ECE 8201: Low-dimensional Signal Models for High-dimensional Data Analysis

Lecture 4: Sparse signal recovery via greedy algorithms

Yuejie Chi
The Ohio State University

THE OHIO STATE UNIVERSITY

Outline

- One-step thresholding
- Orthogonal Matching Pursuit (OMP)
- Compressive Sampling Matching Pursuit (CoSaMP)

References

- Tropp, J. (2004). Greed is good: Algorithmic results for sparse approximation. Information Theory, IEEE Transactions on, 50(10), 2231-2242.
- Needell, D., & Tropp, J. A. (2009). CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Applied and Computational Harmonic Analysis, 26(3), 301-321.

See also:

- Cai, T. T., & Wang, L. (2011). Orthogonal matching pursuit for sparse signal recovery with noise. Information Theory, IEEE Transactions on, 57(7), 4680-4688.

Greedy algorithms

Consider the noise-free case

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is k-sparse, and $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ with unit-norm columns, i.e. $\|\boldsymbol{a}_i\|_2 = 1$.

Our goal is to estimate $\boldsymbol{x}$ from $\boldsymbol{y}$.

If $\boldsymbol{x}$ is one-sparse as $\boldsymbol{x} = \boldsymbol{e}_i$ which is a basis vector in $\mathbb{R}^n$, then $\boldsymbol{y}$ is just $\boldsymbol{a}_i$, and a natural way to determine i is using *matched filter*:

$$i^* = \operatorname*{argmax}_{1 \le i \le n} |\langle \boldsymbol{a}_i, \boldsymbol{y} \rangle|$$

We would like to extend this principle to handle sparsity level greater than one.

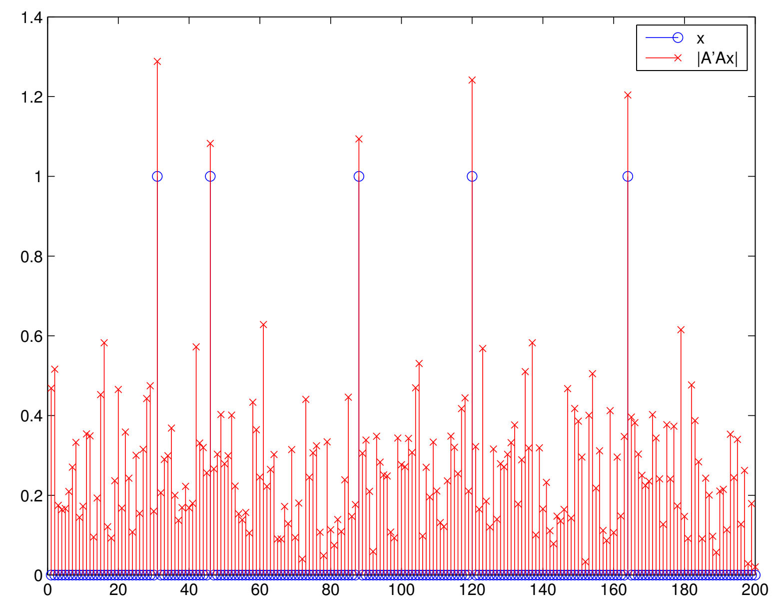
One-step thresholding

One-Step Thresholding (OST) for support recovery (assume k is known):

1. Compute:

$$\boldsymbol{z} = \boldsymbol{A}^{\top}\boldsymbol{y}$$

2. Find the support as the k largest entries of $|\boldsymbol{z}|$.



Performance of OST

It is easy to analyze the performance of OST via mutual coherence, which is defined as

$$\mu = \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|.$$

Note that

$$\boldsymbol{z} = \boldsymbol{A}^\top \boldsymbol{y} = \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x}$$

If the interference from other nonzero entries of $\boldsymbol{x}$ is small enough, it is possible to read off the support of $\boldsymbol{x}$ from the largest entries of $\boldsymbol{z}$.

Without loss of generality, assume $\boldsymbol{x}$ is k-sparse with the nonzero entries indexed by $\{1, \ldots, k\}$, in a descending order $|x_1| \geq |x_2| \geq \ldots \geq |x_k|$.

