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

Lecture 2: Sparse signal recovery: Analysis of ℓ_1 minimization via RIP

Yuejie Chi
The Ohio State University

The Ohio State University

Outline

- Definition of sparse and compressible signals

  Reference: S. Foucart and H. Rauhut. A Mathematical Introduction to Compressive Sensing, Chapter 1.

- Uniqueness and identifiability using spark and coherence

  Reference: Donoho, D. L., & Elad, M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. 2003.

- ℓ_1 minimization, and sufficient condition for recovery using RIP

  Reference: E. J. Candès. The restricted isometry property and its implications for compressed sensing. 2008.

Signals that are exactly sparse

Consider a signal $\boldsymbol{x} \in \mathbb{R}^n$.

Definition 1. [Support] *The* support *of a vector* $\boldsymbol{x} \in \mathbb{R}^n$ *is the* index set *of its nonzero entries, i.e.*

$$\textit{supp}(\boldsymbol{x}) := \{j \in [n] : x_j \neq 0\}$$

where $[n] = \{1, \ldots, n\}$*.*

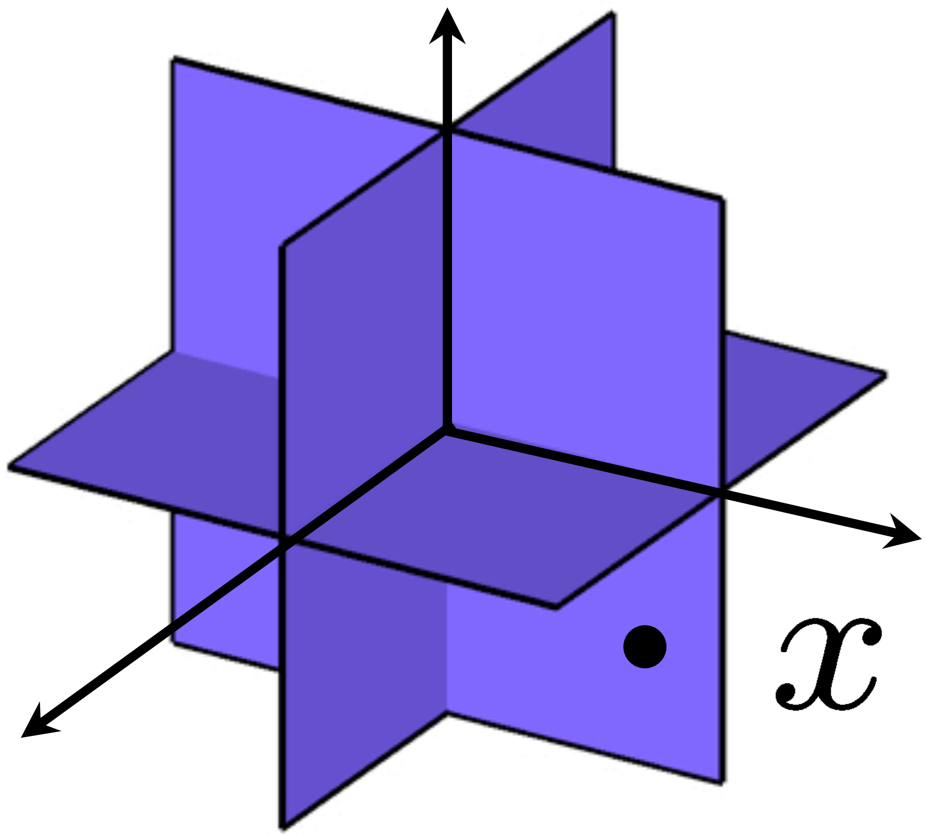
Definition 2. [k-sparse signal] *The signal* $\boldsymbol{x}$ *is called* k*-sparse, if*

$$\|\boldsymbol{x}\|_0 := |\textit{supp}(\boldsymbol{x})| \leq k.$$

Note: $\|\boldsymbol{x}\|_0$ is called the sparsity level of $\boldsymbol{x}$.

Sparse signals belong to union-of-subspace models

There're $\binom{n}{k}$ subspaces of dimension k.



Compressible signals

We're also interested in signals that are *approximately* sparse. This is measured by how well they can be approximated by sparse signals.

