Coherence and RIP Analysis for Greedy Algorithms in Compressive Sensing

Mingrui Yang, Member, IEEE, and Frank de Hoog

Abstract

In this paper we define a new coherence index, named 2-coherence, of a given dictionary and study its relationship with the traditional mutual coherence and the restricted isometry constant. By exploring this relationship, we obtain more general results on sparse signal reconstruction using greedy algorithms in the compressive sensing (CS) framework. In particular, we obtain an improved bound over the best known results on the restricted isometry constant for successful recovery of sparse signals using orthogonal matching pursuit (OMP).

We also initialized a study of a thresholding type greedy algorithm named orthogonal matching pursuit with thresholding (OMPT), which is more feasible in practice than OMP. We analyze its performance in CS framework for both noiseless and noisy cases in terms of coherence indices and the restricted isometry constant. We show that given the same assumptions as required for OMP, it achieves exactly the same reconstruction performance as OMP.

Index Terms

Compressive sensing, mutual coherence, 2-coherence, restricted isometry property, orthogonal matching pursuit (OMP), orthogonal matching pursuit with thresholding (OMPT).

I. INTRODUCTION

Compressive sensing (CS) [1]–[3] is a newly developed and fast growing field of research. It provides a new sampling scheme that breaks the traditional Shannon-Nyquist sampling rate [4] given that the signal of interest is sparse in a certain basis or tight frame. The problem can be formulated as follows. For a vector \( a \in \mathbb{R}^d \), let \( \| a \|_0 \) denote \( \ell_0 \) “norm” of \( a \), which counts the number of nonzero entries in \( a \). We say \( a \) is \( k \)-sparse if \( \| a \|_0 \leq k \). One of the fundamental problems of CS is to solve the following \( \ell_0 \) minimization problem

\[
\min_a \| a \|_0 \text{ subject to } f = \Phi a,
\]

where \( \Phi \in \mathbb{R}^{n \times d} (n \ll d) \) and \( f \in \mathbb{R}^n \). To ensure that the \( k \)-sparse solution is unique, we need the following restricted isometry property introduced by Candes and Tao in [5].

Mingrui Yang is with CSIRO Computational Informatics, Pullenvale, QLD 4069 Australia (email: mingrui.yang@csiro.au).

Frank de Hoog is with CSIRO Computational Informatics, Acton, ACT 2061 Australia (email: frank.dehoog@csiro.au).
**Definition 1.1** (Restricted Isometry Property). A matrix $\Phi$ satisfies the restricted isometry property of order $k$ with the restricted isometry constant $\delta_k$ if $\delta_k$ is the smallest constant such that

$$
(1 - \delta_k)\|a\|_2^2 \leq \|\Phi a\|_2^2 \leq (1 + \delta_k)\|a\|_2^2
$$

(2)

holds for all $k$-sparse signal $a$.

It has been shown in [5] that if $\delta_{2k} < 1$, then the $\ell_0$ minimization problem [1] has a unique $k$-sparse solution. However, solving an $\ell_0$ minimization problem is in general NP-hard.

There are two streams of work in the literature to address this problem. The first one is to relax the $\ell_0$ minimization problem to an $\ell_1$ minimization problem. Candès shows in [6] if $\delta_{2k} < \sqrt{2} - 1$, then $\ell_1$ minimization is equivalent to $\ell_0$ minimization. Better bounds have been developed [7]–[10]. The most recent result along this direction is $\delta_{2k} < 0.4931$ [11]. The other stream of work is to use greedy algorithms to approximate the solution of the $\ell_0$ minimization problem [12]–[18]. Orthogonal matching pursuit (OMP) is one of the most popular algorithms along this line. Different from $\ell_1$ minimization, the metric for a sensing matrix in using greedy algorithms is usually chosen to be coherence indices.

For simplicity, from now on, we always assume that the columns of the matrix (dictionary) $\Phi$ are normalized such that for any column $\phi \in \Phi$, $\|\phi\|_2 = 1$.

**Definition 1.2.** The mutual coherence $M(\Phi)$ of a matrix $\Phi$ is defined by

$$
M(\Phi) := \max_{\phi_i, \phi_j \in \Phi, i \neq j} |\langle \phi_i, \phi_j \rangle|,
$$

(3)

where $\langle \cdot, \cdot \rangle$ represents the usual inner product.