To guarantee the success of OST, we want to show

$$\min_{1 \leq i \leq k} |z_i| > \max_{k+1 \leq i \leq n} |z_i|.$$

Lower bound $\min_{1\le i\le k}|z_i|$

For $1 \le i \le k$,

$$\begin{aligned}
|z_i| &= |\boldsymbol{a}_i^\top \boldsymbol{A}\boldsymbol{x}| \\
&= |\boldsymbol{a}_i^\top(\boldsymbol{a}_i x_i + \sum_{j\ne i} \boldsymbol{a}_j x_j)| \\
&= |x_i + \sum_{j\ne i} \boldsymbol{a}_i^\top \boldsymbol{a}_j x_j| \\
&\ge |x_i| - \sum_{j\ne i} |\boldsymbol{a}_i^\top \boldsymbol{a}_j||x_j| \\
&\ge |x_i| - \mu(\|\boldsymbol{x}\|_1 - |x_i|) \\
&\ge (1+\mu)|x_i| - \mu\|\boldsymbol{x}\|_1,
\end{aligned}$$

therefore, $\min_{1\le i\le k}|z_i| \ge (1+\mu)\min_i |x_i| - \mu\|\boldsymbol{x}\|_1$.

Upper bound $\max_{k+1 \le i \le n} |z_i|$

For $k + 1 \le i \le n$,

$$\begin{aligned}
|z_i| &= |\boldsymbol{a}_i^\top \boldsymbol{A}\boldsymbol{x}| \\
&= |\boldsymbol{a}_i^\top \sum_{j=1}^{k} \boldsymbol{a}_j x_j| \\
&\le \sum_{j=1}^{k} |\boldsymbol{a}_i^\top \boldsymbol{a}_j||x_j| \\
&\le \mu \|\boldsymbol{x}\|_1
\end{aligned}$$

Putting everything together

OST succeeds if

$$(1+\mu)\min_i |x_i| - \mu\|\boldsymbol{x}\|_1 > \mu\|\boldsymbol{x}\|_1$$

which yields

$$(1+\mu)\min_i |x_i| > 2\mu\|\boldsymbol{x}\|_1.$$

or equivalently

$$\frac{\min_i |x_i|}{\|\boldsymbol{x}\|_1} > \frac{2\mu}{(1+\mu)}.$$

- If $|x_1| = \cdots = |x_k|$, the LHS becomes $1/k$ and for success support recovery we require

  $$\frac{1}{k} > \mu \sim \frac{1}{\sqrt{m}}$$

  which requires $m \gtrsim k^2$.

Better strategies

It is obvious better approaches exist, for example, by applying iterations.

The idea is through iterations, we can either iteratively identify new atoms in the sparse representation, or refine our earlier estimate.

- Orthogonal Matching Pursuit (OMP)
- Compressive Sampling Matching Pursuit (CoSaMP)

OMP

OMP (assume k is known):

1. Initialize: the residual $\boldsymbol{r}_0 = \boldsymbol{y}$, and $S_0 = \emptyset$.

2. For $i = 1, \ldots, k$:

   - choose the atom that has the largest correlation with the residual:

$$t = \underset{j}{\operatorname{argmax}} \, |\langle \boldsymbol{a}_j, \boldsymbol{r}_{i-1} \rangle|$$

   - Add t to the support set: $S_i = \{S_{i-1}, t\}$;
   - Update the residual as

$$\boldsymbol{r}_i = (\boldsymbol{I} - \boldsymbol{A}_{S_i} \boldsymbol{A}_{S_i}^{\dagger}) \boldsymbol{y}.$$

OMP doesn't select the same atom twice

If $j \in S_{i-1}$ has been selected,

$$\begin{aligned} \langle \boldsymbol{a}_j, \boldsymbol{r}_{i-1} \rangle &= \langle \boldsymbol{a}_j, (\boldsymbol{I} - \boldsymbol{A}_{S_i} \boldsymbol{A}_{S_i}^{\dagger}) \boldsymbol{y} \rangle \\ &= \boldsymbol{y}^{\mathsf{T}} (\boldsymbol{I} - \boldsymbol{A}_{S_i} \boldsymbol{A}_{S_i}^{\dagger}) \boldsymbol{a}_j = 0, \end{aligned}$$

therefore j won't be selected again by OMP.

If in each step OMP selects a correct index in T, in k iterations it will select all indices in T and terminates.

An alternative way to terminate OMP (without the knowledge of k) is to examine the norm of the residual $\|\boldsymbol{r}_j\|_2 < \epsilon$.

