On the Recovery Limit of Sparse Signal Using Orthogonal Matching Pursuit

Authors: Jian Wang and Byonghyo Shim

Publication: IEEE T. Signal Processing, SEPT.2012
Speaker: Sangjun Park

Short summary: In the paper, the authors give a sufficient condition of the Orthogonal Matching Pursuit (OMP) algorithm. In [2], Wakin and Davenport insisted that OMP can reconstruct any K sparse signal if $\delta_{K+1} < 1/\left(3\sqrt{K}\right)$, where δ_K is the restricted isometry constant. However, in this talk, an improved sufficient condition that guarantees the perfect recovery of OMP is presented

I. History of sufficient conditions of the OMP algorithm

In the below table 1, sufficient conditions that the OMP algorithm reconstructs a K spars signal from a set of linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \Re^{M \times N}$ ($N > M$), are given.

Year	A sufficient condition
2007[1]	$\mu < 1/\left(2K-1\right)$
2010[2]	$\delta_{K+1} < 1/\left(3\sqrt{K}\right)$

Besides, there are many theoretical papers which analyze algorithms based on the OMP algorithm. In here, it is not scope of this seminar. Therefore, we do not care about them.

II. System Model

Let us consider the below equation:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{A} \in \Re^{M \times N}$ ($N > M$), and $\mathbf{x} \in \Re^{N}$ is a K sparse signal, and $\mathbf{y} \in \Re^{M}$ is a set of linear measurements. The smallest constant δ_K called “the restricted isometry constant” satisfies

$$\left(1-\delta_K\right)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le \left(1+\delta_K\right)\|\mathbf{x}\|_2^2 \tag{2}$$

for any K sparse signal $\mathbf{x}$.

III. MAIN RESULTS

A. Improved Recovery Bound of the OMP algorithm

Theorem 1: For any K sparse signal $\mathbf{x}$, the OMP algorithm perfectly reconstruct $\mathbf{x}$ from $\mathbf{y}$ if the isometry constant δ_{K+1} satisfies

$$\delta_{K+1} < \frac{1}{\sqrt{K}+1}. \tag{3}$$

In this talk, we try to understand a proof of Theorem 1.

Before we study the proof, let us consider whether the OMP algorithm perfectly reconstructs $\mathbf{x}$ or not if $\delta_{K+1} = 1/\sqrt{K}$.

B. The OMP algorithm can fail under $\delta_{K+1} = 1/\sqrt{K}$.

Example 1: Let us consider the problem of reconstructing a K sparse signal $\mathbf{x} \in \Re^{K+1}$ such as $x_{K+1} = 0$, and $x_i = 1$ for $i = 1, \cdots, K$ from $\mathbf{y} = \mathbf{A}\mathbf{x}$, where

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & b & \cdots & b \\ b & 1 & & \vdots \\ \vdots & & \ddots & b \\ b & \cdots & b & 1 \end{bmatrix} \in \Re^{(K+1)\times(K+1)}.$$

Obviously, all the Eigen values of $\mathbf{A}^T\mathbf{A}$ are $\lambda_1 = \lambda_2 = \cdots = \lambda_K = 1-b$, and $\lambda_{K+1} = 1+Kb$. (See Example 1 on Appendix). When we assume $b = -1/\left(K\sqrt{K}\right)$, $\mathbf{A}^T\mathbf{A}$ becomes

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & -1/\left(K\sqrt{K}\right) & \cdots & -1/\left(K\sqrt{K}\right) \\ -1/\left(K\sqrt{K}\right) & 1 & & \vdots \\ \vdots & & \ddots & -1/\left(K\sqrt{K}\right) \\ -1/\left(K\sqrt{K}\right) & \cdots & -1/\left(K\sqrt{K}\right) & 1 \end{bmatrix} \in \Re^{(K+1)\times(K+1)}, \tag{4}$$

and the smallest and biggest Eigen values are

$$\lambda_{\min} = 1-1/\sqrt{K}, \text{ and } \lambda_{\max} = 1+1/\left(K\sqrt{K}\right).$$

