Tutorial: Introduction to Compressed Sensing **IIT Delhi**
February 22, 2016
Lecturer: Prof. Mayank Bakshi *Scribes: Xishi Wang, Tongxin Li, Qiaoqiao Zhou, Qiaosheng Zhang, Zhibo Yang*

1 Basics of compressed sensing

1.1 Sparse Recovery

- Examples of sparse recovery problems
- Motivating applications
- How can sparsity be exploited?
  - Number of measurements/tests
- Formal definitions: *sparsity*, *approximate sparsity*, *measurement process*, *noise*

1.2 Review of mathematical basics

- Vector spaces, Linear Transformations, Rank, Nullity
- Normed spaces, L_p-norms
  - Sparse approximation of a vector

1.3 Compressive Sensing

- Linear measurement model
  $y = Ax$
- Need to solve under-determined linear equations
- Formulating Compressive Sensing as an L_0-minimization problem

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & Ax = y \end{aligned} \tag{1}$$

- Equivalent conditions for unique recoverability
- Construction of A as a real valued Vandermonde matrix
  - Why is this not a satisfactory solution? (*Hint:* stability with respect to noise)
- Computational intractibility of L_0-minimization

2 Basis Pursuit (L_1 Minimization)

2.1 Motivation

In compressive sensing, we want to search for the sparsest vector consistent with the measured data. The problem can be stated as L_0 minimization.

$$\text{minimize } \|\mathbf{z}\|_0 \qquad \text{subject to } \mathbf{Az} = \mathbf{y} \tag{P_0}$$

Unfortunately, solving the L_0 minimization (P_0) problem can be shown to be NP-hard. However, computationally efficient algorithms are well-developed to solve a relaxed version of the problem. One of these algorithms is basis pursuit, which can be stated as follows.

$$\text{minimize } \|\mathbf{z}\|_1 \qquad \text{subject to } \mathbf{Az} = \mathbf{y} \tag{P_1}$$

Fortunately, under certain conditions (e.g. null space property), solving P_1 can give us the same solutions as P_0.

2.2 Null Space Property

Definition 1. *A matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *is said to satisfy the null space property relative to a set* $S \subset [N]$ *if*

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\bar{S}}\|_1 \qquad \text{for all} \quad \mathbf{v} \in \ker\mathbf{A}\backslash\{\mathbf{0}\} \tag{1}$$

Theorem 2 (UR≡NSP)**.** *Given a matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *, every vector* $\mathbf{x} \in \mathbb{C}^N$ *supported on a set S is the unique solution of* (P_1) *with* $\mathbf{y} = \mathbf{Ax}$ *if and only if* $\mathbf{A}$ *satisfies the null space property relative to S.*

Proof.

UR⇒NSP. Let us assume $Supp(\mathbf{x}) \subseteq S$ and $\underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{z}\|_1 = \mathbf{x}$ s.t. $\mathbf{Az} = \mathbf{Ax}$. Thus, for any $\mathbf{v} \in \ker(\mathbf{A})\backslash\{\mathbf{0}\}$, the vector $\mathbf{v}_S$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $\mathbf{Az} = \mathbf{Av}_S$. We have $\mathbf{A}(\mathbf{v}_{\bar{S}} + \mathbf{v}_S) = \mathbf{Av} = 0$, which implies $\mathbf{A}(-\mathbf{v}_{\bar{S}}) = \mathbf{Av}_S$ and $-\mathbf{v}_{\bar{S}} \neq \mathbf{v}_S$. Hence, $\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\bar{S}}\|_1$.

UR⇐NSP. Let us assume the null space property relative to S holds. Then, suppose $Supp(\mathbf{x}) \subseteq S$ and $\mathbf{z} \neq \mathbf{x}$ s.t. $\mathbf{Az} = \mathbf{Ax}$. Let $\mathbf{v} := \mathbf{x} - \mathbf{z} \in \ker(\mathbf{A})\backslash\{\mathbf{0}\}$. We have

$$\begin{aligned} \|\mathbf{x}\|_1 &\leq \|\mathbf{x} - \mathbf{z}_S\|_1 + \|\mathbf{z}_S\| = \|\mathbf{v}_S\|_1 + \|\mathbf{z}_S\|_1 \\ &< \|\mathbf{v}_{\bar{S}}\|_1 + \|\mathbf{z}_S\|_1 = \|-\mathbf{z}_{\bar{S}}\|_1 + \|\mathbf{z}_S\|_1 = \|\mathbf{z}\|_1 \end{aligned}$$

□

3 Robustness and Stability

In practice, noise is added to the measurement vector, resulting in following the convex optimization problem

$$\text{minimize} \, \|\mathbf{z}\|_1 \qquad \text{subject to} \ \|\mathbf{Az} - \mathbf{y}\| \leq \eta \tag{$P_{1,\eta}$}$$

