

GREEDY ALGORITHM WITH APPROXIMATION RATIO FOR SAMPLING NOISY GRAPH SIGNALS

Changlong Wu Wenxin Chen June Zhang

University of Hawai'i at Mānoa
Department of Electrical Engineering
Honolulu, HI, USA

ABSTRACT

We study the optimal sampling set selection problem in sampling a noisy k -bandlimited graph signal. To minimize the effect of noise when trying to reconstruct a k -bandlimited graph signal from m samples, the optimal sampling set selection problem has been shown to be equivalent to finding a $m \times k$ submatrix with the maximum smallest singular value, σ_{\min} [3]. As the problem is NP-hard, we present a greedy algorithm inspired by a similar submatrix selection problem known in computer science and to which we add a local search refinement. We show that 1) in experiments, our algorithm finds a submatrix with larger σ_{\min} than prior greedy algorithm [3], and 2) has a proven worst-case approximation ratio of $1/(1 + \epsilon)k$, where ϵ is a constant.

Index Terms— Graph Signal Processing, Greedy Algorithm, Noisy Graph Signal, Graph Signal Sampling, Approximation Ratio

1. INTRODUCTION

Graph signal processing extends classical signal processing in time to signals over an arbitrary graph structure G ; concepts such as Fourier transformation and bandlimitedness have been studied in the general domain of graph signal [1, 2, 3]. Since the graph of interest is often very large (e.g. Facebook network), it is useful to consider sampling the graph signal (i.e., look at a subset of nodes on the graph). Reference [3] showed that for noiseless k -bandlimited graph signal sampling, we can choose to sample the graph signal such that perfect recovery is possible. However, the problem becomes more complicated in the presence of noise. Reference [3] showed that optimally deciding which $m < n$ nodes to sample for noisy graph signal is equivalent to a m -row submatrix selection problem with the goal of maximizing the smallest singular value, σ_{\min} , of the submatrix.

First, we note that in graph signal processing literatures, the problem of sampling noisy graph signal is considered to be NP-hard [3, 4] but has not been formally proven. Second, we note that the problem of optimally choosing a submatrix is similar but not identical to two existing problems studied in

computer science: 1) *matrix column subset selection problem* and 2) *matrix low rank approximation problem*. The matrix column subset selection problem, studied in [5], is the problem of optimally choosing columns from a fat matrix. Matrix low rank approximation, a well-studied problem [6, 7, 8, 9], is the problem choosing best low rank (i.e., subset of rows) approximation of the original matrix.

Inspired by these problems, we study the algorithmic performance of sampling noisy graph signal. First, we prove that the maximum least singular value, σ_{\min} , submatrix selection problem is NP-hard. Second, we give a novel greedy algorithm for optimally sampling noisy graph signal. We show with experiments that this algorithm results in submatrices with larger σ_{\min} than the greedy approach given in [3]. Further, the greedy algorithm we present has an approximation ratio of $1/(1 + \epsilon)k$, where $\epsilon > 0$ is arbitrary constant; the approximation ratio allows us to characterize the ratio of σ_{\min} of the submatrix resulting from our algorithm to the maximum possible σ_{\min} .

2. BACKGROUND AND PREVIOUS WORKS

A graph signal, $s \in \mathbb{R}^n$, consists of both a n -node graph G and a n -length vector denoting the values on each node. The structure of G is described by a graph operator, which is usually either the adjacency matrix or the Laplacian matrix. Assuming that the graph operator is diagonalizable with eigenvector matrix $V = [v_1, v_2, \dots, v_n]$, a graph signal s is k -bandlimited if $s \in \text{Span}\{v_1, \dots, v_k\}$ such that $k < n$ [3].