Definition 3. [Best k-term approximation] *Denote the index set of the k-largest entries of $|\boldsymbol{x}|$ as S_k. The best k-term approximation $\boldsymbol{x}_k$ of $\boldsymbol{x}$ is defined as*

$$\boldsymbol{x}_k(i) = \begin{cases} x_i, & i \in S_k \\ 0, & i \notin S_k \end{cases}$$

The k-term approximation error in ℓ_p norm is then given as

$$\|\boldsymbol{x} - \boldsymbol{x}_k\|_p = \left(\sum_{i \notin S_k} |x_i|^p \right)^{1/p}.$$

Compressibility: A signal is called *compressible* if $\|\boldsymbol{x} - \boldsymbol{x}_k\|_p$ decays fast in k.

Example of compressible signals

Proposition 1. [Compressibility] *For any* $q > p > 0$ *and* $\boldsymbol{x} \in \mathbb{R}^n$*,*

$$\|\boldsymbol{x} - \boldsymbol{x}_k\|_q \leq \frac{1}{k^{1/p-1/q}} \|\boldsymbol{x}\|_p.$$

Example: set $q = 2$ and $0 < p < 1$, we have

$$\|\boldsymbol{x} - \boldsymbol{x}_k\|_2 \leq \frac{1}{k^{1/p-1/2}} \|\boldsymbol{x}\|_p.$$

Consider a signal $\boldsymbol{x} \in B_p^n := \{\boldsymbol{z} \in \mathbb{R}^n : \|\boldsymbol{z}\|_p \leq 1\}$. Then $\boldsymbol{x}$ is compressible when $0 < p < 1$. [Geometrically, the ℓ_p-ball is pointy when $0 < p < 1$ in high dimension.]

Proof of Proposition 1: Without loss of generality we assume the coefficients of $\boldsymbol{x}$ is ordered in descending order of magnitudes. We then have

$$\begin{aligned}
\|\boldsymbol{x}-\boldsymbol{x}_k\|_q^q &= \sum_{j=k+1}^{n} |x_j|^q \quad \text{(by definition)} \\
&= |x_k|^{q-p} \sum_{j=k+1}^{n} |x_j|^p (|x_j|/|x_k|)^{q-p} \\
&\le |x_k|^{q-p} \sum_{j=k+1}^{n} |x_j|^p \quad (|x_j|/|x_k| \le 1) \\
&\le \left(\frac{1}{k}\sum_{j=1}^{k} |x_j|^p\right)^{\frac{q-p}{p}} \left(\sum_{j=k+1}^{n} |x_j|^p\right) \\
&\le \left(\frac{1}{k}\|\boldsymbol{x}\|_p^p\right)^{\frac{q-p}{p}} \|\boldsymbol{x}\|_p^p = \frac{1}{k^{q/p-1}}\|\boldsymbol{x}\|_p^q.
\end{aligned}$$

Compressive acquisition of sparse signals

- Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be the measurement/sensing matrix. Consider, for start, noise-free measurements:
$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m,$$
where $m \ll n$. We are interested in reconstructing $\boldsymbol{x}$ from $\boldsymbol{y}$.

- Since we want to motivate sparse solutions, we could seek the sparsest signal satisfying the observation:
$$\text{(P0:)} \quad \hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x}} \ \|\boldsymbol{x}\|_0 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$
where $\|\cdot\|_0$ counts the number of nonzero entries.

- Although this algorithm is NP-hard, we can still analyze when it is expected to work.

Spark and uniqueness

Question: What properties do we seek in $\boldsymbol{A}$ regardless of complexity of reconstruction algorithms?

Definition 4. [Spark] *Let Spark($\boldsymbol{A}$) be the size of the smallest linearly dependent subset of columns of $\boldsymbol{A}$.*

Basic Fact: $2 \leq \text{Spark}(\boldsymbol{A}) \leq m+1$.

