of $\tilde{\boldsymbol{f}}$ indexed by R={1,...,r} and UVR consists of rows and columns of U indexed by V and columns by R, respectively.

Now, we find a subset of nodes of graphs such that the original graph signal is reconstructed from signal samples on the nodes in the sampling set *S*. Especially when the graph signal is ω-bandlimited, we can define a uniqueness sampling set *S* if the noise-free bandlimited graph signal $\boldsymbol{f}$ can be perfectly recovered from the sampled signal $\boldsymbol{f}_S$ with signal values $f_i, i\in S$ and construct the uniqueness set by choosing r independent row vectors of $\boldsymbol{U}_{VR}$: for instance, selecting the i-th row of $\boldsymbol{U}_{VR}$ is equivalent to selecting the i-th node [4].

Moreover, by using r independent rows selected from $\boldsymbol{U}_{VR}$, we can construct the $|S| \times r$ matrix $\boldsymbol{U}_{SR}$ from which we generate the sampled signal $\boldsymbol{f}_S$ by $\boldsymbol{f}_S = \boldsymbol{U}_{SR}\tilde{\boldsymbol{f}}_R$. Then the least square estimate (LSE) $\hat{\tilde{\boldsymbol{f}}}_R$ of $\tilde{\boldsymbol{f}}_R$ is computed by using the pseudoinverse $\boldsymbol{U}_{SR}^{+}$ of $\boldsymbol{U}_{SR}$: that is $\hat{\tilde{\boldsymbol{f}}}_R = \boldsymbol{U}_{SR}^{+}\boldsymbol{f}_S = (\boldsymbol{U}_{SR}^T\boldsymbol{U}_{SR})^{-1}\boldsymbol{U}_{SR}^T\boldsymbol{f}_S$. Thus, the reconstructed signal can be obtained from $\boldsymbol{f}_S$:

$$\begin{aligned}\hat{\boldsymbol{f}} &= \boldsymbol{U}_{VR}\hat{\tilde{\boldsymbol{f}}}_R = \boldsymbol{U}_{VR}\boldsymbol{U}_{SR}^{+}\boldsymbol{f}_S \\ &= \boldsymbol{U}_{VR}(\boldsymbol{U}_{SR}^T\boldsymbol{U}_{SR})^{-1}\boldsymbol{U}_{SR}^T\boldsymbol{f}_S\end{aligned} \quad (3)$$

In this work, it is assumed that the sampled signal is corrupted by an iid additive noise $\boldsymbol{n} \in \boldsymbol{R}^{|S|}$ with zero mean and unit variance. Then, the reconstruction error vector $\boldsymbol{e} = \hat{\boldsymbol{f}} - \boldsymbol{f} = \boldsymbol{U}_{VR}\boldsymbol{U}_{SR}^{+}\boldsymbol{n}$ since the noise-free bandlimited signal can be perfectly recovered from the sampled signal $\boldsymbol{f}_S$ with uniqueness set S: in other words, $\boldsymbol{f} = \boldsymbol{U}_{VR}\boldsymbol{U}_{SR}^{+}\boldsymbol{f}_S$. From this, the average reconstruction error is given by

$$\begin{aligned}E\|\boldsymbol{e}\|^2 &= tr(E\boldsymbol{e}\boldsymbol{e}^T) \\ &= tr\left(\boldsymbol{U}_{VR}\left(\boldsymbol{U}_{SR}^T\boldsymbol{U}_{SR}\right)^{-1}\boldsymbol{U}_{VR}^T\right) \quad (4) \\ &= tr\left(\left(\boldsymbol{U}_{SR}^T\boldsymbol{U}_{SR}\right)^{-1}\right) \quad (5)\end{aligned}$$

where (4) follows from $\boldsymbol{U}_{SR}^{+} = (\boldsymbol{U}_{SR}^T\boldsymbol{U}_{SR})^{-1}\boldsymbol{U}_{SR}^T$ and the assumption of $E\boldsymbol{n}\boldsymbol{n}^T = \boldsymbol{I}$ and (5) follows since $\boldsymbol{U}_{VR}$ has orthonormal columns. Thus, the best sampling set S^* that minimizes the reconstruction error in (5) can be found as follows:

$$S^* = \arg\min_{S,|S|=r} tr\left(\left(\boldsymbol{U}_{SR}^T\boldsymbol{U}_{SR}\right)^{-1}\right) \quad (6)$$