The problem of sampling (and recovery) of graph signals is to find a subset $I \subset \{1, 2, \dots, n\}$ with $|I| = m \geq k$ (usually $m = k$) such that one can recover the original graph signal $s \in \mathbb{R}^n$ from the sampled signal s_I , where $s_I \in \mathbb{R}^m$ is the m -dimensional vector projected from s to the indices in I . Perfect recovery was shown to be possible when the sampled signal, s_I , is noiseless [3].

However, if the sampled signal s_I is corrupted by some noise, e , so that the observed sampled signal is $s_I + e$, then perfect recovery is no longer possible. In this case, one hopes to recover the signal as close as possible to the original signal

via an estimation function $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ such that the worst case error

$$\max_{e: \|e\|_2 \leq \epsilon, s} \|f(s_I + e) - s\|_2$$

is minimized. In [3], the authors used the following heuristic estimation function

$$f(s_I + e) = VV_I^{-1}(s_I + e),$$

where V_I is a $m \times k$ matrix with rows from V indexed by I (assuming that $m = k$, then V_I is a square matrix). As shown in [3] the worst case error is

$$\sigma_{\max}(V_I^{-1}) = 1/\sigma_{\min}(V_I).$$

Moreover, one can show that this is optimal for worst case error and independent of the choice of orthogonal bases of V . Thus, the optimal sampling set selection problem is reduced to finding

$$I^* = \arg \max_{I \subset \{1, 2, \dots, n\}, |I|=m} \sigma_{\min}(V_I). \quad (1)$$

Assuming that the graph signal $s \in \mathbb{R}^n$ is k -bandlimited, consider the matrix $[v_1, v_2, \dots, v_k] \in \mathbb{R}^{n \times k}$. Our problem is to choose m rows from $[v_1, v_2, \dots, v_k]$ such that the result $m \times k$ matrix (usually $m = k$), V_I , satisfies (1). In [3], the authors provided a greedy algorithm to maximize the least singular value but does not provide any bounds on the approximation ratio. Reference [4] provided a similar greedy algorithms as in [3] that try to maximize the Frobenius norm and volume of V_I^{-1} . The average error version of the optimal sampling set selection problem was studied in [10], where the authors established approximation bounds on mean square errors for their greedy algorithms. However, this analysis only holds for Gaussian noise while the approximation ratio we present here holds for any noise type.

3. GREEDY ALGORITHM FOR SAMPLING NOISY GRAPH SIGNAL

First, we formally prove that solving (1) is NP-hard. Our proof is inspired by the approach in [5] for the matrix column subset selection problem.

Theorem 1. *To find a submatrix $C \in \mathbb{R}^{k \times k}$ from matrix $A \in \mathbb{R}^{n \times k}$ by selecting k rows with $\sigma_{\min}(C)$ maximal is NP hard.*

Proof Sketch. The reduction is from EXACT-3-COVER(X3C) problem: let $X = \{x_1, \dots, x_k\}$ and $C = \{c_1, \dots, c_n\}$ where $c_i \subset X$ with $|c_i| = 3$, we want to find $k/3$ disjoint elements from C such that their union is X . We follow the reduction as in [5], with the differences that for each c_i we construct 3 orthonormal vectors v_i^1, v_i^2, v_i^3 rather than only one vector. With the property that $v_i^r \perp v_j^t$ for all r, t and $i \neq j$ if $c_i \cap c_j \neq \emptyset$, and $v_i^r \perp v_j^t$ if $c_i \cap c_j = \emptyset$. Note that, this implies that the X3C problem has a solution iff one can find k orthonormal vectors from $\{v_i^t\}$. Detailed construction will be given in longer manuscript. \square

Theorem 1 shows that solving (1) is NP hard. Thus one can not hope to find a deterministic polynomial algorithm for this problem unless $P = NP$. In this section, we give a greedy algorithm with approximation ratio $1/(1 + \epsilon)k$ when $m = k$ (i.e., the number of nodes to sample is equal to the bandwidth of the graph signal). First, some definitions:

Definition 1 (ℓ_1 analog of σ_{\min}). *Consider an arbitrary matrix $C \in \mathbb{R}^{m \times k}$. Let $\alpha(C)$ be the ℓ_1 analog of $\sigma_{\min}(C)$. This means that*

$$\alpha(C) = \min_{x \in \mathbb{R}^k, \|x\|_2=1} \sum_{i=1}^m |c_i^T x|,$$

where c_i^T is the i th row of C .