Theorem 1. [Uniqueness, Donoho and Elad 2002] *A representation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ is necessarily the sparsest possible if $\|\boldsymbol{x}\|_0 < \text{Spark}(\boldsymbol{A})/2$.*

Proof: If $\boldsymbol{x}$ and $\boldsymbol{x}'$ satisfy $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}'$, with $\|\boldsymbol{x}'\|_0 \leq \|\boldsymbol{x}\|_0$, then

$$\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}') = 0$$

for $\|\boldsymbol{x} - \boldsymbol{x}'\|_0 < \text{Spark}(\boldsymbol{A})$, which contradicts with definition of Spark. Therefore, $\boldsymbol{x} = \boldsymbol{x}'$ and $\boldsymbol{x}$ is the sparsest solution of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

Mutual coherence

Definition 5. [Mutual Coherence] *Let*

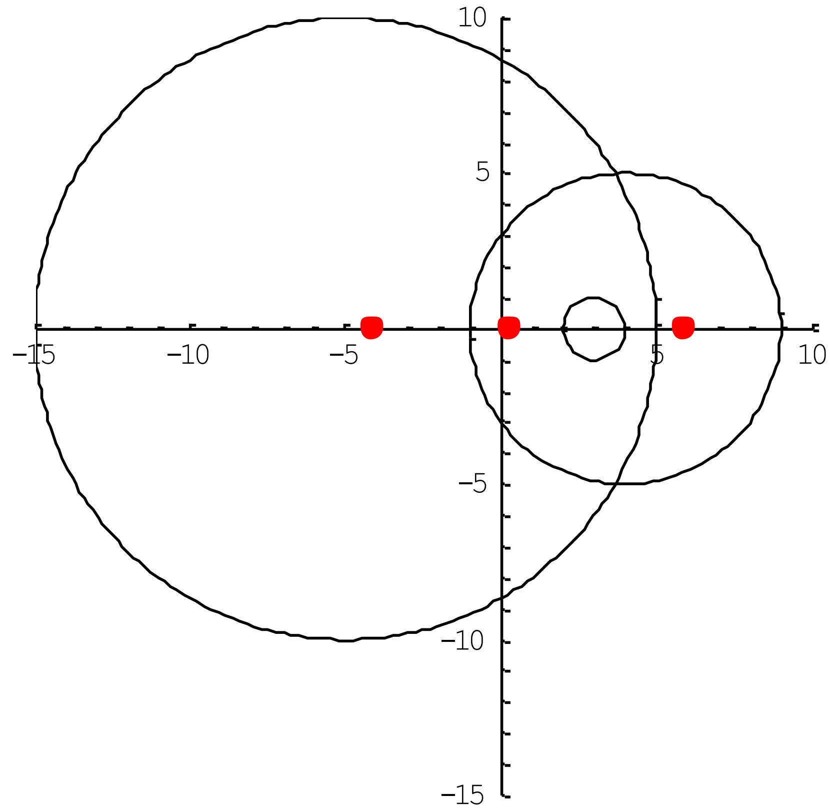
$$\mu = \mu(\boldsymbol{A}) := \max_{i \neq j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle|$$

. *where $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are normalized columns of $\boldsymbol{A}$.*

- $\mu(\boldsymbol{A}) \leq 1$ if the columns of $\boldsymbol{A}$ are pairwise independent.
- Spark$(\boldsymbol{A}) > 1/\mu(\boldsymbol{A})$ [can be shown by the Gershgorin circle's theorem].
- Welch bound asserts
$$\mu^2 \geq \frac{m-n}{n(m-1)},$$
which roughly gives $\mu = O(1/\sqrt{m})$ for a "well-behaved" $\boldsymbol{A}$.

Gershgorin circle's theorem

Lemma 2. [Gershgorin circle's theorem] *The eigenvalues of an $n \times n$ matrix $\boldsymbol{M}$ with entries m_{ij}, $1 \leq i, j \leq n$, lie in the union of n discs $d_i = d_i(c_i, r_i)$, $1 \leq i \leq n$, centered at $c_i = m_{ii}$ and with radius $r_i = \sum_{j \neq i} |m_{ij}|$.*



Example : take $\boldsymbol{M} = \begin{bmatrix} 4 & 2 & 3 \\ -2 & -5 & 8 \\ 1 & 0 & 3 \end{bmatrix}$

Sufficient condition using mutual coherence

Theorem 3. [Equivalence, Donoho and Elad 2002] *The sparsest solution to* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ *is unique if* $\|\boldsymbol{x}\|_0 < \frac{1}{2} + \frac{1}{2\mu(\boldsymbol{A})}$.

