Greedy Algorithm With Approximation Ratio For Sampling Noisy Graph Signals

Changlong Wu, Wenxin Chen, June Zhang {wuchangl, wenxinc, zjz}@hawaii.edu
Department of Electrical Engineering, University of Hawai'i at Mānoa

Motivation

Consider a graph $G = (V, E)$ (e.g. a social networks) with n (large) vertices, the edges of G represents that the connected vertices have similar "attributes". A graph signal is a map $\mathbf{s} : V \to \mathbb{R}$, which gives each vertex a real number (e.g. the preferences of people in a social networks). Assuming s is smooth, meaning that $\mathbf{s}[v_1] \sim \mathbf{s}[v_2]$ if $(v_1, v_2) \in E$.

1. Can we recover the signal by observing a subset of nodes by leveraging the smoothness property?
2. How many such $\mathbf{s}[v]$s needed to observe in order to recover $\mathbf{s}$? Even if some nodes are noisy?

The Noisy Model

$\mathcal{L}$ is the Laplacian matrix of G, with eigendecomposition $\mathcal{L} = \mathcal{V}\Sigma\mathcal{V}^{-1}$, $\mathcal{V}_{(k)}$ are the eigenvectors of the smallest k eigenvlaues. Signal s is said to be k-bandlimited if $s \in \text{Span}\{\mathcal{V}_{(k)}\}$. For $I \subset [n]$ with $|I| = k$, s_I is the projection of s indexed by I, V_I is the submatrix of $\mathcal{V}_{(k)}$ with rows indexed by I.

- If s_I observed exactly, then $\mathcal{V}V_I^{-1}s_I$ recovers s perfectly provide V_I is full rank.
- If $s_I + e$ is observed with noise e, the above recover method gives **worst case error** to be $\sigma_{\max}(V_I^{-1}) = 1/\sigma_{\min}(V_I)$.
- We aim to find I that makes the worst case error minimal, i.e., $\sigma_{\min}(V_I)$ is maximal.

Max $\sigma_{\min}(C)$ Problem [CHEN15]

For a $n \times k$ matrix A with $k \ll n$, how to find a $k \times k$ submatrix C of A by selecting k rows from A, such that the

$$\sigma_{\min}(C)$$

is maximal?

Previous Works

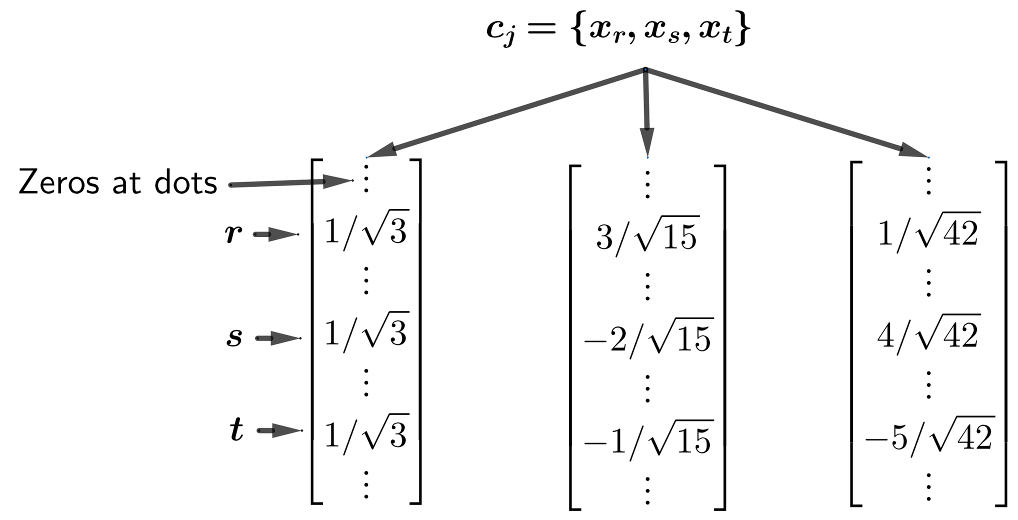
- A greedy algorithm that maximizing $\sigma_{\min}(C)$ is provided in [CHEN15], no bounds given.
- Greedy algorithms for minimizing $\text{Vol}(C^{-1})$ and $||C^{-1}||_F$ studied in [TSIT16], no bounds given.
- An submodular approach for analyzing MSE bound was given in [CHAM17], depends on the factor α and only works for Gaussian noise.
- Maximizing $\text{Vol}(C)$ studied in [CIVR09,13] and [NIKO15], c^k-inapproximability result is given, as well as c^k-approximation algorithm provided.