In this work, we take a greedy selection strategy and seek to select one row (equivalently, one node) at a time that minimizes the reconstruction error at each iteration given by $tr\left(\left(\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\right)^{+}\right)$, $\boldsymbol{U}_{S_iR}^T = [\boldsymbol{u}^{(1)} \cdots \boldsymbol{u}^{(i)}]$ with $\boldsymbol{u}^{(i)}$ the row of $\boldsymbol{U}_{VR}$ selected at the i-th iteration and S_i the set of nodes selected until the i-th iteration. More specifically,

$$j^* = \arg\min_{\boldsymbol{u}^{(i)} = \boldsymbol{u}_j, j \in S_i^C} tr\left(\left(\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\right)^{+}\right) \quad (7)$$

$$S_i^* = S_{i-1}^* + \{j^*\} \quad (8)$$

where $S_i^C = V - S_{i-1}^*$ and S_i^* is the best set constructed at the i-th iteration. The selection process in (7) and (8) is r-times repeated until $|S_i^*| = r$.

Ⅲ. Low-complexity sampling set selection

To simplify the selection process in (7) and (8), we aim to minimize the upper bound of the reconstruction error: denoting by $\boldsymbol{u}^{(i)^*}$ the row minimizing the reconstruction error at the i-th iteration, we have

$$\begin{aligned}\boldsymbol{u}^{(i)^*} &= \arg\min_{\boldsymbol{u}^{(i)}} tr\left(\left(\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\right)^{+}\right) \\ &= \arg\min_{\boldsymbol{u}^{(i)}} \|\boldsymbol{U}_{S_iR}^{+}\|_F^2 \quad (9) \\ &\approx \arg\min_{\boldsymbol{u}^{(i)}} \|\boldsymbol{U}_{S_iR}^{+}\|_2^2 \quad (10) \\ &= \arg\min_{\boldsymbol{u}^{(i)}} \lambda_{\max}\left(\left(\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\right)^{+}\right) \quad (11) \\ &= \arg\max_{\boldsymbol{u}^{(i)}} \lambda_{\min}\left(\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\right) \quad (12)\end{aligned}$$

where (9) follows from the definition of the Frobenius matrix norm and (10) from $\|\boldsymbol{U}_{S_iR}^{+}\|_F^2 \le i\|\boldsymbol{U}_{S_iR}^{+}\|_2^2$. To find $\lambda_{\min}$, we can search **x** with $\|\boldsymbol{x}\| = 1$ that minimizes $\boldsymbol{x}^T\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\boldsymbol{x}$: specifically,

$$\lambda_{\min}(\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}) = \min_{\boldsymbol{x},\|\boldsymbol{x}\|=1} \boldsymbol{x}^T\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\boldsymbol{x} \quad (13)$$

From this,

$$\boldsymbol{u}^{(i)^*} \approx \arg\max_{\boldsymbol{u}^{(i)}} \min_{\boldsymbol{x},\|\boldsymbol{x}\|=1} \left[\boldsymbol{x}^T\boldsymbol{U}_{S_iR}^T\boldsymbol{U}_{S_iR}\boldsymbol{x}\right] \quad (14)$$

We reduce the complexity of searching **x** by limiting the searching space to $\left\{\bar{u}_k \equiv \frac{u_k}{\| u_k \|} : k \in S_i^C\right\}$ and rewrite (14) by rank-one decomposition of $U_{S_iR}^T U_{S_iR}$:

$$\min_{x, \|x\|=1} \left[x^T U_{S_iR}^T U_{S_iR} x\right]$$

$$= \min_{x, \|x\|=1} \left[x^T U_{S_{i-1}R}^T U_{S_{i-1}R} x + x^T u^{(i)} u^{(i)^T} x\right] \quad (15)$$

$$\approx \min_{\bar{u}_k, k \in S_i^C} \left[\bar{u}_k^T U_{S_{i-1}R}^T U_{S_{i-1}R} \bar{u}_k + \bar{u}_k^T u^{(i)} u^{(i)^T} \bar{u}_k\right] \quad (16)$$

$$\approx \bar{u}^{(i)*T} U_{S_{i-1}R}^T U_{S_{i-1}R} \bar{u}^{(i)^*} + \bar{u}^{(i)*T} u^{(i)} u^{(i)^T} \bar{u}^{(i)^*} \quad (17)$$

where $\bar{u}^{(i)^*}$ is given by

$$\bar{u}^{(i)^*} = \arg\min_{\bar{u}_k, k \in S_i^C} \left[\bar{u}_k^T U_{S_{i-1}R}^T U_{S_{i-1}R} \bar{u}_k\right] \quad (18)$$