3.1 Robust null space property

Definition 3. *A matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *is said to satisfy the robust null space property of order s (with respect to* L_2 *norm) with constants* $0 < \rho < 1$ *and* $\tau > 0$ *if, for any set* $S \subset [N]$ *with* $card(S) \leq s$*,*

$$\|\mathbf{v}_S\|_1 < \rho \, \|\mathbf{v}_{\bar{S}}\|_1 + \tau \, \|\mathbf{Av}\|_2 \qquad \text{for all} \quad \mathbf{v} \in \mathbb{C}^N \tag{2}$$

Theorem 4. *The matrix* $\mathbf{A}$ *satisfies the robust null space property relative to S if and only if*

$$\|\mathbf{z} - \mathbf{x}\|_1 \leq \frac{1+\rho}{1-\rho}(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\,\|\mathbf{x}_{\bar{S}}\|_1) + \frac{2\tau}{1-\rho}\,\|\mathbf{A}(\mathbf{z}-\mathbf{x})\|_2 \tag{3}$$

for all vectors $\mathbf{x}$*,* $\mathbf{z} \in \mathbb{C}^N$*.*

3.2 Computational Efficiency

L_1 minimization problem (which belongs to *convex optimization problem*) can solved in polynomial time. In the noiseless case, L_1 minimization can be written as a *linear program*. Linear program can be shown to be solvable in polynomial time. In the noisy case, L_1 minimization can be written as a *second order cone program*, which is harder than linear program but still solvable in polynomial time. Hence, basis pursuit is a computationally efficient algorithm.

- Noiseless Basic Pursuit

  $$\text{minimize} \, \|\mathbf{z}\|_0 \qquad \text{subject to} \ \mathbf{Az} = \mathbf{y},$$

  Note that $\|\mathbf{z}\| = |z_1| + |z_2| + \ldots + |z_n|$. If we choose $t_1, t_2, \ldots, t_n$ such that $t_1 \geq |z_1|, t_2 \geq |z_2|, \ldots, t_n \geq |z_n|$, the above optimization will yield the same solution as the linear program

  $$\begin{aligned} &\text{minimize} \sum_{i=1}^{n} t_i \\ &\text{subject to} \qquad |z_1| \leq t_1 \\ &\qquad\qquad\qquad |z_2| \geq -t_2 \\ &\qquad\qquad\qquad \vdots \\ &\qquad\qquad\qquad |z_n| \geq -t_n \\ &\qquad\qquad\qquad \mathbf{Az} = \mathbf{y} \end{aligned}$$

- Noisy Basic Pursuit

  $$\text{minimize} \, \|\mathbf{z}\|_1 \qquad \text{subject to} \ \|\mathbf{Az} - \mathbf{y}\| \leq \eta$$

Similar to the noiseless case, for noisy basic pursuit, we have

$$\begin{aligned}
&\text{minimize} \sum_{i=1}^{n} t_i \\
&\text{subject to} \quad |z_1| \le t_1 \\
&\qquad |z_2| \ge -t_2 \\
&\qquad \vdots \\
&\qquad |z_n| \ge -t_n \\
&\qquad \|\mathbf{A}\mathbf{z} - \mathbf{y}\| \le \eta
\end{aligned}$$

This optimization problem is a second order cone program, which can be solved in polynomial time.

4 Restricted Isometry Property

Definition 5. *The sth restricted isometry constant* $\delta_s = \delta_s(\mathbf{A})$ *of a matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *is the smallest* $\delta \ge 0$ *such that*

$$(1-\delta)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le (1+\delta)\|\mathbf{x}\|_2^2 \tag{4}$$

for all s-sparse vectors $\mathbf{x} \in \mathbb{C}^N$.

$\mathbf{A}$ satisfies the restricted isometry property if δ_s is small for reasonably large s.

4.1 Sufficient Condition for NSP

Theorem 6. *Suppose that the* $2s$*th restricted isometry constant of the matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *satisfies* $\delta_{2s} < \frac{1}{3}$. *Then every s-sparse vector* $\mathbf{x} \in \mathbb{C}^N$ *is the unique solution of*

$$\text{minimize } \|\mathbf{z}\|_1 \qquad \textit{subject to } \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}. \tag{5}$$

Proof. We will show that RIP implies NSP of oder s, which is in the form:

$$\|\mathbf{v}_S\|_1 < \frac{1}{2}\|\mathbf{v}\|_1 \ \forall\ \mathbf{v} \in \ker\mathbf{A}\backslash\{\mathbf{0}\} \text{ and } \forall\ S \text{ s.t. } \operatorname{card}(S) = s. \tag{6}$$