Lemma 1. *We have*

$$\alpha(C)/\sqrt{m} \leq \sigma_{\min}(C) \leq \alpha(C).$$

Proof. The first inequality comes by Cauchy-Schwarz inequality and the second inequality is by simple algebra. \square

Definition 2 (local $(1 + \epsilon)$ -maximum volume, from [12]). *Consider an arbitrary matrix $A \in \mathbb{R}^{n \times k}$ and a submatrix $C \in \mathbb{R}^{k \times k}$ that is formed by choosing k rows from A . We say C is a local $(1 + \epsilon)$ -maximum volume submatrix, if*

$$(1 + \epsilon)\text{Vol}(C) \geq \text{Vol}(C') = \prod_{i=1}^k \sigma_i(C'),$$

whenever C' is the matrix that replace one row of C by a row in A that is not in C .

Our algorithm for solving (1) and therefore the noisy graph signal sampling problem consists of two phase: **Algorithm 1:** Initialization and **Algorithm 2:** Local Search. We assume that $m = k$ (i.e., the number of nodes to sample is equal to the bandwidth of the graph signal).

Algorithm 1 tries to greedily identify the set I such that the submatrix V_I has the maximum volume of all the possible submatrices in $[v_1, v_2, \dots, v_k] \in \mathbb{R}^{n \times k}$; algorithm 1 was used [5] for approximating maximum volume column submatrix.

Algorithm 1 Initialization

Input: Matrix $A \in \mathbb{R}^{n \times k}$

Output: Matrix $C \in \mathbb{R}^{k \times k}$, such that the rows of C are selected from the rows of A

1. Let $C \leftarrow \phi$ the empty matrix
2. Find a row $a_i^T \in A$ and $a_i^T \notin C$ that maximizes $\text{Vol}(C')$ where

$$C' = \begin{bmatrix} C \\ a_i^T \end{bmatrix}.$$

Set $C \leftarrow C'$.

3. Repeat step 2 until C has k rows. Output C
-

The algorithm 1 will output a submatrix with good volume approximation. However, it is not guaranteed to have a good smallest singular value approximation. We can further improve the output with a local search by looking for an output that has local $(1+\epsilon)$ -maximum volume (see Definition 2). As we will show with experiment in Section 4 and Section 5, algorithm 2 will result in a new output with larger σ_{\min} .

Algorithm 2 Local Search

Input: Matrix $A \in \mathbb{R}^{n \times k}$ and parameter ϵ

Output: Matrix $C \in \mathbb{R}^{k \times k}$, such that the rows of C are in the rows of A

1. Let $C \leftarrow \text{Initialization}(A)$
 2. If there is a row $a_i^T \in A$ and $a_i^T \notin C$, such that $(1 + \epsilon)\text{Vol}(C) \leq \text{Vol}(C')$, then set $C \leftarrow C'$, where C' is the matrix that replace one row in C by a_i^T .
 3. Repeat step 2 and 3 until there is no update. Output C .
-

The output of algorithm 2 is $C = V_I$. A good sampling set I can be identified from the output of algorithm 2 by keeping track of which rows was selected from the input $[v_1, v_2, \dots, v_k] \in \mathbb{R}^{n \times k}$. We give a detailed theoretic analysis to the running time and approximation ratio of this algorithm in the next section.