Our Contribution

1. We prove formally that the Max $\sigma_{\min}(C)$ problem is NP-hard.
2. For any $\epsilon > 0$, we give an approximation algorithm that solve the Max $\sigma_{\min}(C)$ problem, with **provable** approximation ratio of $\frac{1}{(1+\epsilon)k}$.

NP-hardness Reduction

We follow the idea in [CIVR09]. The reduction is from EXACT-3-COVER(X3C): let $X = \{x_1, \cdots, x_k\}$ and $C = \{c_1, \cdots, c_n\}$ where $c_i \subset X$ with $|c_i| = 3$. Decide if there are $k/3$ disjoint elements from C such that their union is X.



Let A be the matrix with rows of all such vectors. One can show that the X3C is true iff there exist C in A such that $C^TC = I$.

The Approx. Algorithm

Step One:

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

Step Two:

1. Let $C \leftarrow$ Step One
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 until there is no update. Output C.

Theorem

Let C_{opt} be the submatrix of A with $\sigma_{\min}$ maximal, C is the output of **Step Two**, then

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

and **Step Two** runs in $\text{poly}(n, k)$.

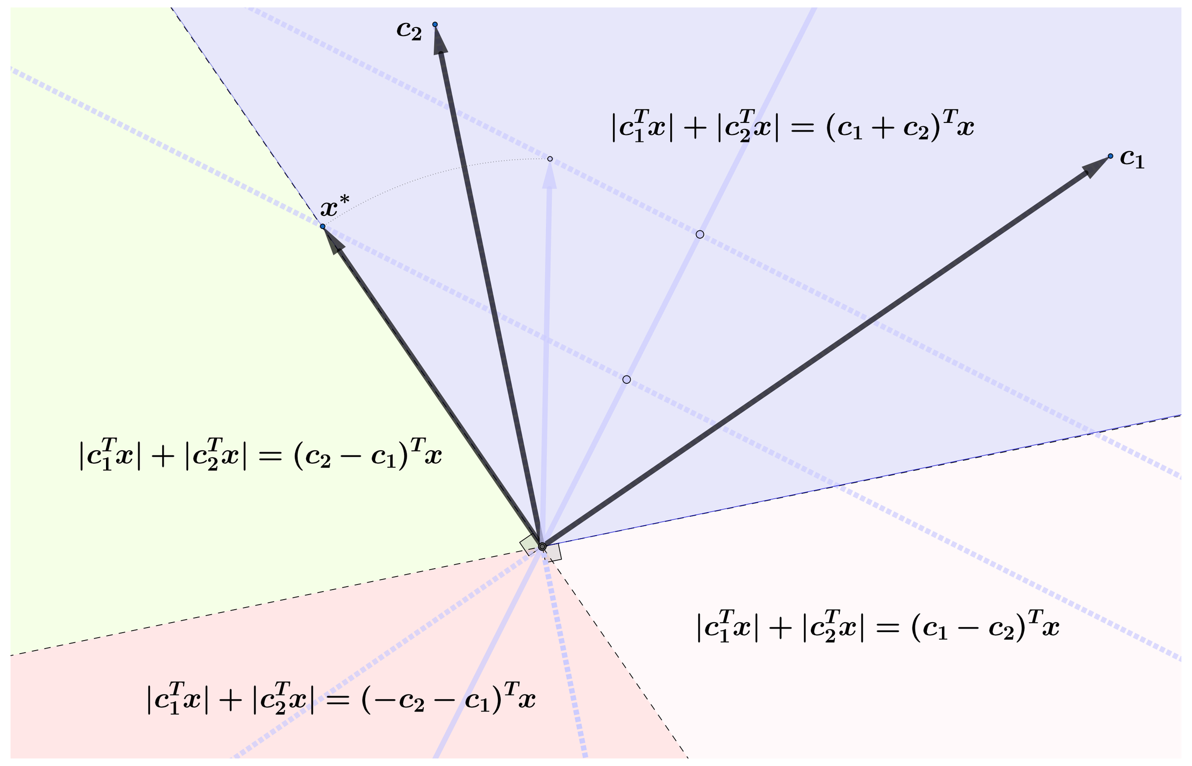
Sketch of Proof