This will follow from the stronger statement

$$\|\mathbf{v}_S\|_2 < \frac{\rho}{2\sqrt{s}}\|\mathbf{v}\|_1 \ \forall\ \mathbf{v} \in \ker\mathbf{A}\backslash\{\mathbf{0}\} \text{ and } \forall\ S \text{ s.t. } \operatorname{card}(S) = s, \tag{7}$$

where

$$\rho := \frac{2\delta_{2s}}{1-\delta_{2s}} < 1. \tag{8}$$

Given $\mathbf{v} \in \ker\mathbf{A}$, we let S_0 be the first s largest absolute entries of the vector $\mathbf{v}$ and S_1 be the second s largest absolute entries of $\mathbf{v}$ etc. Since $\mathbf{A}\mathbf{v} = \mathbf{0}, \mathbf{A}(\mathbf{v}_{S_0} = \mathbf{A}(-\mathbf{v}_{S_1} - \mathbf{v}_{S_2} + \cdots)$. Hence, we have

$$\begin{aligned}
&\|\mathbf{v}_{S_0}\|_2^2 && (9)\\
&\le \frac{1}{1-\delta_{2s}}\|\mathbf{A}(\mathbf{v}_{S_0})\|_2^2 && (10)\\
&= \frac{1}{1-\delta_{2s}}\langle \mathbf{A}(\mathbf{v}_{S_0}), \mathbf{A}(-\mathbf{v}_{S_1}) + \mathbf{A}(-\mathbf{v}_{S_2}) + \cdots\rangle \\
&= \frac{1}{1-\delta_{2s}}\sum_{k\ge1}\langle \mathbf{A}(\mathbf{v}_{S_0}), \mathbf{A}(-\mathbf{v}_{S_k})\rangle && (11)\\
&\le \frac{\delta}{1-\delta}\sum_{k\ge1}\|\mathbf{v}_{S_0}\|_2\|\mathbf{v}_{S_k}\|_2 && (12)\\
&= \frac{\rho}{2}\sum_{k\ge1}\|\mathbf{v}_{S_0}\|_2\|\mathbf{v}_{S_k}\|_2 && (13)\\
& && (14)
\end{aligned}$$

Since

$$\|\mathbf{v}_{S_k}\|_2 \le \frac{1}{\sqrt{s}}\|\mathbf{v}_{S_{k-1}}\|_1, \tag{15}$$

we derive

$$\|\mathbf{v}_{S_0}\|_2 \le \frac{\rho}{2\sqrt{s}}\sum_{k\ge1}\|\mathbf{v}_{S_{k-1}}\|_1 \le \frac{\rho}{2\sqrt{s}}\|\mathbf{v}\|_1. \tag{16}$$

This is the desired inequality. □

4.2 Random Matrix Construction

Testing whether a matrix is satisfies RIP is NP-hard in the worst case [1]. Fortunately, a random constructed matrix satisfies RIP with high probability as long as $m > C_\delta s \log\frac{N}{S}$. Below list the theorems RIP for different types of random matrices without proofs.

Theorem 7. *Let* $\mathbf{A}$ *be an* $m \times N$ *subgaussian random matrix. Then there exists a constant* $C > 0$ *(depending only on the subgaussian parameters* β, κ*) such that the restricted isometry constant of* $\frac{1}{\sqrt{m}}\mathbf{A}$ *satisfies* $\delta_s \leq \delta$ *with probability at least* $1 - \varepsilon$ *provided*

$$m \geq C\delta^{-2}(s\ln(eN/s) + \ln(2\varepsilon^{-1})) \tag{17}$$

Theorem 8. *Let* $\mathbf{A}$ *be an* $m \times N$ *Gaussian or Bernoulli random matrix. Then there exists a universal constant* $C > 0$ *such that the restricted isometry constant of* $\frac{1}{\sqrt{m}}\mathbf{A}$ *satisfies* $\delta_s \leq \delta$ *with probability at least* $1 - \varepsilon$ *provided*

$$m \geq C\delta^{-2}(s\ln(eN/s) + \ln(2\varepsilon^{-1})) \tag{18}$$

5 Orthogonal Matching Pursuit Algorithm

Remember in our second lecture, finding the sparest solution $\mathbf{x}$ which is feasible to sensing constraints given by an encoding matrix $\mathbf{A}$ and observation vector $\mathbf{b}$ can be summarized as the following L_0 minimization problem (which we call P_0).

$$\begin{aligned} &\text{minimize } ||\mathbf{x}||_0 \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned} \tag{P_0}$$

The problem above is hard to solve in general. But we can still get some inspiration by considering the following simplest scenario.