- The largest recoverable sparsity of $\boldsymbol{x}$ is $k \sim O(1/\mu) = O(\sqrt{m})$, which is square-root in the number of measurements.

- This result is deterministic.

- Requires the signal to be exactly sparse, which is not always practical.

Sparse Recovery via ℓ_1 Minimization

Since the above ℓ_0 minimization is NP-hard. We would like to take its convex relaxation, which leads to the ℓ_1 minimization, or basis pursuit:

$$\text{(BP:)} \quad \hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \ \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

- The BP algorithm does not assume knowledge of the sparsity level to perform.

- Compare this with the usual wisdom of ℓ_2 minimization:

$$\hat{\boldsymbol{x}}_{\ell_2} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \ \|\boldsymbol{x}\|_2 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}.$$

  which has a closed form solution

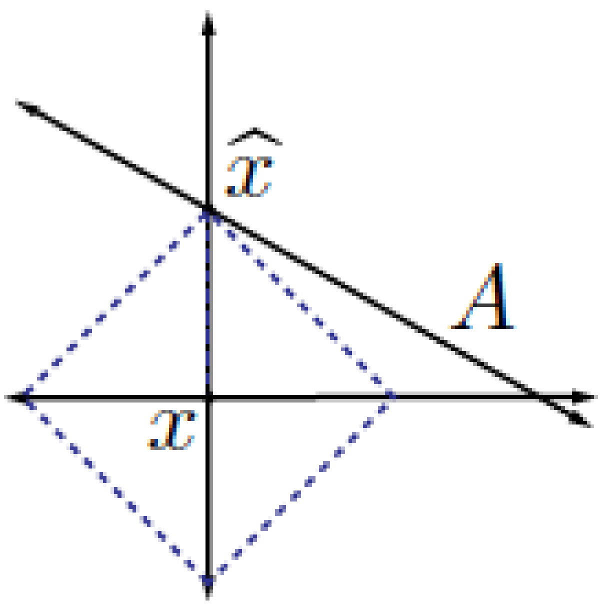
$$\hat{\boldsymbol{x}}_{\ell_2} = \boldsymbol{A}^{\dagger}\boldsymbol{y},$$

  where † denotes pseudo-inverse.

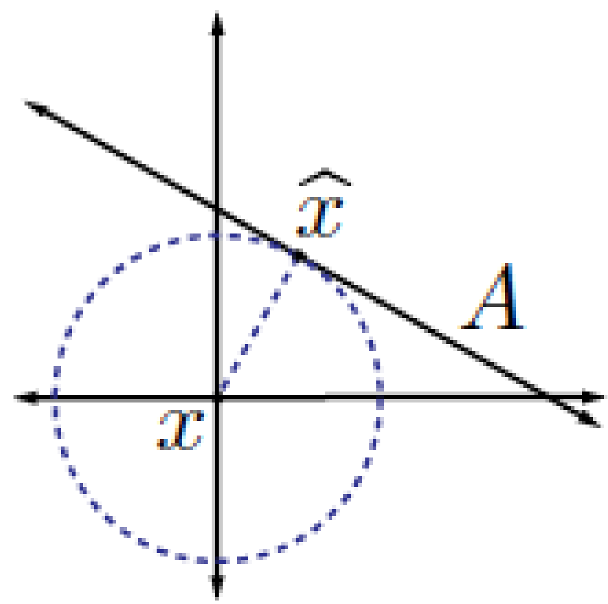
A numerical example

Let's run an example using CVX (`http://cvxr.com/cvx/`).