Therefore, we have $\delta_{K+1} = 1/\sqrt{K}$ (In fact, all the Eigen values of $\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}$ must be contained in the interval $\left[1-\delta_{|\mathcal{S}|}, 1+\delta_{|\mathcal{S}|}\right]$, Thus, $\delta_{K+1} = \max\left\{\lambda_{\max}\left(\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}\right)-1, 1-\lambda_{\min}\left(\mathbf{A}_{\mathcal{S}}^T\mathbf{A}_{\mathcal{S}}\right)\right\}$). Now, we investigate a quantity $\left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right|$ for $i = 1, \cdots, K+1$. For the OMP algorithm to reconstruct $\mathbf{x}$, $\left|\langle \mathbf{a}_{K+1}, \mathbf{y}\rangle\right|$ must be less than any $\left|\langle \mathbf{a}_i, \mathbf{y}\rangle\right|$ for

$i=1,\cdots,K$. This is reason that we investigate the quantities. First, for $i\in\{1,\cdots,K\}$, we have

$$\begin{aligned}\left|\langle \mathbf{a}_i,\mathbf{y}\rangle\right| &\overset{(a)}{=}\left|\langle \mathbf{a}_i,\mathbf{Ax}\rangle\right|\\ &\overset{(b)}{=}\left|\langle \mathbf{A}^T\mathbf{a}_i,\mathbf{x}\rangle\right|\\ &\overset{(c)}{=}1-\frac{K-1}{K\sqrt{K}},\end{aligned} \tag{5}$$

where (a) from the fact $\mathbf{y}=\mathbf{Ax}$, (b) from the fact $\langle \mathbf{a}_i,\mathbf{Ax}\rangle=\mathbf{a}_i^T\mathbf{Ax}=\mathbf{x}^T\mathbf{A}^T\mathbf{a}_i=\langle \mathbf{x},\mathbf{A}^T\mathbf{a}_i\rangle$, and (c) from the fact that $\mathbf{A}^T\mathbf{a}_i$ is the i^{th} column of $\mathbf{A}^T\mathbf{A}$ presented in (4), and $\mathbf{x}$ such as $x_{K+1}=0$, and $x_i=1$ for $i=1,\cdots,K$. Second, for $i=K+1$, we have

$$\begin{aligned}\left|\langle \mathbf{a}_{K+1},\mathbf{y}\rangle\right| &=\left|\langle \mathbf{a}_{K+1},\mathbf{Ax}\rangle\right|\\ &=\left|\langle \mathbf{A}^T\mathbf{a}_{K+1},\mathbf{x}\rangle\right|\\ &=\frac{1}{\sqrt{K}}.\end{aligned} \tag{6}$$

Obviously, the OMP algorithm must fail in the first iteration if an inequality $\left|\langle \mathbf{a}_{K+1},\mathbf{y}\rangle\right|\geq\left|\langle \mathbf{a}_i,\mathbf{y}\rangle\right|$ for all $i\in\{1,\cdots,K\}$. The inequity becomes

$$\frac{1}{\sqrt{K}}\geq 1-\frac{K-1}{K\sqrt{K}}$$

which is always true if $K=2$. Thus, the OMP algorithm in the first iteration selects an incorrect index.

IV. Proof of Theorem 1

A. Notations

The below notations will be used throughout the rest of this presentation. $\mathcal{T}=\operatorname{supp}(\mathbf{x}):=\{i|x_i\neq 0\}$ is the set of indices corresponding to non-zero coefficients of $\mathbf{x}$. $|\mathcal{T}|$ is the cardinality of $\mathcal{T}$, and $\mathcal{T}\setminus\mathcal{I}$ is the set of elements belonging to $\mathcal{T}$ but not to $\mathcal{I}$. $\mathbf{A}_{\mathcal{T}}\in\Re^{M\times|\mathcal{T}|}$ is a sub-matrix of $\mathbf{A}$ which contains columns corresponding to indices of $\mathcal{T}$. $\mathbf{x}_{\mathcal{T}}\in\Re^{|\mathcal{T}|}$ is a restriction of $\mathbf{x}$ to the elements indexed by $\mathcal{T}$. $span(\mathbf{A}_{\mathcal{T}})$ is the span of columns in $\mathbf{A}_{\mathcal{T}}$, $\mathbf{A}_{\mathcal{T}}^T$ is the transpose of $\mathbf{A}_{\mathcal{T}}$, and $\mathbf{A}_{\mathcal{T}}^{\dagger}=\left(\mathbf{A}_{\mathcal{T}}^T\mathbf{A}_{\mathcal{T}}\right)^{-1}\mathbf{A}_{\mathcal{T}}^T$ is the pseudo inverse of $\mathbf{A}_{\mathcal{T}}$.