Suppose one Genie from the story of Aladdin revealed us the unique solution $\mathbf{x}$ except for one entry in it. Now we are required to resolve this remaining entry without knowing its location and value. In this example, it suffices to assume $\mathbf{b}$ is a scalar multiple of some column $\mathbf{a}_j$ of the matrix $\mathbf{A}$, viz, $x_j\mathbf{a}_j = \mathbf{b}$ for some constant x_j.

Instead of relaxing P_0 to a L_1 minimization problem as we did in previous lecture, we consider a different approach by updating the solution from the Genie in a greedy manner. That is, we minimize the error term in Euclidean distance $\varepsilon(j) = \min_{x_j} ||x_j\mathbf{a}_j - \mathbf{b}||_2$ by setting the minimizer $x_j^* = \frac{\mathbf{a}_j^*\mathbf{b}}{||\mathbf{a}_j||_2^2}$. If the resulting expression $\varepsilon(j) = ||\frac{\mathbf{a}_j^*\mathbf{b}\mathbf{a}_j}{||\mathbf{a}_j||_2^2} - \mathbf{b}||_2^2$ is zero for some j, we shall conclude that $\mathbf{x} = [0, 0, 0, \cdots, x_j, \cdots, 0, 0, 0]^T$.

[1]This property explains why the algorithm is called Orthogonal Matching Pursuit.

Now, in a more general case, denote $\mathcal{S}$ the support set of $\mathbf{x}$ and suppose the Genie already resolved the locations of i out of $|\mathcal{S}|$ many non-zero entries in $\mathbf{x}$. For example, the Genie can select a subset $\hat{\mathcal{S}} \subseteq \mathcal{S}$ with cardinality $|\hat{\mathcal{S}}| = i < |\mathcal{S}|$ and reveal this subset to us. Based on the subset $\hat{\mathcal{S}}$ we can find a approximation $\hat{\mathbf{x}}$ for $\mathbf{x}$ with $\text{supp}(\hat{\mathbf{x}}) \subseteq \hat{\mathcal{S}}$ by minimizing the Euclidean distance between $\mathbf{A}\hat{\mathbf{x}}$ and $\mathbf{b}$. To be more precise, let $\hat{\mathbf{x}} = \underset{\mathbf{x}:\text{supp}(\mathbf{x}) \subseteq \hat{\mathcal{S}}}{\text{argmin}} ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2$ and we can calculate a corresponding residual $\hat{\mathbf{r}} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$.

Then we can mimic the approach that we used to find the single entry above as follows. Firstly we start by finding a column $\mathbf{a}^j$ outside the "Genie subset" $\hat{\mathcal{S}}$ in the matrix $\mathbf{A}$ which has the most "similarities" with the observation vector $\mathbf{b}$. Using Euclidean distance as the metric, the above can be done by plugging a minimizer $x_j^* = \frac{\mathbf{a}_j^*\hat{\mathbf{r}}}{||\mathbf{a}_j||_2^2}$ to a similar error term $\varepsilon(j) = \min_{x_j, j \in \{1,2,\ldots,\text{length}(\mathbf{x})\} \setminus \hat{\mathcal{S}}} ||x_j\mathbf{a}_j - \hat{\mathbf{r}}||_2$ as the single-entry case. We have the following simplification.

$$\min_{x_j, j \in \{1,2,\ldots,\text{length}(\mathbf{x})\} \setminus \hat{\mathcal{S}}} \varepsilon(j)$$

$$= \min_{x_j, j \in \{1,2,\ldots,\text{length}(\mathbf{x})\} \setminus \hat{\mathcal{S}}} ||x_j\mathbf{a}_j - \hat{\mathbf{r}}||_2 \tag{19}$$

$$= \min_{j \in \{1,2,\ldots,\text{length}(\mathbf{x})\} \setminus \hat{\mathcal{S}}} ||\frac{\mathbf{a}_j^*\hat{\mathbf{r}}\mathbf{a}_j}{||\mathbf{a}_j||_2^2} - \hat{\mathbf{r}}||_2^2 \tag{20}$$

$$= \min_j ||\hat{\mathbf{r}}||_2^2 - 2\frac{\left(\mathbf{a}_j^*\hat{\mathbf{r}}\right)^2}{||\mathbf{a}_j||_2^2} + \frac{\left(\mathbf{a}_j^*\hat{\mathbf{r}}\right)^2}{||\mathbf{a}_j||_2^2} \tag{21}$$

$$= \min_j ||\hat{\mathbf{r}}||_2^2 - \frac{\left(\mathbf{a}_j^*\hat{\mathbf{r}}\right)^2}{||\mathbf{a}_j||_2^2} \tag{22}$$

$$= \max_j \frac{|\langle\hat{\mathbf{r}}, \mathbf{a}_j\rangle|}{||\mathbf{a}_j||_2}. \tag{23}$$