Geometry of basis pursuit



$p = 1$

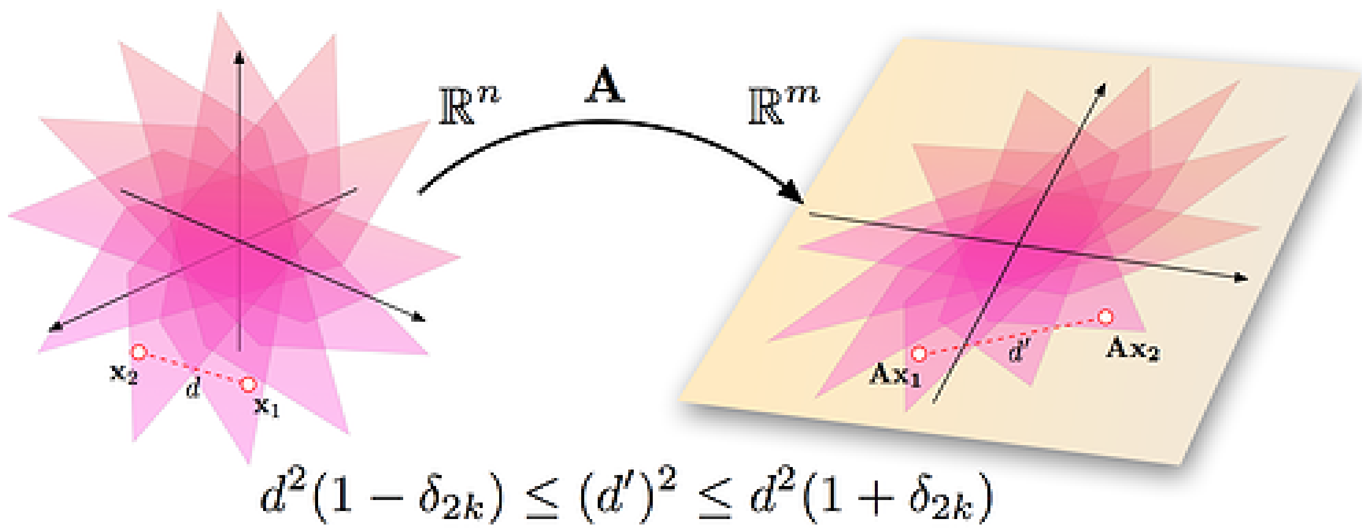


$p = 2$

Restricted isometry property

Definition 6. [Restricted Isometry Property (RIP)] *If $\boldsymbol{A}$ satisfies the restricted isometry property (RIP) with δ_{2k}, then for any two k-sparse vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:*

$$1 - \delta_{2k} \leq \frac{\|\boldsymbol{A}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_2^2}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2^2} \leq 1 + \delta_{2k}.$$



If $\delta_{2k} < 1$, this implies the ℓ_0 problem has a unique k-sparse solution.

RIP matrices preserve orthogonality between sparse vectors

Proposition 2.

$$|\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2\rangle| \le \delta_{s_1+s_2}\|\boldsymbol{x}_1\|_2\|\boldsymbol{x}_2\|_2$$

for all $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ that are supported on disjoint subsets $T_1, T_2 \subset [n]$ with $|T_1| \le s_1$ and $|T_2| \le s_2$.

Proof: Without loss of generality assume $\|\boldsymbol{x}_1\|_2 = \|\boldsymbol{x}_2\|_2 = 1$. Applying the parallelogram identity, which says

$$\begin{aligned} |\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2\rangle| &= \frac{1}{4}\left|\|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2\|_2^2 - \|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{A}\boldsymbol{x}_2\|_2^2\right| \\ &\le \frac{1}{4}|2(1+\delta_{s_1+s_2}) - 2(1-\delta_{s_1+s_2})| \le \delta_{s_1+s_2}. \end{aligned}$$

Restricted isometry property

Theorem 4. [Performance of BP via RIP, Candès, Tao, Romberg, 2006]
If $\delta_{2k} < \sqrt{2} - 1$, then for any vector $\boldsymbol{x}$, the solution to basis pursuit satisfies

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \le C_0 k^{-1/2} \|\boldsymbol{x} - \boldsymbol{x}_k\|_1.$$

where $\boldsymbol{x}_k$ is the best k-term approximation of $\boldsymbol{x}$ for some constant C_0.

- exact recovery if $\boldsymbol{x}$ is exactly k-sparse.
- Many random ensembles (e.g. Gaussian, sub-Gaussian, partial DFT) satisfies the RIP as soon as (we'll return to this point)

$$m \sim \Theta(k \log(n/k))$$

- The proof of theorem is particularly elegant.