Tropp's Exact Recovery Condition (ERC) for OMP

Theorem 1. [ERC] *Suppose that* $\boldsymbol{x}$ *be a* k*-sparse signal supported on* T*. OMP recovers the* k*-term representation of* $\boldsymbol{x}$ *whenever*

$$\max_{\boldsymbol{a} \in T^c} \|\boldsymbol{A}_T^{\dagger} \boldsymbol{a}\|_1 < 1$$

where $\dagger$ *denotes pseudo-inverse.*

- This condition also guarantees the success of BP, see [Tropp 2004].
- Interestingly enough, this condition only depends on $\boldsymbol{A}$, not on the coefficients of $\boldsymbol{x}$ - much improved from OST.
- A natural question is when does this condition hold?

Tropp's Exact Recovery Condition (ERC) for OMP

Theorem 2. *ERC holds for* every *superposition of k atoms from $\boldsymbol{A}$ whenever*

$$k < \frac{1}{2}(\mu^{-1} + 1)$$

or more generally, whenever

$$\mu_1(k-1) + \mu_1(k) < 1$$

where $\mu_1(m)$ is defined as the Babel function of $\boldsymbol{A}$:

$$\mu_1(k) := \max_{|\Lambda|=k} \max_{i \in \Lambda^c} \sum_{\lambda \in \Lambda} |\langle \boldsymbol{a}_i, \boldsymbol{a}_\lambda \rangle|.$$

Remark: Since $\mu = \mu_1(1)$ and $\mu_1(k) \leq k\mu$, the latter condition implies the former condition.

Proof for ERC

Recall the support of $\boldsymbol{x}$ is T.

After i steps, assume OMP has already identified i correct indices in T. We would like to develop a condition that guarantees the next selected atom is also in T.

Motivated by our earlier discussions with OST, we only need to examine if the ratio

$$\rho(\boldsymbol{r}_k) = \frac{\left\|\boldsymbol{A}_{T^c}^{\mathsf{T}}\boldsymbol{r}_k\right\|_\infty}{\left\|\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{r}_k\right\|_\infty} < 1.$$

Realizing that $\boldsymbol{r}_k \in \mathsf{Span}(\boldsymbol{A}_T)$, we write

$$\boldsymbol{r}_k = \boldsymbol{A}_T\boldsymbol{A}_T^{\dagger}\boldsymbol{r}_k = \boldsymbol{A}_T(\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{A}_T)^{-1}\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{r}_k = (\boldsymbol{A}_T^{\dagger})^{\mathsf{T}}\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{r}_k.$$

This allows us to bound

$$\rho(\boldsymbol{r}_k) = \frac{\left\|\boldsymbol{A}_{T^c}^{\mathsf{T}}\boldsymbol{r}_k\right\|_\infty}{\left\|\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{r}_k\right\|_\infty} \leq \frac{\left\|\boldsymbol{A}_{T^c}^{\mathsf{T}}(\boldsymbol{A}_T^\dagger)^{\mathsf{T}}\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{r}_k\right\|_\infty}{\left\|\boldsymbol{A}_T^{\mathsf{T}}\boldsymbol{r}_k\right\|_\infty} \leq \left\|\boldsymbol{A}_{T^c}^{\mathsf{T}}(\boldsymbol{A}_T^\dagger)^{\mathsf{T}}\right\|_{\infty,\infty}$$

where the matrix norm $\|\cdot\|_{p,p}$ is defined as

$$\|\boldsymbol{R}\|_{p,p} := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{R}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}.$$

It is easy to check (by yourself) that

- $\|\boldsymbol{R}\|_{\infty,\infty}$ equals the maximum absolute row sum of $\boldsymbol{R}$;
- $\|\boldsymbol{R}\|_{1,1}$ equals the maximum absolute column sum of $\boldsymbol{R}$;

we have

$$\rho(\boldsymbol{r}_k) \leq \left\|\boldsymbol{A}_{T^c}^{\mathsf{T}}(\boldsymbol{A}_T^\dagger)^{\mathsf{T}}\right\|_{\infty,\infty} = \left\|\boldsymbol{A}_T^\dagger\boldsymbol{A}_{T^c}\right\|_{1,1} = \max_{i\in T^c}\left\|\boldsymbol{A}_T^\dagger\boldsymbol{a}_i\right\|_1.$$