where (19) comes from plugging the minimizer $x_j^* = \frac{\mathbf{a}_j^*\hat{\mathbf{r}}}{||\mathbf{a}_j||_2^2}$ to (19). Moreover (21) holds since $\hat{\mathbf{r}}$ is a minimizer of $||\mathbf{A}_{\hat{\mathcal{S}}}\mathbf{x}_{\hat{\mathcal{S}}} - \mathbf{b}||_2$ where $\mathbf{z}_{\hat{\mathcal{S}}}$ is the non-zero portion of the vector $\mathbf{x}$ based on the "Genie subset" $\hat{\mathcal{S}}$. Then the optimized $\hat{\mathbf{r}}$ is given by setting the derivative of this quadratic form be zero. In detail, $\nabla||\mathbf{A}_{\hat{\mathcal{S}}}\mathbf{x}_{\hat{\mathcal{S}}} - \mathbf{b}||_2 = 2\mathbf{A}_{\hat{\mathcal{S}}}{}^*\left(\mathbf{A}_{\hat{\mathcal{S}}}\mathbf{x}_{\hat{\mathcal{S}}} - \mathbf{b}\right) = -2\mathbf{A}_{\hat{\mathcal{S}}}{}^*\hat{\mathbf{r}} = 0$. The relation above suggests that the part of columns in $\mathbf{A}$ according to the support set $\hat{\mathcal{S}}$ are orthogonal[1] to the residual $\hat{\mathbf{r}}$. This in turn implies that if we relax the constraint in (1) such that index j can be chosen from the set $\{1, 2, \ldots, \text{length}(\mathbf{x})\}$, we will still get the same answer.

The above argument inspires us to find an index j by maximizing the inner product of the residual $\hat{\mathbf{r}}$ and columns in $\mathbf{A}$ if those columns are L_2-normalized. Then we add that index j to the "Genie subset" $\hat{\mathcal{S}}$ to result a larger support set. We start with setting $\hat{\mathcal{S}}$ as an empty set and repeating the iterations over and over again, which suggests the following greedy algorithm with a preassigned error threshold[2] ε_0.

5.1 Least Square Estimator

Note that the step $\mathbf{x}^{i+1} = \underset{\mathbf{x}:\text{supp}(\mathbf{x})\subseteq\mathcal{S}^{i+1}}{\text{argmin}} ||\mathbf{A}\mathbf{x}-\mathbf{b}||_2$ in above algorithm is exhaustive if we implement it directly. Hence we adopt a simpler expression for this step as the below lemma states.

Lemma 9. *In the i-th iteration* $(i \geq 1)$ *we have* $\mathbf{x}^i_{\mathcal{S}^i} = \mathbf{A}^{\dagger}_{\mathcal{S}^i}\mathbf{b}$ *where* $\mathbf{x}^i_{\mathcal{S}^i}$ *denotes the restriction of* $\mathbf{x}^i$ *to its support set* $\mathcal{S}^i$ *and where* $\dagger$ *is the operator for pseudo-inversion.*

Proof. This lemma simply says that $\mathbf{z} = \mathbf{x}^i_{\mathcal{S}^i}$ is a solution of the equation $\mathbf{A}^*_{\mathcal{S}^i}\mathbf{A}_{\mathcal{S}^i}\mathbf{z} = \mathbf{A}^*_{\mathcal{S}^i}\mathbf{b}$. We already proved this equation must be true for $\mathbf{z} = \mathbf{x}^i_{\mathcal{S}^i}$ since the gradient of $||\mathbf{A}_{\mathcal{S}^i}\mathbf{x}_{\mathcal{S}^i} - \mathbf{b}||_2$ need to be zero, from which the orthogonality condition follows. □

5.2 Performance Guarantee

Note that we mentioned before that if $\mathbf{x}$ is exactly s-sparse and there is no noise, then OMP terminates in the s-step with the correct output $\mathbf{x}$ or an error. This is a direct result since each time we pick a unique index j and then $|\mathcal{S}^i| = i$ means s steps are enough for decoding. Therefore one may ask under which condition we can guarantee OMP to succeed with s iterations. The following theorem gives the sufficient and necessary conditions.

5.2.1 Exact Recovery Conditions

Theorem 10 (Exact Recovery Conditions)**.** *Given a matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$*, every nonzero vector* $\mathbf{x} \in \mathbb{C}^N$ *supported on a set* $\mathcal{S}$ *of size s is successfully recovered from OMP after at most s iterations if and only if the following conditions hold.*

1. $\mathbf{A}_{\mathcal{S}}$ *is injective.*

2. $\max_{j\in\mathcal{S}} \frac{|\langle \mathbf{r},\mathbf{a}_j\rangle|}{||\mathbf{a}_j||_2} > \max_{j\in\mathcal{S}^c} \frac{|\langle \mathbf{r},\mathbf{a}_l\rangle|}{||\mathbf{a}_l||_2}$ *for all nonzero* $\mathbf{r} \in \{\mathbf{A}\mathbf{z}, \text{supp}(\mathbf{z}) \subseteq \mathcal{S}\}$.