It has been shown that if $(2k - 1)M < 1$, then OMP can recover every $k$-sparse signal exactly in $k$ iterations [13], [15]. Recently, researchers have started to investigate the performance of OMP using the RIP constant. Davenport and Wakin [19] have proved that $\delta_{k+1} < \frac{1}{3\sqrt{k}}$ is sufficient for OMP to recover any $k$-sparse signal in $k$ iterations. Later, improvements over this bound have been achieved in [20]–[22]. The latest result in [22] improves the bound to $\delta_{k+1} < \frac{1}{1 + \sqrt{k}}$. Mo and Shen [22] also give an example that OMP fails when $\delta_{k+1} = \frac{1}{\sqrt{k}}$, as was conjectured by Dai and Milenkovic in [17].

In this paper, we define a new coherence index, named 2-coherence. We first establish a bridge connecting the mutual coherence, the 2-coherence, and the restricted isometry constant. Then by using this newly defined coherence index, we analyze the performance of weak orthogonal matching pursuit (WOMP), a weak version of OMP, for both noiseless and noisy scenario. In particular, we show that $\delta_k + \sqrt{k}\delta_{k+1} < 1$ is sufficient for OMP to recover any $k$-sparse signal in $k$ iterations, which provides an improved bound over the best known result from [22].

We also initialize a study of a thresholding type greedy algorithm named orthogonal matching pursuit with thresholding (OMPT) in the CS framework. OMPT is easier than OMP to implement. It does not require the more expensive greedy step, which calculates in each iteration the inner products between the residual and all the
atoms from the dictionary. Instead, it only needs to calculate the $\ell_2$ norm of the residual once in each iteration. We show that by carefully choosing the thresholding parameter, OMPT maintains exactly the same reconstruction performance as OMP, for both ideal noiseless and noisy cases. More details are given in Section \textbf{V}.

II. A New Coherence Index: 2-COHERRENCE

We first define a new coherence index, the 2-coherence, $\nu_k(\Phi)$ for a given dictionary $\Phi$. Then based on this new coherence index, we establish the connections among the coherence indices and the restricted isometry constant $\delta_k$.

**Definition II.1.** Denote $[d]$ the index set $\{1, 2, \ldots, d\}$. The 2-coherence of a dictionary $\Phi \in \mathbb{R}^{n \times d}$ is defined as

$$\nu_k(\Phi) := \max_{i \in [d]} \max_{|\Lambda| \leq k} \left( \sum_{j \in \Lambda} \langle \phi_i, \phi_j \rangle^2 \right)^{1/2},$$

(4)

where $\phi_i, \phi_j$ are columns from the dictionary $\Phi$.

Notice that the 2-coherence $\nu_k(\Phi)$ defined above is more general than the mutual coherence defined in Definition \textbf{I.2}. In fact, when $k = 1$, the 2-coherence defined in Definition \textbf{II.1} is exactly the mutual coherence.

Now we are ready to establish the following lemma, which describes the relations among the mutual coherence $M$, the 2-coherence $\nu_k$, and the restricted isometry constant $\delta_k$.

**Lemma II.2.** For $k > 1$, we have

$$M \leq \nu_{k-1} \leq \delta_k \leq \sqrt{k-1} \nu_{k-1} \leq (k-1)M.$$  

(5)

**Proof:** It is easy to show that $\nu_k$ increases with $k$ while $\frac{\nu_k}{\sqrt{k}}$ decreases with $k$. Therefore, the first and the last relations follow immediately.

We now prove the second inequality.

$$\nu_{k-1}(\Phi) = \max_{i \in [d]} \max_{|\Lambda| \leq k-1, |\Lambda| \leq k} \left( \sum_{j \in \Lambda} \langle \phi_i, \phi_j \rangle^2 \right)^{1/2}$$

$$= \max_{|\Lambda| \leq k} \max_{i \in \Lambda} \left( \sum_{j \in \Lambda \setminus \{i\}} \langle \phi_i, \phi_j \rangle^2 \right)^{1/2}$$

$$= \max_{|\Lambda| \leq k} \| \Phi^T \Lambda \Phi_{\Lambda} - I \|_{\infty, 2},$$

where $\Phi_{\Lambda} \in \mathbb{R}^{n \times |\Lambda|}$ is a submatrix of $\Phi$ with columns indexed in $\Lambda$. 


On the other hand, according to Proposition 2.5 in [23], one has

\[
\delta_k = \max_{\Lambda \subseteq [d], |\Lambda| \leq k} \|\Phi^T_\Lambda \Phi_\Lambda - I\|_{2,2}
\]

\[
\geq \max_{\Lambda \subseteq [d], |\Lambda| \leq k} \|\Phi^T_\Lambda \Phi_\Lambda - I\|_{\infty,2}
\]

\[
= \nu_{k-1}(\Phi),
\]

which completes the proof for the second inequality.