Proof of Theorem 4

Proof of Theorem 4: Set $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{h}$. We already show $\boldsymbol{A}\boldsymbol{h} = 0$. The goal is to establish that $\boldsymbol{h} = 0$ when $\boldsymbol{A}$ satisfies the desired RIP.

The first step is to decompose $\boldsymbol{h}$ into a sum of vectors $\boldsymbol{h}_{T_0}$, $\boldsymbol{h}_{T_1}$, $\boldsymbol{h}_{T_2}$, $\ldots$, each of sparsity at most k. Here, T_0 corresponds to the locations of the k largest coefficients of $\boldsymbol{x}$; T_1 to the locations of the k largest coefficients of $\boldsymbol{h}_{T_0^c}$, T_2 to the locations of the next k largest coefficients of $\boldsymbol{h}_{T_0^c}$, and so on.

The proof proceeds in two steps:

1. the first step shows that the size of $\boldsymbol{h}$ outside of $T_0 \cup T_1$ is essentially bounded by that of $\boldsymbol{h}$ on $T_0 \cup T_1$.

2. the second step shows that $\|\boldsymbol{h}_{T_0 \cup T_1}\|_2$ is appropriately small.

Proof continued

Step 1: Note that for each $j \geq 2$,

$$\|\boldsymbol{h}_{T_j}\|_2 \leq \sqrt{k}\|\boldsymbol{h}_{T_j}\|_\infty \leq \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_{j-1}}\|_1$$

therefore

$$\sum_{j\geq 2} \|\boldsymbol{h}_{T_j}\|_2 \leq \frac{1}{\sqrt{k}}\sum_{j\geq 1} \|\boldsymbol{h}_{T_j}\|_1 = \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0^c}\|_1.$$

This allows us to bound

$$\|\boldsymbol{h}_{(T_0\cup T_1)^c}\|_2 \leq \|\sum_{j\geq 2} \boldsymbol{h}_{T_j}\|_2 \leq \sum_{j\geq 2} \|\boldsymbol{h}_{T_j}\|_2 \leq \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0^c}\|_1.$$

Given $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{h}$ is the optimal solution, we have

$$\begin{aligned}\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x} + \boldsymbol{h}\|_1 &= \sum_{i\in T_0} |x_i + h_i| + \sum_{i\in T_0^c} |x_i + h_i| \\ &\geq \|\boldsymbol{x}_{T_0}\|_1 - \|\boldsymbol{h}_{T_0}\|_1 + \|\boldsymbol{h}_{T_0^c}\|_1 - \|\boldsymbol{x}_{T_0^c}\|_1, \quad (*)\end{aligned}$$

which gives

$$\begin{aligned}\|\boldsymbol{h}_{T_0^c}\|_1 &\leq \|\boldsymbol{h}_{T_0}\|_1 + \|\boldsymbol{x}\|_1 - \|\boldsymbol{x}_{T_0}\|_1 + \|\boldsymbol{x}_{T_0^c}\|_1 \\ &\leq \|\boldsymbol{h}_{T_0}\|_1 + 2\|\boldsymbol{x}_{T_0^c}\|_1 := \|\boldsymbol{h}_{T_0}\|_1 + 2\|\boldsymbol{x} - \boldsymbol{x}_k\|_1.\end{aligned}$$

Combining with (*), we have

$$\|\boldsymbol{h}_{(T_0 \cup T_1)^c}\|_2 \leq \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0^c}\|_1 \leq \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0}\|_1 + \frac{2}{\sqrt{k}}\|\boldsymbol{x} - \boldsymbol{x}_k\|_1.$$

Step 2: We next bound $\|\boldsymbol{h}_{T_0 \cup T_1}\|_2$. Note that

$$0 = \boldsymbol{A}\boldsymbol{h} = \boldsymbol{A}\boldsymbol{h}_{T_0 \cup T_1} + \sum_{j \geq 2} \boldsymbol{A}\boldsymbol{h}_{T_j},$$

we have by RIP

$$(1 - \delta_{2k})\|\boldsymbol{h}_{T_0 \cup T_1}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{h}_{T_0 \cup T_1}\|_2^2 = |\langle \boldsymbol{A}\boldsymbol{h}_{T_0 \cup T_1}, \sum_{j \geq 2} \boldsymbol{A}\boldsymbol{h}_{T_j}\rangle|.$$