4. ANALYSIS OF THE ALGORITHM

In this section, we show that the submatrix output of algorithm 2 give a good approximation ratio. To best select the $m = k$ nodes to sample in the presence of noise, we need to solve (1). Suppose that there is some optimal sampling set I_{opt} that solves (1) corresponding to the optimal submatrix C_{opt} ; this means that C_{opt} has the largest σ_{\min} of all the submatrices of $[v_1, v_2, \dots, v_k]$ with dimension $k \times k$.

Theorem 2. *Let A , ϵ and C be the input and output of algorithm 2: local search. We have*

$$\alpha(C) \geq \frac{\sigma_{\min}(C_{\text{opt}})}{(1 + \epsilon)\sqrt{k}}, \quad (2)$$

where $\alpha(C)$ is the ℓ_1 analog of $\sigma_{\min}(C)$ as defined in Definition 1. The algorithm runs in

$$O(k^5(n - k) \log_{1+\epsilon} k). \quad (3)$$

We delay the proof of this theorem later. Assuming that Theorem 2 is correct, a natural corollary follows:

Corollary 1. *There exist a polynomial time algorithm that finds a submatrix $C \in \mathbb{R}^{k \times k}$ from $A \in \mathbb{R}^{n \times k}$ by selecting k rows, such that*

$$\sigma_{\min}(C) \geq \frac{\sigma_{\min}(C_{\text{opt}})}{(1 + \epsilon)k},$$

where C_{opt} is the optimal submatrix with the maximum σ_{\min} .

Proof. By Lemma 1, we have $\sigma_{\min}(C) \geq \frac{\alpha(C)}{\sqrt{k}}$. Together with (2) in Theorem 2, one will finish the proof. \square

Corollary 1 shows that the approximation ratio between the smallest singular value of the output of algorithm 2, $\sigma_{\min}(C)$, and $\sigma_{\min}(C_{\text{opt}})$ is lowerbounded by $\frac{1}{(1+\epsilon)k}$, where ϵ is some (small) precision parameter and k is the bandwidth of the graph signal. To prove Theorem 2, we first need the following lemma:

Lemma 2 (Key lemma). *Let $C \in \mathbb{R}^{k \times k}$ be a matrix, and c_i^T be the i th row of C . Then*

$$\alpha(C) = \min_{x \in \mathbb{R}^k, \|x\|_2=1} \sum_{i=1}^k |c_i^T x| = \min_j d(c_j^T, H_{-j}), \quad (4)$$

where H_{-j} is the hyperplane spanned by $\{c_i^T\}_{i \neq j}$ and $d(c, H)$ is the Euclidean distance from point $c \in \mathbb{R}^k$ to hyperplane H .

Proof. Consider x^* such that

$$x^* = \arg \min_{x \in \mathbb{R}^k, \|x\|_2=1} \sum_{i=1}^k |c_i^T x|. \quad (5)$$

We only need to prove that $x^* \perp H_{-j}$ for some j . We do so by induction on the dimension k . Consider the base case of $k = 2$, then C consists of 2 rows: c_1^T and c_2^T . There must be two lines that are perpendicular to c_1^T and c_2^T , respectively. The intersection of these lines (assuming c_1^T and c_2^T are not parallel) partition the \mathbb{R}^2 plane into 4 convex cones. The point x^* must on the boundary of one of these cones. To see this, assume that x^* lies in the interior of one of the convex cone, say B . By the construction of B , we have there exist $s_1, s_2 \in \{-1, 1\}$, such that

$$\forall y \in B, |c_1^T y| + |c_2^T y| = (s_1 c_1 + s_2 c_2)^T y.$$

Denote $c' = s_1 c_1 + s_2 c_2$. We know that there must be some small Δx such that $c'^T \Delta x < 0$, $y = x + \Delta x \in B$ and $\|y\|_2 = 1$, as x is in the interior of B . However, we will now have $c'^T y < c'^T x^*$, which contradicts to (5), this proves the base case.