Proof. $\Longrightarrow$ Suppose OMP succeeds after s iterations. Then any $l \in \mathcal{S}^c$ can not be chosen in the first iteration, which in turn implies the second condition that $\max_{j\in\mathcal{S}} \frac{|\langle \mathbf{r},\mathbf{a}_j\rangle|}{||\mathbf{a}_j||_2} > \max_{j\in\mathcal{S}^c} \frac{|\langle \mathbf{r},\mathbf{a}_l\rangle|}{||\mathbf{a}_l||_2}$ for all nonzero $\mathbf{r} \in \{\mathbf{A}\mathbf{z}, \text{supp}(\mathbf{z}) \subseteq \mathcal{S}\}$ holds. Moreover, since two vectors supported on $\mathcal{S}$ which have the same corresponding measurement vector $\mathbf{b}$ must be equal, the matrix $\mathbf{A}_{\mathcal{S}}$ is injective.

$\Longleftarrow$ To prove the sufficiency, we are going to prove that the set $\mathcal{S}^i$ is always a subset of $\mathcal{S}$ for all $0 \leq i \leq s$. This will imply $\mathcal{S}^s = \mathcal{S}$; hence as a result $\mathbf{A}\mathbf{x}^s = \mathbf{b}$ and because of the injectivity of $\mathbf{A}_{\mathcal{S}}$, we conclude that $\mathbf{x}^s = \mathbf{x}$. The claim above can be shown by induction. Suppose up to the n-th step with $0 \leq n \leq s-1$ the support recovery is correct. We notice that at any n-th step, induction hypothesis gives $\mathcal{S}^n \subseteq \mathcal{S}$, which yields $\mathbf{r}^n = \mathbf{b} - \mathbf{A}\mathbf{x}^n \in \{\mathbf{A}\mathbf{z}, \text{supp}(\mathbf{z}) \subseteq \mathcal{S}\}$. Hence by the second condition the index j^{n+1} lies in $\mathcal{S}$ as well. Therefore we get $\mathcal{S}^{n+1} \subseteq \mathcal{S}$. Noticing the fact that the cardinality of $\mathcal{S}^i$ satisfies $|\mathcal{S}^i| = i$ finishes our proof. □

We have already seen the *exact recovery condition* for OMP in previous context. One may still wonder if the standard *restricted isometry property* (R.I.P.) isenough to guarantee the recovery of all s-sparse vectors within at most s iterations using OMP. A counter-example is presented below, in which we construct a matrix $\mathbf{A}$ that satisfies R.I.P but OMP will fail at the first step.

5.2.2 R.I.P is not sufficient

Example 11. *Our recovery goal is standard. Given any s-sparse vector, we expect a matrix* $\mathbf{A}$ *which staisfies R.I.P but OMP can not output the correct answer using s iterations.*

[2] Later we will see that if $\mathbf{x}$ is exactly s-sparse and there is no noise, OMP either stops in exactly s steps with the correct solution or gives a wrong solution after more than $s+1$ steps. Hence if s is known *a priori*, this can be used as a natural stopping rule. For approximate sparsity and presence of noise, stopping rules based on convergence of the residual are typically used. In Algorithm 1 we use the vanilla version such that after each iteration we compare the error threshold with L_2 norm of the residual.

Consider constructing a $(s+1)\times(s+1)$ L_2 normalized matrix in the following way. Fix $1<\eta<\sqrt{s}$, the matrix $\mathbf{A}$ *is defined by*

$$\mathbf{A}=\left[\begin{array}{c|c} \mathbf{Id} & \begin{matrix}\frac{\eta}{s}\\ \vdots \\ \frac{\eta}{s}\end{matrix} \\ \hline 0\cdots 0 & \sqrt{\frac{s-\eta^2}{s}}\end{array}\right]. \tag{3}$$

From (3) *we calculate that*

$$\mathbf{A}^*\mathbf{A}-\mathbf{Id}=\left[\begin{array}{c|c} \mathbf{0} & \begin{matrix}\frac{\eta}{s}\\ \vdots \\ \frac{\eta}{s}\end{matrix} \\ \hline \frac{\eta}{s}\cdots\frac{\eta}{s} & 0\end{array}\right]. \tag{4}$$