Next we prove the third inequality. Consider the Gram matrix \( G = \Phi^T_\Lambda \Phi_\Lambda \), where its entries \( g_{ij} = \langle \phi_i, \phi_j \rangle \).

Clearly its diagonal entries \( g_{ii} = 1 \). Then by the Gershgorin Circle Theorem, each eigenvalue \( \lambda \) of \( G \) is in at least one of the disks \( \{ z : |z - 1| \leq R_i \} \), where \( R_i = \sum_{j \neq i} |g_{ij}| \). Equivalently, we have

\[
1 - R_i \leq \lambda \leq 1 + R_i
\]

for some \( i \). Therefore,

\[
\delta_k \leq \max_i R_i
\]

\[
= \max_i \sum_{j \in \Lambda, j \neq i} |g_{ij}|
\]

\[
\leq \max_i \sqrt{k - 1} \left( \sum_{j \in \Lambda, j \neq i} |g_{ij}|^2 \right)^{\frac{1}{2}}
\]

\[
\leq \sqrt{k - 1} \nu_{k-1}.
\]

Before proceeding to the analysis of greedy algorithms using the new coherence index \( \nu_k \), we need the following lemmas.

**Lemma II.3.** Let \( \Lambda \subseteq [d] \) with \( |\Lambda| = k \) and \( f = \Phi a \) where \( a \) is a \( k \)-sparse vector with \( \text{supp}(a) = \Lambda \). In addition, assume that there exits \( \Omega \subseteq \Lambda \) with \( |\Omega| = m \), such that

\[
\langle f, \phi_i \rangle = 0, \quad \text{for } i \in \Lambda \setminus \Omega.
\]

Then

\[
\max_{i \in [d] \setminus \Lambda} |\langle f, \phi_i \rangle| \leq \nu_k |a|_2, \quad (6)
\]

\[
\max_{i \in \Lambda} |\langle f, \phi_i \rangle| \geq \frac{\sqrt{1 - \delta_k}}{\sqrt{m}} \|f\|_2. \quad (7)
\]
Proof: For \( i \in [d] \setminus \Lambda \),
\[
|\langle f, \phi_i \rangle| = \left| \sum_{j \in \Lambda} a_j \langle \phi_j, \phi_i \rangle \right|
\leq \|a\|_2 \left( \sum_{j \in \Lambda} |\langle \phi_j, \phi_i \rangle|^2 \right)^{\frac{1}{2}}
\leq \nu_k \|a\|_2.
\]
This establishes the first inequality.

For (7), we have
\[
\|f\|_2^2 = \langle f, \sum_{i \in \Lambda} a_i \phi_i \rangle
= \sum_{i \in \Lambda} a_i \langle f, \phi_i \rangle
= \sum_{i \in \Omega} a_i \langle f, \phi_i \rangle
\leq \sum_{i \in \Omega} |a_i| \cdot |\langle f, \phi_i \rangle|
\leq \sqrt{m} \|a\|_2 \max_{i \in \Lambda} |\langle \Phi a, \phi_i \rangle|.
\]
Equivalently,
\[
\max_{i \in \Lambda} |\langle \Phi a, \phi_i \rangle| \geq \frac{\|f\|_2^2}{\sqrt{m} \|a\|_2}.
\]
The result now follows from the definition of the restricted isometry constant \( \delta_k \).

This can easily be extended to the noisy case where a signal is contaminated by a perturbation. Specifically, we have:

**Lemma II.4.** Let \( \Lambda \subset [d] \) with \( |\Lambda| = k \). Let \( f = \Phi a + w \) with \( \text{supp}(a) = \Lambda \) and \( \|w\|_2 \leq \epsilon \). In addition, assume that there exists \( \Omega \subseteq \Lambda \) with \( |\Omega| = m \), such that
\[
\langle \Phi a, \phi_i \rangle = 0, \text{ for } i \in \Lambda \setminus \Omega.
\]
Then
\[
\max_{i \in [d] \setminus \Lambda} |\langle f, \phi_i \rangle| \leq \nu_k \|a\|_2 + \epsilon,
\]
\[
\max_{i \in \Lambda} |\langle f, \phi_i \rangle| \geq \frac{\sqrt{1 - \delta_k}}{\sqrt{m}} \|\Phi a\|_2 - \epsilon.
\]
**Proof:** For $i \in [d] \setminus \Lambda$, we have

\[
|\langle f, \phi_i \rangle| = |\langle \Phi a + w, \phi_i \rangle| \\
\leq |\langle \Phi a, \phi_i \rangle| + |\langle w, \phi_i \rangle| \\
\leq \nu_k \|a\|_2 + \|w\|_2 \|\phi_i\|_2 \\
\leq \nu_k \|a\|_2 + \epsilon.
\]

Taking maximum on both sides completes the proof of the first inequality.