Proof of Theorem 2

It is sufficient to show that ERC holds when

$$\mu_1(k-1) + \mu_1(k) < 1$$

where $\mu_1(m)$ is defined as the Babel function of $\boldsymbol{A}$:

$$\mu_1(k) := \max_{|\Lambda|=k} \max_{i \in \Lambda^c} \sum_{\lambda \in \Lambda} |\langle \boldsymbol{a}_i, \boldsymbol{a}_\lambda \rangle|.$$

Recall the ERC can be upper bounded as

$$\begin{aligned} \max_{i \in T^c} \left\| \boldsymbol{A}_T^\dagger \boldsymbol{a}_i \right\|_1 &= \max_{i \in T^c} \left\| (\boldsymbol{A}_T^\top \boldsymbol{A}_T)^{-1} \boldsymbol{A}_T^\top \boldsymbol{a}_i \right\|_1 \\ &\leq \left\| (\boldsymbol{A}_T^\top \boldsymbol{A}_T)^{-1} \right\|_{1,1} \max_{i \in T^c} \left\| \boldsymbol{A}_T^\top \boldsymbol{a}_i \right\|_1, \qquad (*) \end{aligned}$$

where the second term can be bounded by the Babel function

$$\max_{i \in T^c} \left\| \boldsymbol{A}_T^\top \boldsymbol{a}_i \right\|_1 = \max_{i \in T^c} \sum_{j \in T} |\langle \boldsymbol{a}_j, \boldsymbol{a}_i \rangle| \leq \mu_1(k).$$

For the first term, we set off to write $\boldsymbol{A}_T^\top \boldsymbol{A}_T$ as

$$\boldsymbol{A}_T^\top \boldsymbol{A}_T = \boldsymbol{I} + \boldsymbol{\Phi}$$

where $\phi_{ij} = \langle \boldsymbol{a}_{T_i}, \boldsymbol{a}_{T_j} \rangle$, and

$$\|\boldsymbol{\Phi}\|_{1,1} = \max_{l} \sum_{j \neq l} |\langle \boldsymbol{a}_{T_l}, \boldsymbol{a}_{T_j}| \leq \mu_1(k-1).$$

If $\|\boldsymbol{\Phi}\|_{1,1} < 1$, the von Neumann series $\sum_{k=0}^{\infty} (-\boldsymbol{\Phi})^k$ converges to $(\boldsymbol{I} + \boldsymbol{\Phi})^{-1}$, we can compute

$$\begin{aligned}\left\|(\boldsymbol{A}_T^\mathsf{T}\boldsymbol{A}_T)^{-1}\right\|_{1,1} &= \left\|(\boldsymbol{I}+\boldsymbol{\Phi})^{-1}\right\|_{1,1}\\ &= \left\|\sum_{k=0}^{\infty}(-\boldsymbol{\Phi})^k\right\|_{1,1}\\ &\le \sum_{k=0}^{\infty}\|(-\boldsymbol{\Phi})\|_{1,1}^k = \frac{1}{1-\|\boldsymbol{\Phi}\|_{1,1}} \le \frac{1}{1-\mu_1(k-1)}.\end{aligned}$$

Plugging this into (*), a sufficient condition to guarantee ERC is

$$\frac{\mu_1(k)}{1-\mu_1(k-1)} < 1$$

which gives

$$\mu_1(k-1)+\mu_1(k) < 1.$$

CoSaMP

Compressive Sampling Matching Pursuit (CoSaMP) with k known

1. Initialization: the residual $\boldsymbol{r}_0 = \boldsymbol{y}$, signal estimation $\boldsymbol{x}_0 = \boldsymbol{0}$,

2. For $i = 1, 2, \ldots$

   - Identify the $2k$ largest coefficients of the signal proxy $\boldsymbol{z}_i = \boldsymbol{A}^\top \boldsymbol{r}_{i-1}$:

$$\Omega = \text{supp}(\boldsymbol{z}_{2k})$$

   - Merge support: $S = \Omega \cup \text{supp}(\boldsymbol{x}_{i-1})$;
   - Estimation by least-squares:

$$\boldsymbol{b}_S = \boldsymbol{A}_S^\dagger \boldsymbol{y}, \quad \boldsymbol{b}_{S^c} = 0;$$

   - Pruning to obtain the next estimate: $\boldsymbol{x}_i = \boldsymbol{b}_k$ as the k-term approximation to $\boldsymbol{b}$;
   - Residual update:

$$\boldsymbol{r}_i = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_i$$

Performance of CoSaMP

- The stopping criteria of CoSaMP can be either based on residual energy, or a fixed number of iterations;

We will analyze CoSaMP for exactly k-sparse signals without measurement noise. It is not difficult to extend the analysis to the general case.