The matrix above in (4) *has eigenvalues* $\frac{-\eta}{\sqrt{s}}$, $\frac{\eta}{\sqrt{s}}$ *and* 0 *with multiplicity* $s-1$. *Thus the restricted isometry constant*[3] δ_{s+1} *satisfies*

$$\delta_{s+1}=||\mathbf{A}^*\mathbf{A}-\mathbf{Id}||_{2\to 2}=\frac{\eta}{\sqrt{s}}<1.$$

However, the s-sparse vector $\mathbf{x}=[1,\ldots,1,0]^T$ *can not be recovered from* $\mathbf{b}=\mathbf{A}\mathbf{x}$ *after s iterations, since OMP even fails at the first iteration, i.e., the wrong index (the last one) is picked at the first iteration. To see this, notice that*

$$\mathbf{A}^*\left(\mathbf{b}-\mathbf{A}\mathbf{x}^0\right)=\mathbf{A}^*\mathbf{A}\mathbf{x}=\left[\begin{array}{c|c} \mathbf{Id} & \begin{matrix}\frac{\eta}{s}\\ \vdots \\ \frac{\eta}{s}\end{matrix} \\ \hline 0\cdots 0 & \sqrt{\frac{s-\eta^2}{s}}\end{array}\right]\left[\begin{array}{c}1\\ \vdots\\ 1\\ \hline 0\end{array}\right]=\left[\begin{array}{c}1\\ \vdots\\ 1\\ \hline \eta\end{array}\right].$$

Since $\eta>1$, *we get the first index* $j^1=s+1$, *which gives the incorrect answer.*

[3]The restricted isometry constant is the infimum of all δ for a given matrix to satisfy R.I.P. This comes from the fact that for any matrix satisfies R.I.P with restricted isometry constant δ_k, we have the following relation —- $1-\delta_k\le\lambda_{\min}(\mathbf{A}^*\mathbf{A})\le\lambda_{\max}(\mathbf{A}^*\mathbf{A})\le 1+\delta_k$.

5.2.3 Mutual Coherence

Instead of using the exact recovery conditions in Theorem 3.2 (which is in general hard to check for every possible support set), we can relax ourselves to sufficient conditions that are relatively easy to check. We know that R.I.P is not sufficient for OMP to work., this brings up the question that if we can find an efficient way to check sufficient conditions for OMP. It turns our the *mutual coherence* defined below becomes useful.

Definition 12 (Mutual Coherence)**.** *The mutual coherence of a given matrix* $\mathbf{A}\in\mathbb{C}^{m\times N}$ *is the largest absolute normalized inner product between different columns from* $\mathbf{A}$. *It is defined by* $\mu(\mathbf{A})=\max_{1\le i,j\le N, i\ne j}\frac{|\langle\mathbf{a}_i,\mathbf{a}_j\rangle|}{||\mathbf{a}_i||_2\cdot||\mathbf{a}_j||_2}$.

Based on the definition above which characterizes the dependency between columns of the matrix $\mathbf{A}$, we have the following sufficient conditions by Elad [RM13].

Theorem 13 (Sufficient Recovery Condition)**.** *Given a full-rank matrix* $\mathbf{A}\in\mathbb{C}^{m\times N}$ *with* $(m<N)$, *if a solution* $\mathbf{x}$ *of the linear system* $\mathbf{A}\mathbf{x}=\mathbf{b}$ *exists obeying*

$$||\mathbf{x}||_0<\frac{1}{2}\left(1+\frac{1}{\mu(\mathbf{A})}\right), \tag{5}$$

then OMP runs with threshold parameter $\varepsilon_0=0$ *is guaranteed to give* $\mathbf{x}$ *exactly.*

Remark 14. *Once the solution* $\mathbf{x}$ *exists and satisfies the condition in* (5), *it is necessarily the sparsest (the unique) solution.*

Proof. Suppose that without loss of generality, the sparest solution is such that all its s nonzero entries are at the beginning of the vector, in decreasing order of the values $|x_j|$. Thus we have

$$\mathbf{b}=\mathbf{A}\mathbf{x}=\sum_{i=1}^{s}x_i\mathbf{a}_i. \tag{6}$$

Note that in the first step for selecting the index j^1, we must require the following conditions for OMP to succeed.

$$|\mathbf{a}_1^*\mathbf{b}|>|\mathbf{a}_k^*\mathbf{b}| \text{ for all } k>s. \tag{7}$$

Combining (6) and (7) we get for all $k>s$

$$|\sum_{i=1}^{s}x_i\mathbf{a}_1^*\mathbf{a}_i|>|\sum_{i=1}^{s}x_i\mathbf{a}_k^*\mathbf{a}_i|. \tag{*}$$

Our goal is to find a sufficient condition in term of $\mu(\mathbf{A})$ that the inequality $(*)$ above holds. The LHS of $(*)$ can be bounded from below as follows.

$$
\begin{aligned}
|\sum_{i=1}^{s} x_i \mathbf{a}_1^* \mathbf{a}_i| &\geq |x_1| - \sum_{i=2}^{s} |x_i| \cdot |\mathbf{a}_1^* \mathbf{a}_i \\
&\geq |x_1| - \sum_{i=2}^{s} |x_i| \cdot \mu(\mathbf{A}) \\
&\geq |x_1|(1 - \mu(\mathbf{A})(s-1)) \qquad (8)
\end{aligned}
$$

where (8) comes from that x_i are ranked as the descending order.