For general dimension k , we know that \mathbb{R}^k will be partitioning into 2^k convex cones by the hyperplanes that is perpendicular to c_i^T for $i = 1, \dots, k$. By the same argument as for $k = 2$, we know that x^* must be in hyperplane H_{-i} that is perpendicular to c_i for some i . We now project the c_j^T for $j \neq i$ to H_{-i} with projection image $c_j'^T$. Note that, we have $c_j^T x^* = c_j'^T x^*$. We now reduced the problem to the problem of finding $x^* \in H_{-i}$ with $\|x^*\|_2 = 1$ such that $\sum_{j \neq i} |c_j'^T x^*|$ is minimal. By induction hypothesis on $k - 1$, we know that there exist a $t \neq i$, such that $x^* \perp c_j'^T$ thus $x^* \perp c_j^T$ for all $j \neq i, t$. Note that x^* is also perpendicular to c_i , we know that $x^* \perp H_{-t}$, the lemma now follows. \square

Proof of theorem 2. By lemma 2 we know that $\alpha(C)$ is exactly the minimum distance from c_j^T to the hyperplane H_{-j} , which is spanned by $\{c_i\}_{i \neq j}$. Let x be the unit vector that is perpendicular to H_{-j} . We know that $\alpha(C) = |c_j^T x|$. By definition of local $(1 + \epsilon)$ -maximum volume (Definition 2), we know that for any row $r^T \in C_{opt}$ we have $|r^T x| \leq (1 + \epsilon)|c_j^T x| = (1 + \epsilon)\alpha(C)$ (otherwise one can replace the row c_j^T in C with r^T that violate the definition of local $(1 + \epsilon)$ -maximum volume). Therefore $\sigma_{\min}(C_{opt}) \leq \sqrt{\sum_{r \in C_{opt}} |r^T x|^2} \leq \sqrt{k}(1 + \epsilon)\alpha(C)$ (i.e. $\alpha(C) \geq \frac{1}{\sqrt{k}(1 + \epsilon)}\sigma_{\min}(C_{opt})$). This proves equation (2) of Theorem 2.

Now, we have to analyze the running time of the algorithm. The only unclear thing is the number of iterations. Let $C^* \in \mathbb{R}^{k \times k}$ be the submatrix of A with $\text{Vol}(C^*)$ maximum. Using the result in [5], we have

$$\text{Vol}(C) \geq \frac{\text{Vol}(C^*)}{k!}, \quad (6)$$

where C is the output of algorithm 1. Note that, after each iteration in algorithm 2, the $\text{Vol}(C)$ will increase by a factor of $(1 + \epsilon)$. Using (6) we know that there are at most

$$\log_{1+\epsilon} k! = k \log_{1+\epsilon} k,$$

iterations. This proves equation (3) of Theorem 2. \square

We have shown that our local search algorithm will give the approximation ratio on the order of $\frac{1}{k}$. This seems to be loose. However, we have intuition that this is the best we can get for any naive greedy approach to solve (1).

Theorem 3. For any $\delta > 0$, we can find an instance of the least singular value subset selection problem, such that

$$\alpha(C) \leq \frac{\alpha(C_{opt})}{k} - \delta,$$

where C is the output of algorithm 2 and C_{opt} is the optimal submatrix that maximize α .

The proof of theorem 3 is by explicit construction. Details will be given in longer manuscripts. Moreover, this construction also gives a upper bound on $\sigma_{\min}(C)$.

5. EXPERIMENT

In this section, we provide empirical comparisons to optimally sampling noisy graph signal. We consider three algorithms: 1) **Greedy without Refinement**: algorithm 1 in this paper, which appeared previously in [5], 2) **Greedy**: the greedy algorithm introduced in [3], and 3) **Greedy with Refinement** ($\epsilon = 0.001$): algorithm 1+2 that we introduced in this paper (where we only use the algorithms to identify the first k position then greedily choosing the following positions

as in [3], when $m > k$). We consider a k -bandlimited graph signal where $k = 10$ (corresponding the 10 largest eigenvector of the adjacency matrix) on an Erdős-Rényi graph with 50 nodes. The x -axis of Figure 1 shows increasing sample sizes m from $m = k = 10$ to $m = 50$. The y -axis shows minimum singular value, σ_{\min} , of the different algorithms; in the context of graph signal processing, the larger the minimum singular value, the smaller the worst case error of reconstruction from noisy graph signal samples.