Theorem 3. [Needell and Tropp, 2008] *Assume $\boldsymbol{A}$ satisfies the RIP with $\delta_{2k} \leq 0.05$. For any k-sparse signal $\boldsymbol{x}$, the reconstruction in the ith iteration $\boldsymbol{x}_i$ is k-sparse, and satisfies*

$$\|\boldsymbol{x} - \boldsymbol{x}_{i+1}\|_2 \leq 0.26 \cdot \|\boldsymbol{x} - \boldsymbol{x}_i\|_2.$$

Moreover, CoSaMP is exact after at most $6(k+1)$ iterations.

Similar performance as BP order-wise.

A useful lemma

Proposition 1. *Suppose* $\boldsymbol{A}$ *has restricted isometry constant* δ_r. *Let* T *be a set of indices, and let* $\boldsymbol{x}$ *be a vector. Provided that* $r \geq |T \cup \operatorname{supp}(\boldsymbol{x})|$,

$$\|\boldsymbol{A}_T^\top \boldsymbol{A}_{T^c} \boldsymbol{x}_{T^c}\|_2 \leq \delta_r \|\boldsymbol{x}_{T^c}\|_2.$$

Proof: Define $S = \operatorname{supp}(\boldsymbol{x}) \backslash T$, we have $\boldsymbol{x}_S = \boldsymbol{x}_{T^c}$. Thus,

$$\begin{aligned}
\|\boldsymbol{A}_T^\top \boldsymbol{A}_{T^c} \boldsymbol{x}_{T^c}\|_2 &= \|\boldsymbol{A}_T^\top \boldsymbol{A}_S \boldsymbol{x}_S\|_2 \\
&\leq \|\boldsymbol{A}_T^\top \boldsymbol{A}_S\| \cdot \|\boldsymbol{x}_S\|_2 \leq \delta_r \|\boldsymbol{x}_{T^c}\|_2.
\end{aligned}$$

where we used $\|\boldsymbol{A}_T^\top \boldsymbol{A}_S\| \leq \delta_r$ from the near orthogonality-preserving property of RIP.

Proposition 2. *For positive integers* c *and* r, $\delta_{cr} \leq c\delta_{2r}$.

Progress of CoSaMP in one iteration

Without loss of generality we write $\boldsymbol{x}^o = \boldsymbol{x}_i$ as the previous reconstruction, and the new reconstruction as $\boldsymbol{x}^n = \boldsymbol{x}_{i+1}$.

Denote the error as $\boldsymbol{\nu} = \boldsymbol{x} - \boldsymbol{x}^o$, which is $2k$-sparse. The measurement residual can be written as

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^o = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^o) := \boldsymbol{A}\boldsymbol{\nu}.$$

1. Identification: the identified indices captures most of the energy in $\boldsymbol{s}$.

$$\|\boldsymbol{\nu}_{\Omega^c}\|_2 \leq 0.1053\|\boldsymbol{\nu}\|_2$$

Proof: Denote the support of $\boldsymbol{\nu}$ as $R = \text{supp}(\boldsymbol{\nu})$. By the choice Ω, we have $\|\boldsymbol{z}_R\|_2 \leq \|\boldsymbol{z}_\Omega\|_2$. Squaring the inequality and canceling the terms in $R \cap \Omega$, we have

$$\|\boldsymbol{z}_{R\setminus\Omega}\|_2 \leq \|\boldsymbol{z}_{\Omega\setminus R}\|_2.$$

On one side,

$$\|\boldsymbol{z}_{\Omega \backslash R}\|_2 = \|\boldsymbol{A}_{\Omega \backslash R}^{\mathsf{T}} \boldsymbol{A}\boldsymbol{\nu}\|_2 \leq \delta_{2k} \|\boldsymbol{\nu}\|_2$$