$\mathbf{P}_{\mathcal{T}} = \mathbf{A}_{\mathcal{T}}\mathbf{A}_{\mathcal{T}}^{\dagger}$ is the orthogonal projection onto $span(\mathbf{A}_{\mathcal{T}})$, and $\mathbf{P}_{\mathcal{T}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathcal{T}}$ is the orthogonal projection onto the orthogonal complement of $span(\mathbf{A}_{\mathcal{T}})$.

B. Lemmas

We need the below lemmas to prove Theorem 1.

Lemma 1: For a set $\mathcal{I}$, if $\delta_{|\mathcal{I}|} < 1$, then

$$\left(1-\delta_{|\mathcal{I}|}\right)\|\mathbf{v}\|_2 \le \left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{A}_{\mathcal{I}} \mathbf{v}_{\mathcal{I}}\right\|_2 \le \left(1+\delta_{|\mathcal{I}|}\right)\|\mathbf{v}\|_2$$

holds for any $\mathbf{v}$ supported on $\mathcal{I}$.

Lemma 2: For disjoint sets $\mathcal{I}, \mathcal{J}$, if $\delta_{|\mathcal{I}|+|\mathcal{J}|} < 1$, then

$$\left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{A}\mathbf{v}\right\|_2 = \left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{A}_{\mathcal{J}} \mathbf{v}_{\mathcal{J}}\right\|_2 \le \delta_{|\mathcal{I}|+|\mathcal{J}|}\|\mathbf{v}\|_2$$

holds for any $\mathbf{v}$ supported on $\mathcal{J}$.

Lemma 3: If the sensing matrix satisfies the RIP of both orders K_1 and K_2, then $\delta_{K_1} \le \delta_{K_2}$ for any $K_1 \le K_2$

All proofs of the above lemmas are given in [3].

C. Proof of Theorem 1

1) We provide a condition under which the OMP algorithm selects a correct index in the first iteration. 2) We show that the residual in the general iteration preservers the sparsity of a K sparse signal. 3) The condition for the first iteration can be extended to the general iteration. 4) Theorem 1 is established from the conditions. The statements are an overall strategy of Proof of Theorem1.

First, we need investigate the condition when the OMP algorithm selects a correct index in the first iteration. Let us denote t^k be the index of the column maximally correlated with the residual $\mathbf{r}^{k-1}$. In the first iteration, we have

$$t^1 = \arg\max_i \left\|\langle \mathbf{a}_i, \mathbf{r}^0 \rangle\right\| = \arg\max_i \left\|\langle \mathbf{a}_i, \mathbf{y} \rangle\right\|. \tag{7}$$

Now, let us suppose that t^1 always belong to the support set $\mathcal{I}$ of $\mathbf{x}$. From (7), we have

$$\begin{aligned}\left\|\langle \mathbf{a}_{t^1}, \mathbf{y} \rangle\right\| &= \left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{y}\right\|_\infty \\ &\overset{(a)}{\ge} \frac{1}{\sqrt{K}}\left\|\mathbf{A}_{\mathcal{I}}^T \mathbf{y}\right\|_2 \\ &\overset{(b)}{\ge} \frac{1}{\sqrt{K}}\left(1-\delta_K\right)\|\mathbf{x}_{\mathcal{I}}\|_2,\end{aligned} \tag{8}$$

where (a) from the norm inequalities, and (b) from the fact that $y = \mathbf{A}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}$ and Lemma 1.Suppose that t^1 does not belong to the support set $\mathcal{I}$, then