We see that both **Greedy without Refinement** (algorithm 1) and **Greedy with Refinement** (algorithm 1+2) result in submatrices with larger σ_{\min} than the method used in [3]. Our inclusion of local search in algorithm 2 results in further improvement on algorithm 1.

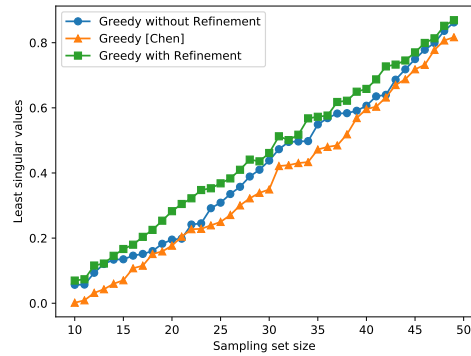


Fig. 1. Comparison of σ_{\min} for different algorithms

6. DISCUSSION

Optimally sampling a k -bandlimited noisy graph signal is equivalent of finding an $m \times k$ submatrix from $[v_1, \dots, v_k]$ with maximum least singular value. Similar type of problem is well-studied in computer science. We show a greedy search algorithm that modifies an existing algorithm for approximating maximum volume column submatrix in [5] with an added local search refinement. Experimentally, our algorithm performs better than an existing greedy algorithm for sampling noisy graph signal. Furthermore, we were able to derive an approximation ratio on the smallest singular value for our algorithm when $m = k$ by leveraging the geometry structure of σ_{\min} . We will try to derive similar type of approximation bounds for greedy sampling when $m > k$ in our future work.

7. REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Sig-*

nal Processing Magazine, vol. 30, no. 3, pp. 83–98, 2013.

- [2] A. Sandryhaila and J. M. Moura, “Discrete signal processing on graphs,” *IEEE transactions on signal processing*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [3] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, “Discrete signal processing on graphs: Sampling theory,” *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.
- [4] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, “Signals on graphs: Uncertainty principle and sampling,” *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.
- [5] A. Çivril and M. Magdon-Ismail, “On selecting a maximum volume sub-matrix of a matrix and related problems,” *Theoretical Computer Science*, vol. 410, no. 47–49, pp. 4801–4811, 2009.
- [6] A. Deshpande and L. Rademacher, “Efficient volume sampling for row/column subset selection,” in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. IEEE, 2010, pp. 329–338.
- [7] H. Avron and C. Boutsidis, “Faster subset selection for matrices and applications,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 4, pp. 1464–1499, 2013.
- [8] A. Deshpande, L. Rademacher, S. Vempala, and G. Wang, “Matrix approximation and projective clustering via volume sampling,” in *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*. Society for Industrial and Applied Mathematics, 2006, pp. 1117–1126.
- [9] Y. Wang and A. Singh, “An empirical comparison of sampling techniques for matrix column subset selection,” in *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*. IEEE, 2015, pp. 1069–1074.
- [10] L. F. Chamon and A. Ribeiro, “Greedy sampling of graph signals,” *arXiv preprint arXiv:1704.01223*, 2017.
- [11] D. S. Johnson, J. K. Lenstra, and A. Kan, “The complexity of the network design problem,” *Networks*, vol. 8, no. 4, pp. 279–285, 1978.
- [12] C.-T. Pan, “On the existence and computation of rank-revealing lu factorizations,” *Linear Algebra and its Applications*, vol. 316, no. 1–3, pp. 199–222, 2000.