Since the first tem of (17) is irrelevant in finding $u^{(i)^*}$, we have

$$u^{(i)^*} \approx \arg\max_{u^{(i)} = u_k, k \in S_i^C} \left|\left\langle \bar{u}^{(i)^*}, u^{(i)} \right\rangle\right|^2 \quad (19)$$

$$\approx \arg\max_{u^{(i)} = u_k, k \in S_i^C} \| \bar{u}^{(i)^*} \|^2 \| u^{(i)} \|^2 \quad (20)$$

$$= \arg\max_{u^{(i)} = u_k, k \in S_i^C} \| u^{(i)} \|^2 \quad (21)$$

where (20) follows from the Cauchy–Schwarz inequality $\left|\left\langle \bar{u}^{(i)^*}, u^{(i)} \right\rangle\right|^2 \leq \| \bar{u}^{(i)^*} \|^2 \| u^{(i)} \|^2$ and (21) from $\| \bar{u}^{(i)^*} \| = 1$.

Obviously, the proposed algorithm in (21) enables us to find $u^{(i)^*}$ without the computation of $\bar{u}^{(i)^*}$ in (18), providing a low weight selection process which is evaluated for various graphs in terms of reconstruction performance and complexity by comparing with other methods in the following section.

Ⅳ. Simulation results

In the experiments, we examine how various selection methods perform for three graphs such as random sensor graph (RSG), random regular graph (RRG) with each vertex connected to six vertices and random Erdös-Rényi graph (RERG) with the edge connecting probability p=0.05. We use the graph signal processing toolbox (GSPBox) for Matlab [9] to generate graph realizations with N=1000, the combinatorial Laplacian as a variation operator, and its eigenvectors and eigenvalues. To construct sampling sets S with size $|S|$ from 30 to 100, we use three different techniques, denoted by random sampling method (RSM) which selected randomly $|S|$ nodes from N nodes, the efficient sampling method (ESM) [4] in which the nodes are selected by a column-wise Gaussian elimination of U_{VR} and the proposed method given by (21), respectively. To generate signals residing on the nodes of graphs, we consider random graph signals assumed to follow the Gaussian joint distribution:

$$p(f) \propto \exp\left(-f^T K^{-1} f\right) = \exp\left(-f^T (L + \delta I) f\right) \quad (22)$$

where $(L + \delta I)^{-1}$ plays a role of the covariance matrix K and δ takes a small value (=0.01 in this experiment) to ensure non-singularity of $(L + \delta I)$. Since graph signals are practically non-bandlimited and noise-corrupted, we test the three sampling methods for non-bandlimited graph signals in the presence of an iid additive noise with $N(0, \sigma^2)$.

For each method, we compute the average reconstruction error $E\frac{\| f - \hat{f} \|^2}{N}$ in which the average is taken over 100 graph signal values at each node for each of 50 graph realizations. Note that for the case of non-bandlimited signals, the reconstructed signal can be similarly obtained by $\hat{f} = U U_{SV}^{+} f_S$ (see (3) for the case of bandlimited signals) where U_{SV}^{+} is the pseudoinverse of the matrix U_{SV} with rows of U indexed by S. In fig. 1, the reconstruction performance curves with respect to sample size are plotted to show its superior reconstruction performance to RSM and ESM. Note that for the case of RRG, the algorithms optimized to reduce the reconstruction error yield the performance similar to the non-optimized one (e.g., RSM) because of the regular structure of the graph. Furthermore, the running time in second for the sampling methods is tabulated in Table 1. As expected, the proposed algorithm runs over three times faster than ESM for the graphs tested in the experiments, offering a practical solution to the sampling set selection problem for real-time network applications.