Now for $i \in \Lambda$,

\[
\max_{i \in \Lambda} |\langle f, \phi_i \rangle| = \max_{i \in \Lambda} |\langle \Phi a + w, \phi_i \rangle| \\
\geq \max_{i \in \Lambda} |\langle \Phi a, \phi_i \rangle| - \max_{i \in \Lambda} |\langle w, \phi_i \rangle| \\
\geq \frac{\sqrt{1 - \delta_k}}{\sqrt{m}} \|\Phi a\|_2 - \max_{i \in \Lambda} \|w\|_2 \|\phi_i\|_2 \\
\geq \frac{\sqrt{1 - \delta_k}}{\sqrt{m}} \|\Phi a\|_2 - \epsilon.
\]

This completes the proof of the second inequality.

Now we are ready to analyze the performance of some greedy algorithms in CS framework using the new coherence index $\nu_k$.

### III. Weak Orthogonal Matching Pursuit

We first begin with a well known greedy algorithm, the weak orthogonal matching pursuit (WOMP), which was defined in [24]. Here we present a simple version in Algorithm 1 where the weak parameter $\rho$ is a constant for each iteration.

Notice that OMP is a special case of WOMP when $\rho = 1$.

The following recovery property of WOMP was proved in [13].

**Theorem III.1.** Let $\Lambda \subset [d]$ with $|\Lambda| = k$. Let $f = \Phi a + w$ with $\text{supp}(a) = \Lambda$ and $\|w\|_2 \leq \epsilon$. Denote $a_{\text{min}}$ the nonzero coefficient with the least magnitude, and $\hat{a}_{\text{womp}}$ the recovered representation of $f$ in $\Phi$ by WOMP after $k$ iterations. If

\[
k < \frac{1 + M^{-1}}{1 + \rho^{-1}}
\]

and the noise level obeys

\[
\epsilon < \frac{\rho(1 - (k - 1)M) - kM}{1 + \rho} |a_{\text{min}}|,
\]

then

a) $\hat{a}_{\text{womp}}$ has the correct sparsity pattern

\[
\text{supp}(\hat{a}_{\text{womp}}) = \text{supp}(a);
\]

\[
\hat{a}_{\text{womp}} = \sum_{i \in \Lambda} \alpha_i \Phi_i.
\]

Theorem III.1.

Let $\Lambda \subset [d]$ with $|\Lambda| = k$. Let $f = \Phi a + w$ with $\text{supp}(a) = \Lambda$ and $\|w\|_2 \leq \epsilon$. Denote $a_{\text{min}}$ the nonzero coefficient with the least magnitude, and $\hat{a}_{\text{womp}}$ the recovered representation of $f$ in $\Phi$ by WOMP after $k$ iterations. If

\[
k < \frac{1 + M^{-1}}{1 + \rho^{-1}}
\]

and the noise level obeys

\[
\epsilon < \frac{\rho(1 - (k - 1)M) - kM}{1 + \rho} |a_{\text{min}}|,
\]

then

a) $\hat{a}_{\text{womp}}$ has the correct sparsity pattern

\[
\text{supp}(\hat{a}_{\text{womp}}) = \text{supp}(a);
\]
Algorithm 1 Weak Orthogonal Matching Pursuit (WOMP)

1: **Input:** weak parameter $\rho \in (0, 1]$, dictionary $\Phi$, signal $f$, and the noise level $\epsilon$.
2: **Initialization:** $r_0 := f$, $x_0 := 0$, $\Lambda_0 := \emptyset$, $s := 0$.
3: while $\|r_s\|_2 > \epsilon$ do
4:      Find an index $i$ such that $|\langle r_s, \phi_i \rangle| \geq \rho \cdot \max_{\phi} |\langle r_s, \phi \rangle|$
where $\phi$ is any column of $\Phi$;
5:      Update the support:
     \[ \Lambda_{s+1} = \Lambda_s \cup \{i\}; \]
6:      Update the estimate:
     \[ x_{s+1} = \arg\min_z \|f - \Phi_{\Lambda_{s+1}} z\|_2; \]
7:      Update the residual:
     \[ r_{s+1} = f - \Phi_{\Lambda_{s+1}} x_{s+1}; \]
8:      $s = s + 1$;
9:  end while
10: **Output:** If the algorithm is stopped after $k$ iterations, then the output estimate $\hat{a}$ of $a$ is $\hat{a}_{\Lambda_k} = x_k$ and $\hat{a}_{\Lambda^c_k} = 0$.

b) $\hat{a}_{\text{womp}}$ approximates the ideal noiseless representation
\[ \|\hat{a}_{\text{womp}} - a\|_2^2 \leq \frac{\epsilon^2}{1 - (k - 1)M}. \] (10)

Similar results hold for OMP as shown in [15].