$$\begin{aligned} \left\| \left\langle \mathbf{a}_{t^1}, \mathbf{y} \right\rangle \right\| &= \left\| \mathbf{a}_{t^1}^T \mathbf{A}_{\mathcal{I}} \mathbf{x}_{\mathcal{I}} \right\| \\ &\overset{(a)}{\leq} \left(1-\delta_{K+1}\right) \left\| \mathbf{x}_{\mathcal{I}} \right\|_2, \end{aligned} \tag{9}$$

where (a) from Lemma 2. Clearly, t^1 must belong to the support set $\mathcal{I}$. Thus, if

$$\frac{1}{\sqrt{K}}\left(1-\delta_K\right)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2 > \left(1-\delta_{K+1}\right)\left\|\mathbf{x}_{\mathcal{I}}\right\|_2 \tag{10}$$

then, the OMP algorithm selects a correct index in the first iteration. The equation (10) becomes $\sqrt{K}\delta_{K+1} + \delta_K < 1$. From Lemma 3, the inequality becomes $\sqrt{K}\delta_{K+1} + \delta_{K+1} < 1$ which leads to

$$\delta_{K+1} < \frac{1}{\sqrt{K}+1} \tag{11}$$

In short, if (11) is true, then the OMP algorithm always selects a correct index in the first iteration.

Now, we investigate a condition such that the OMP algorithm selects a correct index in the $(k+1)$ th iteration. Let us suppose that initial k iterations of the OMP algorithm are successful. Namely, $\mathcal{T}^k = \left\{t^1, \cdots, t^k\right\} \in \mathcal{I}$. Then, $\mathbf{r}^k = \mathbf{y} - \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} \in span\left(\mathbf{A}_{\mathcal{I}}\right)$ because $\mathbf{y} = \mathbf{A}_{\mathcal{I}}\mathbf{x}_{\mathcal{I}}$ and is a sub matrix of $\mathbf{A}_{\mathcal{I}}$. Thus, $\mathbf{r}^k$ can be expressed as $\mathbf{r}^k = \mathbf{A}\mathbf{x}^k$ (i.e., $\mathbf{r}^k$ is a linear combination of the K columns of $\mathbf{A}_{\mathcal{I}}$), where the support set of $\mathbf{x}^k$ belongs to the support set of $\mathbf{x}$. If the OMP algorithm selects a correct index belonging to the support set of $\mathbf{x}^k$, then the OMP algorithm also selects a correct index belonging to the support set of $\mathbf{x}$. Clearly, if $\sqrt{K}\delta_{K+1} + \delta_{K+1} < 1$ is satisfied, then the OMP algorithm success in the $(k+1)$ th iteration.

Last, we need to show that the index t^{k+1} selected at the $(k+1)$ th iteration of the OMP algorithm does not belong to T^k. First, we have $\hat{\mathbf{x}}_{\mathcal{T}^k} = \mathbf{A}_{\mathcal{T}^k}^{\dagger}\mathbf{y}$, and $\mathbf{r}^k = \mathbf{y} - \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} = \mathbf{P}_{\mathcal{T}^k}^{\perp}\mathbf{y}$. Second, for all $i \in \mathcal{T}^k$, we have

$$\begin{aligned} \left\langle \mathbf{a}_i, \mathbf{r}^k \right\rangle &= \left\langle \mathbf{a}_i, \mathbf{y} - \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} \right\rangle \\ &= \left\langle \mathbf{a}_i, \mathbf{y} \right\rangle - \left\{ \mathbf{a}_i, \mathbf{A}_{\mathcal{T}^k}\hat{\mathbf{x}}_{\mathcal{T}^k} \right\} \\ &= \mathbf{a}_i^T A_{\mathcal{I}} x_{\mathcal{I}} - \mathbf{a}_i^T \mathbf{A}_{\mathcal{T}^k} \mathbf{A}_{\mathcal{T}^k}^{\dagger} \mathbf{y} \\ &= 0. \end{aligned}$$

Therefore, we conclude that $\mathbf{r}^k$ is orthogonal to the columns $\mathbf{a}_i$ for all $i \in T^k$. It leads to $t^{k+1} \notin \mathcal{T}^k$. Furthermore, if $\mathbf{r}^k \neq \mathbf{0}$ and $\mathbf{r}^k \in span(\mathbf{A}_{\mathcal{I}})$, then there exists $i \in \mathcal{I}$ such as $\langle \mathbf{a}_i, \mathbf{r}^k \rangle \neq 0$. Therefore, the OMP algorithm selects $i \in \mathcal{I} \setminus \mathcal{T}^k$.