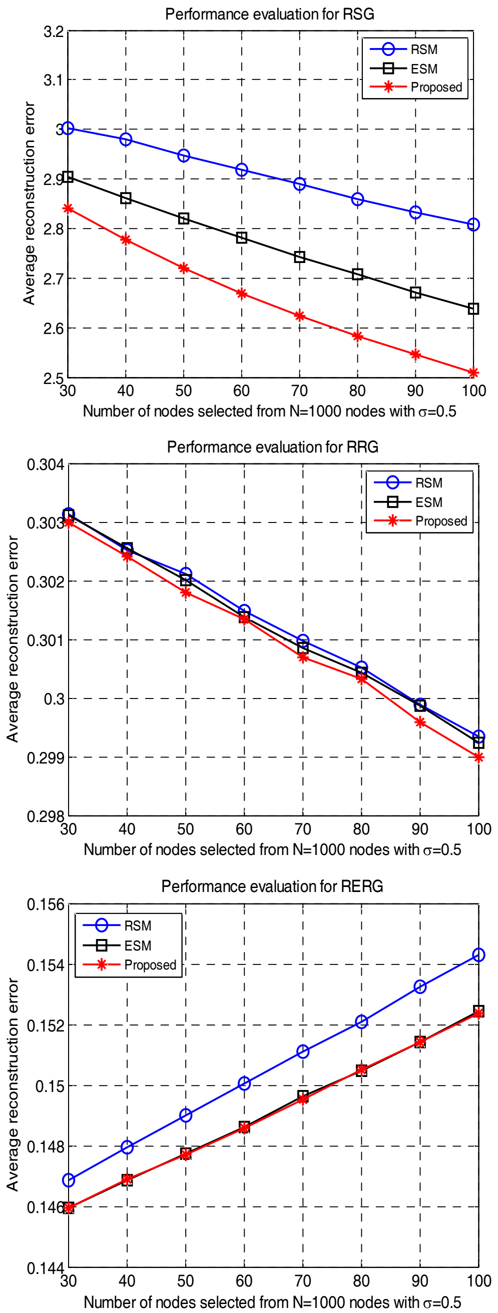

Fig. 1 Performance evaluation of different sampling methods for various graphs by varying sample size with signal noise level $\sigma = 0.5$.

Table. 1 Comparison of running time: the ratio of average running time in second is provided for various graphs: RT(ESM) and RT(Prop) denote the average running time for ESM and Proposed method, respectively.

	RSG	RRG	RERG
$\frac{RT(ESM)}{RT(\mathrm{Prop})}$	3.52	3.49	3.38

Ⅴ. Conclusion

We proposed a low-complexity sampling set selection algorithm based on a simplified metric that approximates the worst case of the reconstruction error. We inspected the performance of the proposed algorithm in comparison with previous different methods for noisy random graph signal, showing that the proposed method performs well in a practical aspect.

References

[1] D. Shuman, S. 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 Processing Magazine*, vol. 30, no. 3, pp. 83-98, 2013.

[2] A. Ortega, P. Frossard, J. Kovaevic, J. M. F. Moura, P. Vandergheynst, “Graph signal processing: overview, challenges and applications,” *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808-828, 2018.

[3] A. Anis, A. Gadde, A. Ortega, “Towards a sampling theorem for signals on arbitrary graphs,” *IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, Florence, Italy, pp. 3864-3858, 2014.

[4] A. Anis, A. Gadde, A. Ortega, “Efficient sampling set selection for bandlimited graph signals using graph spectral proxies,” *IEEE Transactions on Signal Processing*, vol. 64,

no. 14, pp. 3775-3789, 2016.

[5] S. Chen, R. Varma, A. Sandryhaila, J. Kovaevic, "Discrete signal processing on graphs: sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp.6510-6523, 2015.

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

[7] A. Sakiyama, Y. Tanaka, T. Tanaka, A. Ortega, "Eigendecompostion-free sampling set selection for graph signals," *IEEE Transactions on Signal Processing*, vol.67, no. 10, pp. 2679 – 2692, 2019.

[8] F. Wang, G. Cheung, Y. Wang, "Low-complexity graph sampling with noise and signal reconstruction via Neumann series," *IEEE Transactions. On Signal Processing*, vol. 67, no. 21, pp. 5511 – 5526, 2019.

[9] N. Perraudin, J. Paratte, D. Shuman, L. Martin, V. Kalofolias, P. Vandergheynst, and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," ArXiv e-prints arXiv:1408.5781, Aug. 2014.

김윤학(Yoon Hak Kim)

1992: BS degree in Electronic Engineering, Yonsei University
1994: MS degree in Electronic Engineering, Yonsei University
2007: Ph.D. degree in Electrical Engineering, University of Southern California
2012 - Present: Associate professor, Chosun University