Next, we present our new results based on the 2-coherence $\nu_k$.

A. Results for Noiseless Ensembles

We first consider the ideal noiseless case where the signal is $k$-sparse. Specifically, let $\Lambda \subset [d]$ with $|\Lambda| = k$. We consider a signal $f = \Phi a$, where $f \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{n \times d}$, and $a \in \mathbb{R}^d$ with $\text{supp}(a) = \Lambda$. We have the following results.

Theorem III.2. Let $f = \Phi a$ with $\|a\|_0 = k$. If
\[ \sqrt{k}\nu_k < \rho(1 - \delta_k), \] (11)
then $a$ is the unique sparsest representation of $f$ and moreover, WOMP recovers $a$ exactly in $k$ steps.
Proof: First, we need to show for the first iteration, we have
\[ \max_{i \in [d] \setminus \Lambda} \| r_0, \phi_i \| < \rho \max_{i \in \Lambda} \| r_0, \phi_i \|. \]
By Lemma II.3 with \( m = k \) and (11), we have
\[
\max_{i \in [d] \setminus \Lambda} | \langle f, \phi_i \rangle | \leq \nu_k \| a \|_2 \\
\leq \nu_k \frac{\| f \|_2}{\sqrt{1 - \delta_k}} \\
\leq \frac{\nu_k}{\sqrt{1 - \delta_k}} \sqrt{1 - \delta_k} \max_{i \in \Lambda} | \langle f, \phi_i \rangle | \\
< \rho \max_{i \in \Lambda} | \langle f, \phi_i \rangle |.
\]
Therefore, WOMP only chooses one of the atoms from \( \phi_i \) for \( i \in \Lambda \) at the first iteration.

Arguing as above, we prove that in passing from iteration 1 to iteration \( k \), it lands us back each time in the same situation of studying \( r_1, \ldots, r_{k-1} \), for which we have the same structure at each iteration, namely \( r_s = \sum_{j \in \Lambda} c_j^{(s)} \phi_j \), with possibly different coefficients \( c_j^{(s)} \). Moreover, the orthogonal projection procedure guarantees that the atom selected in each iteration will not repeat the ones selected in previous iterations. Therefore, \( a \) is the unique sparsest representation of \( f \), and WOMP recovers \( a \) exactly in \( k \) iterations. \( \blacksquare \)

Note that the bound given in (11) is the best result in the literature. From Lemma II.2, it follows that

**Corollary III.3.** Let \( f = \Phi a \) with \( \| a \|_0 = k \). If
\[
\sqrt{k} \delta_{k+1} < \rho(1 - \delta_k),
\] (12)
or
\[
\sqrt{k} \nu_k < \rho(1 - \nu_{k-1} \sqrt{k - 1}),
\]
or
\[
kM < \rho(1 - (k - 1)M),
\]
then, \( a \) is the unique sparsest representation of \( f \) and moreover, WOMP recovers \( a \) exactly in \( k \) iterations.

Notice that the performance of WOMP decreases as \( \rho \) decreases. Now if we set \( \rho = 1 \) in WOMP, then we obtain the following two corollaries for OMP immediately.

**Corollary III.4.** Let \( f = \Phi a \) with \( \| a \|_0 = k \). If
\[
\delta_k + \sqrt{k} \delta_{k+1} < 1,
\] (13)
then, \( a \) is the unique sparsest representation of \( f \) and moreover, OMP recovers \( a \) exactly in \( k \) iterations.
Remark III.5. The condition in (13) gives an improved bound on the restricted isometry constant compared to the bound obtained in [25] and [22] for OMP for successful recovery, where the bound was $\delta_{k+1} < \frac{1}{\sqrt{k+1}}$.

Corollary III.6. Let $f = \Phi a$ with $\|a\|_0 = k$. If
\[
    k < \frac{1}{2} \left( 1 + \frac{1}{M} \right),
\]
then, $a$ is the unique sparsest representation of $f$ and moreover, OMP recovers $a$ exactly in $k$ iterations.

Note that Corollary III.6 has been previously proved in [26], [13], and [14].