Using Proposition 2, we have for $j \geq 2$

$$\begin{aligned}
|\langle \boldsymbol{A}\boldsymbol{h}_{T_0 \cup T_1}, \boldsymbol{A}\boldsymbol{h}_{T_j} \rangle| &\leq |\langle \boldsymbol{A}\boldsymbol{h}_{T_0}, \boldsymbol{A}\boldsymbol{h}_{T_j} \rangle| + |\langle \boldsymbol{A}\boldsymbol{h}_{T_1}, \boldsymbol{A}\boldsymbol{h}_{T_j} \rangle| \\
&\leq \delta_{2k}(\|\boldsymbol{h}_{T_0}\|_2 + \|\boldsymbol{h}_{T_1}\|_2)\|\boldsymbol{h}_{T_j}\|_2 \\
&\leq \delta_{2k}\sqrt{2}\|\boldsymbol{h}_{T_0 \cup T_1}\|_2\|\boldsymbol{h}_{T_j}\|_2,
\end{aligned}$$

which gives

$$\begin{aligned}
(1 - \delta_{2k})\|\boldsymbol{h}_{T_0 \cup T_1}\|_2^2 &\leq \sum_{j \geq 2} |\langle \boldsymbol{A}\boldsymbol{h}_{T_0 \cup T_1}, \boldsymbol{A}\boldsymbol{h}_{T_j} \rangle| \\
&\leq \sqrt{2}\delta_{2k}\|\boldsymbol{h}_{T_0 \cup T_1}\|_2 \sum_{j \geq 2} \|\boldsymbol{h}_{T_j}\|_2 \\
&\leq \sqrt{2}\delta_{2k}\|\boldsymbol{h}_{T_0 \cup T_1}\|_2 \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0^c}\|_1,
\end{aligned}$$

therefore

$$\|\boldsymbol{h}_{T_0 \cup T_1}\|_2 \leq \frac{\sqrt{2}\delta_{2k}}{(1 - \delta_{2k})} \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0^c}\|_1 \leq \rho \frac{1}{\sqrt{k}}(\|\boldsymbol{h}_{T_0}\|_1 + 2\|\boldsymbol{x} - \boldsymbol{x}_k\|_1)$$

where $\rho := \frac{\sqrt{2}\delta_{2k}}{(1-\delta_{2k})}$. Since $\|\boldsymbol{h}_{T_0}\|_1 \leq \sqrt{k}\|\boldsymbol{h}_{T_0}\|_2 \leq \sqrt{k}\|\boldsymbol{h}_{T_0 \cup T_1}\|_2$, we can bound

$$\|\boldsymbol{h}_{T_0 \cup T_1}\|_2 \leq \frac{2\rho}{1-\rho}\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|_1}{\sqrt{k}}.$$

Finally,

$$\begin{aligned}
\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2 = \|\boldsymbol{h}\|_2 &\leq \|\boldsymbol{h}_{T_0 \cup T_1}\|_2 + \|\boldsymbol{h}_{(T_0 \cup T_1)^c}\|_2 \\
&\leq \|\boldsymbol{h}_{T_0 \cup T_1}\|_2 + \frac{1}{\sqrt{k}}\|\boldsymbol{h}_{T_0}\|_1 + \frac{2}{\sqrt{k}}\|\boldsymbol{x}-\boldsymbol{x}_k\|_1 \\
&\leq 2\|\boldsymbol{h}_{T_0 \cup T_1}\|_2 + \frac{2}{\sqrt{k}}\|\boldsymbol{x}-\boldsymbol{x}_k\|_1 \\
&\leq \frac{2(1+\rho)}{1-\rho}\frac{\|\boldsymbol{x}-\boldsymbol{x}_k\|_1}{\sqrt{k}}.
\end{aligned}$$

Therefore, $C_0 := \frac{2(1+\rho)}{1-\rho}$. The requirement on δ_{2k} comes from the fact that we need $1-\rho > 0$ to avoid the bound to blow up.

ℓ_1 recovery in the noisy case

In the presence of additive measurement noise,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w},$$

where $\|\boldsymbol{w}\|_2 \leq \epsilon$ is assumed to be bounded.