1. Let $\alpha(C) = \min_{x \in \mathbb{R}^k : ||x||_2 = 1} \sum_{c_i \in C} |c_i^T x|$, c_i is the ith row of C. Let x^* achieves the minimal, show that $x^* \perp H_{-j}$ with $H_{-j} = \text{Span}\{c_i \in C | i \neq j\}$.
2. Show that $\alpha(C)/\sqrt{k} \leq \sigma_{\min}(C) \leq \alpha(C)$.
3. Show that if **Step Two** stops with output C, one has $\alpha(C) \geq \sigma_{\min}(C_{opt})/(1+\epsilon)\sqrt{k}$.
4. Note that the $\text{Vol}(C)$ in **Step Two** will increase by factor $(1+\epsilon)$ after each iteration.

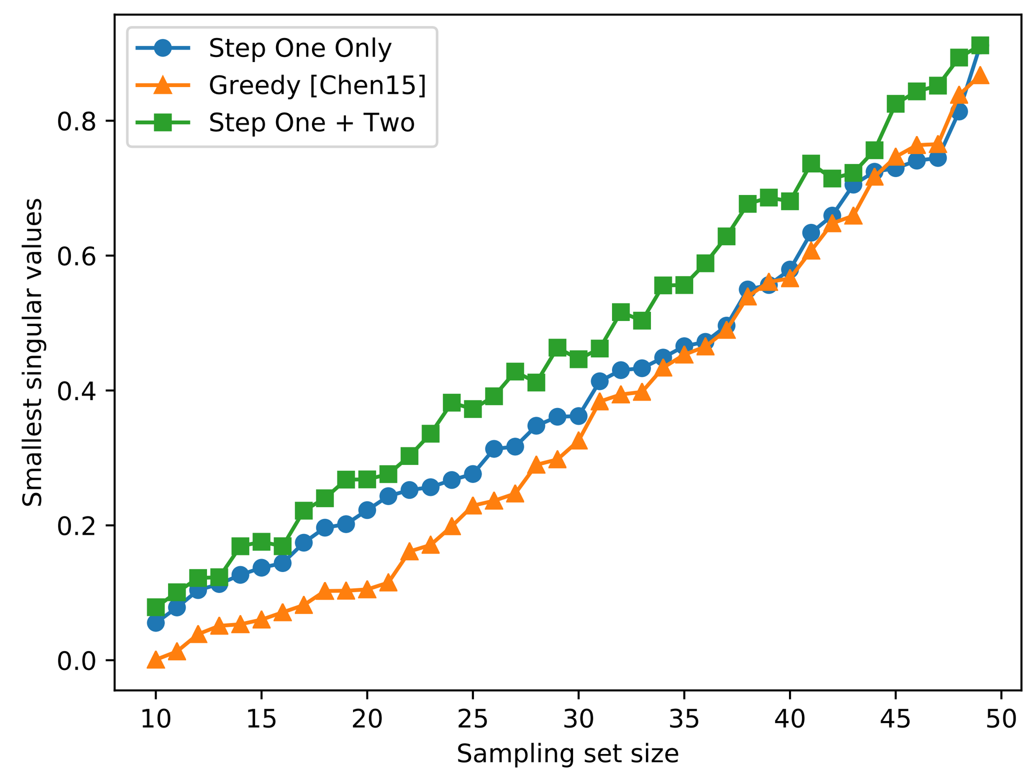
Sketch of Proof(Cont.)

5. By [CIVR09], the output of **Step One** has volume approximation of $k!$, thus **Step Two** has at most $k \log_{1+\epsilon} k$ iterations.

The 1st step of the proof can be illustrated as following picture:



Empirical Simulation



We run our algorithms over Erdos-Renyi graph ($p = 0.5$) with $n = 50$ nodes and $k = 10$ to choose the first 10 samples, then use naive greedy to choose the following nodes when sample size greater than 10.

Acknowledgments: The authors were partially supported by Center for Science of Information(CSoI), an NSF Science and Technology center, under grant agreement CCF-0939370.