On the other side,

$$\begin{aligned}
\|\boldsymbol{z}_{R \backslash \Omega}\|_2 &= \|\boldsymbol{A}_{R \backslash \Omega}^{\mathsf{T}} \boldsymbol{A}\boldsymbol{\nu}\|_2 = \|\boldsymbol{A}_{R \backslash \Omega}^{\mathsf{T}} \boldsymbol{A}(\boldsymbol{\nu}_{R \backslash \Omega} + \boldsymbol{\nu}_{\Omega})\|_2 \\
&\geq \|\boldsymbol{A}_{R \backslash \Omega}^{\mathsf{T}} \boldsymbol{A}_{R \backslash \Omega} \boldsymbol{\nu}_{R \backslash \Omega}\|_2 - \|\boldsymbol{A}_{R \backslash \Omega}^{\mathsf{T}} \boldsymbol{A}\boldsymbol{\nu}_{\Omega})\|_2 \\
&\geq (1 - \delta_{2k}) \|\boldsymbol{\nu}_{R \backslash \Omega}\|_2 - \delta_{2k} \|\boldsymbol{\nu}\|_2 \\
&= (1 - \delta_{2k}) \|\boldsymbol{\nu}_{\Omega^c}\|_2 - \delta_{2k} \|\boldsymbol{\nu}\|_2.
\end{aligned}$$

Combining these we have

$$(1 - \delta_{2k}) \|\boldsymbol{\nu}_{\Omega^c}\|_2 - \delta_{2k} \|\boldsymbol{\nu}\|_2 \leq \delta_{2k} \|\boldsymbol{\nu}\|_2$$

which gives

$$\|\boldsymbol{\nu}_{\Omega^c}\|_2 \leq \frac{2\delta_{2k}}{1 - \delta_{2k}} \|\boldsymbol{\nu}\|_2 \leq 0.1053 \|\boldsymbol{\nu}\|_2.$$

2. Merge support: The signal $\boldsymbol{x}$ has little energy outside the merged support $S = \Omega \cup \mathsf{supp}(\boldsymbol{x}^o)$.

$$\|\boldsymbol{x}_{S^c}\|_2 \leq \|\boldsymbol{\nu}_{\Omega^c}\|_2$$

Proof: $\|\boldsymbol{x}_{S^c}\|_2 = \|(\boldsymbol{x} - \boldsymbol{x}^o)_{S^c}\|_2 = \|\boldsymbol{\nu}_{S^c}\|_2 \leq \|\boldsymbol{\nu}_{\Omega^c}\|_2$.

3. Estimation by least-squares on $\boldsymbol{A}_S$:

$$\|\boldsymbol{x} - \boldsymbol{b}\|_2 \leq 1.2352\|\boldsymbol{x}_{S^c}\|_2$$

Proof: Note that $\|\boldsymbol{x} - \boldsymbol{b}\|_2 \leq \|(\boldsymbol{x} - \boldsymbol{b})_S\|_2 + \|\boldsymbol{x}_{S^c}\|_2$. To bound the first term, we have ($\boldsymbol{A}_S^\dagger \boldsymbol{A}_S = \boldsymbol{I}$)

$$
\begin{aligned}
\|\boldsymbol{x}_S - \boldsymbol{b}_S\|_2 = \|\boldsymbol{x}_S - \boldsymbol{A}_S^\dagger \boldsymbol{y}\|_2 &= \|\boldsymbol{x}_S - \boldsymbol{A}_S^\dagger \boldsymbol{A}\boldsymbol{x}\|_2 \\
&= \|\boldsymbol{x}_S - \boldsymbol{A}_S^\dagger (\boldsymbol{A}_S \boldsymbol{x}_S + \boldsymbol{A}_{S^c} \boldsymbol{x}_{S^c})\|_2 \\
&= \|(\boldsymbol{A}_S^\top \boldsymbol{A}_S)^{-1} \boldsymbol{A}_S^\top \boldsymbol{A} \boldsymbol{x}_{S^c}\|_2 \\
&\leq \|(\boldsymbol{A}_S^\top \boldsymbol{A}_S)^{-1}\|_2 \|\boldsymbol{A}_S^\top \boldsymbol{A} \boldsymbol{x}_{S^c}\|_2 \\
&\leq \frac{\delta_{4k}}{1 - \delta_{3k}} \|\boldsymbol{x}_{S^c}\|_2 \leq 0.2352 \|\boldsymbol{x}_{S^c}\|_2.
\end{aligned}
$$

4. Pruning: the error introduced by pruning is small.

   Proof:

$$
\|\boldsymbol{x} - \boldsymbol{x}^n\|_2 = \|\boldsymbol{x} - \boldsymbol{b} + \boldsymbol{b} - \boldsymbol{x}^n\|_2 \leq \|\boldsymbol{x} - \boldsymbol{b}\|_2 + \|\boldsymbol{b} - \boldsymbol{x}^n\|_2 \leq 2\|\boldsymbol{x} - \boldsymbol{b}\|_2
$$

   since $\boldsymbol{x}^n$ is the best k-term approximation of $\boldsymbol{b}$.