On the other hand, the RHS of $(*)$ can also be bounded above by

$$
\begin{aligned}
|\sum_{i=1}^{s} x_i \mathbf{a}_k^* \mathbf{a}_i| &\leq \sum_{i=1}^{s} |x_i| \cdot |\mathbf{a}_k^* \mathbf{a}_i| \\
&\leq \sum_{i=1}^{s} |x_i| \cdot \mu(\mathbf{A}) \\
&\leq |x_1| \cdot \mu(\mathbf{A}) s. \qquad (9)
\end{aligned}
$$

The two bounds we get from (8) and (9) lead to a sufficient inequality allowing us to derive $(*)$ from it.

$$|x_1|(1 - \mu(\mathbf{A})(s-1)) > |x_1| \cdot \mu(\mathbf{A}) s, \qquad (9)$$

which yields $1 + \mu(\mathbf{A}) > 2\mu(\mathbf{A})s$ and hence $||x||_0 = s < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{A})}\right)$ is a sufficient condition we need for the first step.

Once the first iteration succeeds, the remaining steps follow because the residual is also also a linear combination of at most s columns in $\mathbf{A}$. Thus the condition in (5) guarantees that for all $k > s$, $|\sum_{i=1}^{s} x_i \mathbf{a}_2^* \mathbf{a}_i| > |\sum_{i=1}^{s} x_i \mathbf{a}_k^* \mathbf{a}_i|$ at the second step and so on. By orthogonality of OMP, the same index will never be chosen again hence after s iterations, the residual term becomes zero and the correct solution is recovered. □

5.2.4 Random Measurements

We know that R.I.P. is not sufficient for the success of OMP. However, if we restrict more on the restricted isometry constant δ_s and the number of measurements m, similarly as Basis Pursuit, random construction on the encoding matrix $\mathbf{A}$ works for OMP (asymptotically). We have the following theorem without proving it.

Theorem 15 (Random Measurements)**.** *Let* $\delta_s \in (0, 0.36)$ *and* $m \geq C \cdot s \log(\frac{N}{s})$ *for some constant* $C > 0$. *Fix a s-sparse vector* $\mathbf{x}$. *Given any matrix* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *satisfies the following conditions*

1. *(M0). Independence: Two columns* $\mathbf{a}_i$ *and* $\mathbf{a}_j$ *are statistically independent for all* $i \neq j$.
2. *(M1). Normalization:* $\mathbb{E}||\mathbf{a}_j||_2^2 = 1$ *for all* $j = 1, 2, \ldots, N$.
3. *(M2). Joint Correlation: Let* $\{\mathbf{x}_i\}$ *be a sequence of length-N vectors whose* L_2 *norm do not exceed one. Let* $\mathbf{a}_j$ *be a column of* $\mathbf{A}$ *which is independent from this sequence. Then* $\Pr[\max_j |\langle \mathbf{a}_j, \mathbf{x}_i \rangle| \leq \varepsilon] \geq 1 - 2s \cdot e^{-c\varepsilon^2 m}$.
4. *(M3). Smallest singular value: For a given* $m \times N$ *submatrix* $\mathbf{B}$ *from* $\mathbf{A}$, *the s-th largest singular value* $\sigma_s(\mathbf{B})$ *satisfies*

$$\Pr[\sigma_s(\mathbf{B}) \geq 0.5] \geq 1 - e^{c \cdot m},$$

we have OMP succeeds with high probability $\Pr_s > 1 - \delta_s$.

References

[1] A.S. Bandeira, E. Dobriban, D.G. Mixon, W.F. Sawin, "Certifying the restricted isometry property is hard," *arXiv eprint arXiv:1204.1580 [math.FA]*, 2012.

[TG07] TROPP, JOEL and GILBERT, ANNA C, "Signal recovery from random measurements via orthogonal matching pursuit", *IEEE Trans. Inf. theory*, vol. 53, no. 12, pp. 4655-4666, 2007.

[RM13] GIRYES, RAJA and ELAD, MICHAEL, "OMP with Highly Coherent Dictionaries", *10th Int. Conf. on Sampling Theory Appl*, 2013