B. Results for noisy ensembles

Next, we consider the case where a sparse signal is contaminated by a perturbation. Specifically, Let $\Lambda \subset [d]$ with $|\Lambda| = k$. We consider a signal $f = \Phi a + w$, where $a \in \mathbb{R}^d$ with $\text{supp}(a) = \Lambda$ and $\|w\|_2 \leq \epsilon$.

Theorem III.7. Denote by $a_{\text{min}}$ the nonzero entry of $a$ with the least magnitude, and $\hat{a}_{\text{womp}}$ the recovered representation of $f$ in $\Phi$ by WOMP after $k$ iterations. If
\[
    \sqrt{k}\nu_k < \rho(1 - \delta_k)
\]
and the noise level obeys
\[
    \epsilon < \frac{\rho(1 - \delta_k) - \sqrt{k}\nu_k}{1 + \rho}|a_{\text{min}}|,
\]
then
a) $\hat{a}_{\text{womp}}$ has the correct sparsity pattern
\[
    \text{supp}(\hat{a}_{\text{womp}}) = \text{supp}(a);
\]
b) $\hat{a}_{\text{womp}}$ approximates the ideal noiseless representation
\[
    \|\hat{a}_{\text{womp}} - a\|_2^2 \leq \frac{\epsilon^2}{1 - \delta_k}.
\]

Proof: First, we show that WOMP recovers the correct support of $a$.

We start with the first iteration. Note that $r_0 = f$. We need to show
\[
    \max_{i \in [d]\setminus\Lambda} |\langle f, \phi_i \rangle| < \rho \max_{i \in \Lambda} |\langle f, \phi_i \rangle|.
\]
By Lemma [23] we have
\[
    \max_{i \in [d]\setminus\Lambda} |\langle f, \phi_i \rangle| \leq \nu_k \|a\|_2 + \epsilon,
\]
and
\[
    \max_{i \in \Lambda} |\langle f, \phi_i \rangle| \geq \frac{1 - \delta_k}{\sqrt{k}} \|\Phi a\|_2 - \epsilon
\]
\[
    \geq \frac{1 - \delta_k}{\sqrt{k}} \|a\|_2 - \epsilon.
\]
Now since \( |a|_2 \geq \sqrt{k} |a_{\min}| \), by imposing conditions (15) and (16), we get
\[
\nu_k |a|_2 + \epsilon < \rho \left( \frac{1 - \delta_k}{\sqrt{k}} |a|_2 - \epsilon \right),
\]
and relation (18) follows from the two bounds (19) and (20). Hence, WOMP only selects one atom from \( \{\phi_i\}_{i \in \Lambda} \) in the first iteration.

Now we argue that by repeatedly applying the above procedure, we are able to correctly recover the support of \( a \). In fact, we have for the \( s \)-th iteration
\[
r_s = f - P_{\Lambda_s}(f)
= \Phi a + w - (P_{\Lambda_s}(\Phi a) + P_{\Lambda_s}(w))
= (I - P_{\Lambda_s})\Phi a + (I - P_{\Lambda_s})w
= \Phi a_s + w_s
\]
where
\[
\Phi a_s = (I - P_{\Lambda_s})\Phi a
\]
and
\[
w_s = (I - P_{\Lambda_s})w.
\]
Therefore, \( \langle \Phi a_s, \phi_i \rangle = 0 \) for \( i \in \Lambda_s \). Note that \( (k - s) \) components of \( a_s \) are the same as that of \( a \). Then the result follows from the inequality \( |a_s|_2 \geq \sqrt{k - s} |a_{\min}| \) and Lemma II.4 for \( s \)-th iteration. In addition, the orthogonal projection step guarantees that the procedure will not repeat the atoms already chosen in previous iterations. Therefore, the correct support of the noiseless representation \( a \) can be recovered exactly after \( k \) iterations.

Next, we prove the error bound (17). The proof follows the idea of the proof of Theorem 5.1 in (15). Let \( a_T \) denote \( a \) restricted to its support. Similarly, let \( \Phi_T \) denote the dictionary \( \Phi \) restricted to the support of \( a \). The orthogonal projection step tells that WOMP solves for
\[
\hat{a}_T = \arg \min_{a_T} \| f - \Phi_T a_T \|_2 = \Phi_T^\dagger f
\]
where \( \Phi_T^\dagger \) denotes the Moore-Penrose generalized inverse of \( \Phi_T \). Then we have
\[
\hat{a}_T = \Phi_T^\dagger f
= \Phi_T^\dagger (\Phi a + w)
= \Phi_T^\dagger (\Phi_T a_T + w)
= a_T + \Phi_T^\dagger w.
\]
The term $\Phi^\dagger_T w$ denotes the reconstruction error. It can be bounded by

$$
\|\hat{a}_{\text{womp}} - a\|_2 = \|\hat{a}_T - a_T\|_2
$$

$$
= \|\Phi^\dagger_T w\|_2
$$

$$
\leq \|\Phi^\dagger\|_2 \cdot \|w\|_2
$$

$$
\leq \epsilon / \sigma_{\text{min}}
$$

where we bound the norm of $\Phi^\dagger_T$ by the smallest singular value $\sigma_{\text{min}}$ of $\Phi$. Now by RIP, we have $\sigma_{\text{min}}^2 \geq 1 - \delta_k$, and the error bound (17) follows.