We can modify the BP algorithm in the following manner:

$$\text{(BP-noisy:)} \quad \hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon.$$

Theorem 5. [Performance of BP via RIP, noisy case] *If $\delta_{2k} < \sqrt{2} - 1$, then for any vector $\boldsymbol{x}$, the solution to basis pursuit (noisy case) satisfies*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq C_0 k^{-1/2} \|\boldsymbol{x} - \boldsymbol{x}_k\|_1 + C_1 \epsilon.$$

where $\boldsymbol{x}_k$ is the best k-term approximation of $\boldsymbol{x}$ for some constants C_0 and C_1.

Proof of Theorem 5

Again let's start by assuming $\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{h}$. The key difference from the noiseless case is that in Step 2, we now have

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{h}\|_2 = \|\boldsymbol{A}(\hat{\boldsymbol{x}} - \boldsymbol{x})\|_2 &= \|(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}) - (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\|_2 \\ &\leq \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}\|_2 + \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq 2\epsilon.\end{aligned}$$

Therefore, we need to bound

$$\begin{aligned}\|\boldsymbol{A}\boldsymbol{h}_{T_0\cup T_1}\|_2^2 &= \langle \boldsymbol{A}\boldsymbol{h} - \sum_{j\geq 2} \boldsymbol{A}\boldsymbol{h}_{T_j}, \boldsymbol{A}\boldsymbol{h}_{T_0\cup T_1}\rangle \\ &\leq \underbrace{\langle \boldsymbol{A}\boldsymbol{h}, \boldsymbol{A}\boldsymbol{h}_{T_0\cup T_1}\rangle}_{\leq 2\epsilon\delta_{2k}\|\boldsymbol{h}_{T_0\cup T_1}\|_2} \underbrace{- \sum_{j\geq 2} \langle \boldsymbol{A}\boldsymbol{h}_{T_j}, \boldsymbol{A}\boldsymbol{h}_{T_0\cup T_1}\rangle}_{\text{bounded as before}}\end{aligned}$$

By plugging in this modification, we show

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 = \|\boldsymbol{h}\|_2 \leq \frac{2(1+\rho)}{1-\rho} \frac{\|\boldsymbol{x} - \boldsymbol{x}_k\|_1}{\sqrt{k}} + \frac{2\alpha}{1-\rho}\epsilon,$$

where

$$\alpha = \frac{2\sqrt{1+\delta_{2k}}}{1-\delta_{2k}}.$$

Remarks

- The theorems are quite strong, in the sense it holds for *all* signals once $\boldsymbol{A}$ satisfies RIP.

- The reconstruction quality relies on two quantities: the best k-term approximation error and the noise level.

- Our generalization of the performance guarantee from the noise-free case to the noisy case is essentially effortless. However, we do need an upper bound of the noise level in order to perform the algorithm.

- A related algorithm is called LASSO, which has the form of

$$\hat{\boldsymbol{x}}_{lasso} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \ \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1,$$

  where $\lambda > 0$ is called a regularization parameter. Another related algorithm is called *Dantizg selector*. Both can be analyzed in a similar manner as the BP using RIP.

Which matrices satisfy RIP?

- Random matrices with i.i.d. Gaussian entries satisfy RIP with high probability, as long as
$$m \gtrsim k \log(n/k).$$
- Random Partial DFT matrices, $\boldsymbol{A} = \boldsymbol{I}_\Omega \boldsymbol{F}$, where $\boldsymbol{I}_\Omega$ is an partial identity matrix with rows indexed by the random subset Ω, and $\boldsymbol{F}$ is the DFT matrix, satisfy RIP with high probability, as long as
$$m = |\Omega| \gtrsim k \log^4 n.$$
- Similar results hold for random Partial Circulant/Toeplitz matrices, random matrices with i.i.d. sub-Gaussian entries, etc...
- All these are probabilistic, in the sense if we draw a random matrix following the stated distribution, it will satisfy the RIP with high probability (i.e. $1 - \exp(-cm)$).

Deterministic matrices satisfying RIP

Constructing deterministic matrices that satisfy RIP is difficult.

There're many benefits of having deterministic constructions: fast computation, less storage, etc..