Now, we apply the mathematical induction. First, we proved that the OMP algorithm select a correct index if $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Second, when we assume that the initial k iterations of the OMP algorithm are successful, the OMP algorithm select a correct index in the $(k+1)^{\text{th}}$ iteration if $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$. Thus, the OMP algorithm will terminate after the K^{th} iteration if $\delta_{K+1} < \frac{1}{\sqrt{K}+1}$.

V. Discussion on Theorem 1

It is hard for us to determine δ_{K+1} from a sensing matrix because we need to examine all possible K sparse signal. However, the below result is known

Result 1[ref]: If an $M \times N$ sensing matrix $\mathbf{A}$ whose entries are i.i.d. $\mathcal{N}(0,1/M)$, then $\mathbf{A}$ obeys the RIP condition $\delta_K \leq \varepsilon$ with high probability under

$$M \geq \frac{\rho K \log\left(\frac{N}{K}\right)}{\varepsilon^2} \tag{12}$$

where ρ is a positive constant. When we utilize the above inequalities, we indirectly compare the result obtained by [2] and the result presented in this talk.

	A sufficient condition	A sufficient condition on M
[1]	$\delta_{K+1} < 1/(3\sqrt{K})$	$M \geq \rho 9K(K+1)\log\left(\frac{N}{K+1}\right)$
The paper	$\delta_{K+1} < 1/(\sqrt{K}+1)$	$M \geq \rho(K+1)(\sqrt{K}+1)^2 \log\frac{N}{K+1}$

Appendix

Example 1) computing all the Eigen values of $\begin{bmatrix} 1 & b & \cdots & b \\ b & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & 1 \end{bmatrix}$.

$$\begin{vmatrix} 1-\lambda & b & b \\ b & 1-\lambda & b \\ b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 2b & 1-\lambda+b \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 0 & 1-\lambda+2b \end{vmatrix}$$

$$= (1-\lambda-b)^2 (1-\lambda+2b)$$

Therefore, $\lambda_1 = \lambda_2 = 1-b$, and $\lambda_3 = 1+2b$.

$$\begin{vmatrix} 1-\lambda & b & b & b \\ b & 1-\lambda & b & b \\ b & b & 1-\lambda & b \\ b & b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ b & b & 1-\lambda & b \\ b & b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & b & b & b \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & b & 1-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ b & b & b & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & b & b & 1-\lambda+b \end{vmatrix}$$

$$= \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 0 & 2b & 1-\lambda+b \end{vmatrix} = \begin{vmatrix} 1-\lambda-b & 0 & 0 & b-(1-\lambda) \\ 0 & 1-\lambda-b & b-(1-\lambda) & 0 \\ 0 & 0 & 1-\lambda-b & b-(1-\lambda) \\ 0 & 0 & 0 & 1-\lambda+3b \end{vmatrix}$$

$$= (1-\lambda-b)^3 (1-\lambda+3b)$$

Therefore, $\lambda_1 = \lambda_2 = \lambda_3 = 1-b$, and $\lambda_4 = 1+3b$

Thus, we concluded all the Eigen values of a $(K+1)\times(K+1)$ $\begin{bmatrix} 1 & b & \cdots & b \\ b & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & 1 \end{bmatrix}$ are

$\lambda_1 = \cdots = \lambda_K = 1-b,$ and $\lambda_{K+1} = 1+Kb$.

Reference

[1] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Trans. Inf. Theory, vol. 53, pp. 4655 – 4666, Dec. 2007.

[2] M. A. Davenport and M. B. Wakin, "Analysis of Orthogonal Matching Pursuit using the restricted i sometry property," IEEE Trans. Inf. Theory, vol. 56, pp. 589 – 592, 2008.

[3] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2230 – 2249, May 2009.

[4] R. Baraniuk, M. Davenport, R. De Vore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," Construct. Approx., vol. 28, no. 3, pp. 253– 263, 2008.