Theorem III.7 is a generalization of Theorem III.1. In fact, Theorem III.7 follows immediately from Theorem III.7 by applying Lemma II.2.

IV. Orthogonal Matching Pursuit with Thresholding (OMPT)

We now introduce the orthogonal matching pursuit with thresholding (OMPT). This is a thresholding type modification of OMP and WOMP. It replaces the expensive greedy step in OMP and WOMP with a thresholding step. Details are presented in Algorithm 2.

An initial study of this algorithm in Hilbert space was presented in [27]. For the case of Banach space, a similar version of the algorithm was studied in [28] (see also [29] and [30]). We are the first to study the performance of this algorithm in the CS framework.

Next, we present our new results for both noiseless and noisy cases.

A. Results for Noiseless Ensembles

We first examine the performance of OMPT for the noiseless case. Consider the ideal noiseless case where the signal is $k$-sparse. Specifically, let $\Lambda \subset [d]$ with $|\Lambda| = k$. We consider a signal $f = \Phi a$, where $f \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{n \times d}$, and $a \in \mathbb{R}^d$ with $\text{supp}(a) = \Lambda$. Again from Lemma II.3 we get the following result.

**Theorem IV.1.** Let $f = \Phi a$ with $\|a\|_0 = k$. If

$$
\frac{\nu_k}{\sqrt{1 - \delta_k}} < t \leq \sqrt{1 - \delta_k} / \sqrt{k},
$$

then $a$ is the unique sparsest representation of $f$ and moreover, OMPT recovers $a$ exactly in $k$ iterations.

**Proof:** First we show that for the first iteration, the following two inequalities hold.

$$
\max_{i \in \Lambda} |\langle f, \phi_i \rangle| \geq t \|f\|_2, \quad (22)
$$

$$
\max_{i \in [d] \setminus \Lambda} |\langle f, \phi_i \rangle| < t \|f\|_2. \quad (23)
$$

To show (22), we have by Lemma II.3 and (21)

$$
\max_{i \in \Lambda} |\langle f, \phi_i \rangle| \geq \frac{\sqrt{1 - \delta_k}}{\sqrt{k}} \|f\|_2 \geq t \|f\|_2.
$$
Algorithm 2 Orthogonal Matching Pursuit with Thresholding (OMPT)

1: Input: threshold $t$, dictionary $\Phi$, signal $f$, and noise level $\epsilon$.
2: Initialization: $r_0 := f$, $x_0 := 0$, $\Lambda_0 := \emptyset$, $s := 0$.
3: while $\|r_s\|_2 > \epsilon$ do
4:   Find an index $i$ such that $|\langle r_s, \phi_i \rangle| \geq t\|r_s\|_2$;
5:   Update the support: $\Lambda_{s+1} = \Lambda_s \cup \{i\}$;
6:   Update the estimate: $x_{s+1} = \arg\min_z \|f - \Phi_{\Lambda_{s+1}}z\|_2$;
7:   Update the residual: $r_{s+1} = f - \Phi_{\Lambda_{s+1}}x_{s+1}$;
8:   $s = s + 1$;
9: end while
10: Output: If the algorithm is stopped after $k$ iterations, then the output estimate $\hat{a}$ of $a$ is $\hat{a}_{\Lambda_k} = x_k$ and $\hat{a}_{\Lambda^C_k} = 0$.

For (23), we have by Lemma II.3 and (21)
\[
\max_{i \in [d] \setminus \Lambda} |\langle f, \phi_i \rangle| \leq \nu_k \|a\|_2 \\
(RIP) \leq \frac{\nu_k}{\sqrt{1 - \delta_k}} \|f\|_2 \\
< t \|f\|_2.
\]
Hence, at the first iteration, OMPT only picks one atom $\phi_i$ with $i \in \Lambda$.