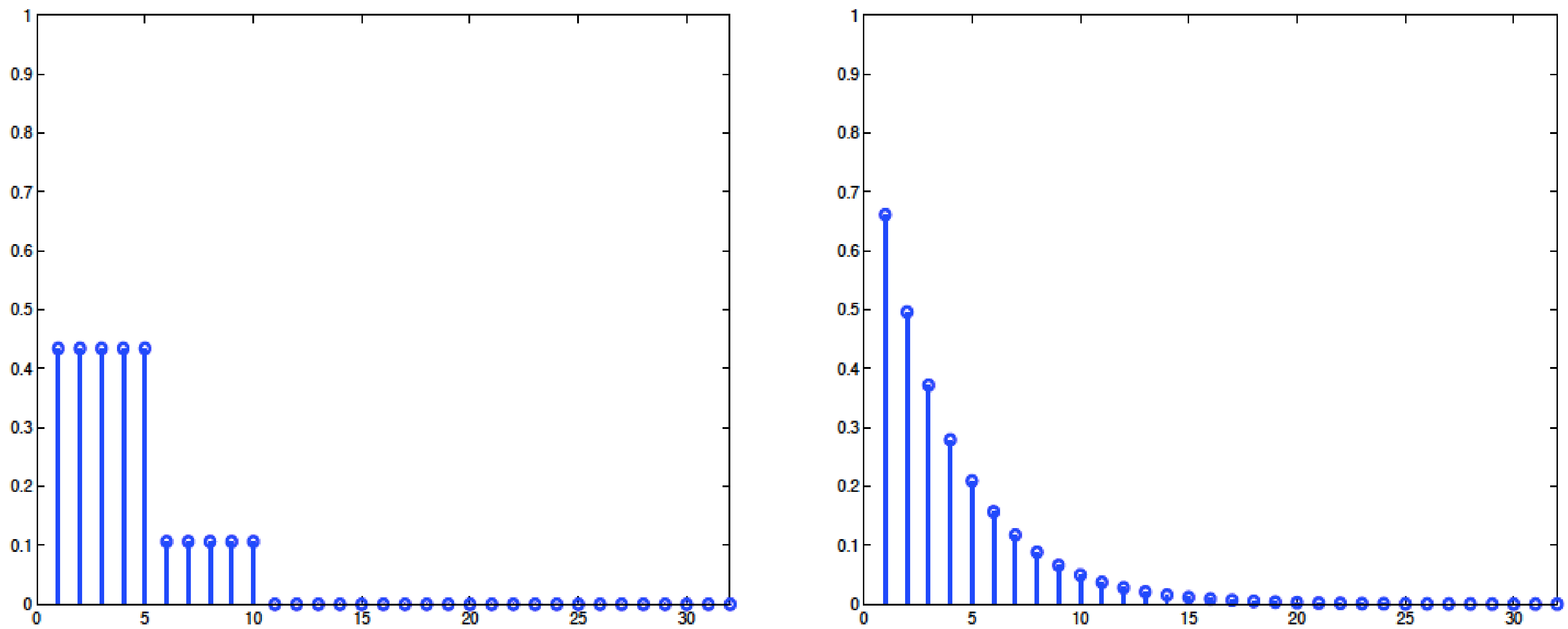
Putting everything together, we have

$$
\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{x}^n\|_2 &\leq 2\|\boldsymbol{x} - \boldsymbol{b}\|_2 \qquad \text{(pruning)} \\
&\leq 2 \cdot 1.2352\|\boldsymbol{x}_{S^c}\|_2 \qquad \text{(estimation)} \\
&\leq 2.4706\|\boldsymbol{\nu}_{\Omega^c}\|_2 \qquad \text{(merge support)} \\
&\leq 2.4706 \cdot 0.1053\|\boldsymbol{\nu}\|_2 \qquad \text{(identification)} \\
&< 0.26\|\boldsymbol{\nu}\|_2 = 0.26\|\boldsymbol{x}^o\|_2.
\end{aligned}
$$

Iteration Count

The number of iterations is at most $6(k+1)$, and could be as small as $\log k$.

It heavily relies on the coefficient profile.



(a) Low profile (b) High profile

FIGURE 1. Illustration of two unit-norm signals with sharply different profiles.