Arguing as above, we prove that in passing from iteration 1 to iteration $k$, it lands us back each time in the same situation of studying $r_1, \ldots, r_{k-1}$, which we have the same structure at each iteration, namely $r_s = \sum_{j \in \Lambda} c_j^{(s)} \phi_j$, with possibly different coefficients $c^{(s)}$. Moreover, the orthogonal projection procedure guarantees that the atom selected in each iteration will not repeat the ones selected in previous iterations. Therefore, $a$ is the unique sparsest representation of $f$, and WOMP recovers $a$ exactly in $k$ iterations.  

From Lemma II.2, it immediately follows that:

**Corollary IV.2.** Let $f = \Phi a$ with $\|a\|_0 = k$. If any of the following three conditions is satisfied:
i) 

\[ \delta_k + \sqrt{k} \delta_{k+1} < 1 \]  

and 

\[ \frac{\delta_{k+1}}{\sqrt{1 - \delta_k}} < t \leq \frac{\sqrt{1 - \delta_k}}{\sqrt{k}}, \]  

ii) 

\[ \nu_k \sqrt{k} + \nu_{k-1} \sqrt{k-1} < 1 \]  

and 

\[ \frac{\nu_k}{\sqrt{1 - \nu_{k-1} \sqrt{k-1}}} < t \leq \frac{\sqrt{1 - \nu_{k-1} \sqrt{k-1}}}{\sqrt{k}}, \]  

iii) 

\[ M < \frac{1}{2k-1} \]  

and 

\[ \frac{\sqrt{k} M}{\sqrt{1 - (k-1) M}} < t \leq \frac{\sqrt{1 - (k-1) M}}{\sqrt{k}}, \]

then, \( a \) is the unique sparsest representation of \( f \) and moreover, OMPT recovers \( a \) exactly in \( k \) iterations.

As we can see from the above corollary, although we are replacing the most difficult (expensive) step of OMP, namely the greedy step, by a very simple thresholding step making it more practically feasible, there is no performance degrading at all compared to OMP. The bound on the restricted isometry constant \( \delta_k \) in (24) is exactly the same as (13) in Corollary III.4. The bound on the mutual coherence \( M \) in (26) also coincides with the bound (14) in Corollary III.6, which is known to be optimal for OMP [30].

B. Results for noisy ensembles

Again, we consider the case where a sparse signal is contaminated by a perturbation. Specifically, Let \( \Lambda \subset [d] \) with \( |\Lambda| = k \). We consider a signal \( f = \Phi a + w \), where \( a \in \mathbb{R}^d \) with \( \text{supp}(a) = \Lambda \) and \( \|w\|_2 \leq \epsilon \). To show the main result for the noisy case, we will need the following lemmas.

**Lemma IV.3.** Consider the residual at the \( s \)-th iteration of OMPT \( r_s = \Phi a_s + w_s \). If 

\[ t > \frac{\nu_k |a_{\min}| + \epsilon}{\sqrt{1 - \delta_k |a_{\min}|} - \epsilon}, \]  

then 

\[ \max_{i \in [d] \setminus \Lambda} |\langle r_s, \phi_i \rangle| < t ||r_s||_2. \]
Proof: By using Lemma II.4 and the fact that $\|a_s\|_2 \geq \sqrt{k-s}|a_{\text{min}}|$, it is easy to derive

$$\max_{i \in [d] \setminus \Lambda} |\langle r_s, \phi_i \rangle| \leq \nu_k a_s + \epsilon$$

$$< t \left( \sqrt{1 - \delta_k} \|a_s\|_2 - \epsilon \right)$$

$$\leq t (\|\Phi a_s\|_2 - \epsilon)$$

$$\leq t \|r_s\|_2 .$$

Lemma IV.4. Consider the residual at the $s$-th iteration of OMPT $r_s = \Phi a_s + w_s$. Given

$$t \leq \frac{(1 - \delta_k)|a_{\text{min}}| - \epsilon}{\sqrt{k(1 - \delta_k)|a_{\text{min}}| + \epsilon}} ,$$

we have

$$\max_{i \in \Lambda} |\langle r_s, \phi_i \rangle| \geq t \|r_s\|_2 .$$

Proof: By using Lemma II.4 and the inequality

$$\sqrt{(k-s)(1 - \delta_k)|a_{\text{min}}| \leq \sqrt{1 - \delta_k}\|a\|_2 \leq \|\Phi a\|_2}$$

we have

$$\max_{i \in \Lambda} |\langle r_s, \phi_i \rangle| \geq \frac{\sqrt{1 - \delta_k}}{\sqrt{k - s}} \|\Phi a_s\|_2 - \epsilon$$

$$\geq t (\|\Phi a_s\|_2 + \epsilon)$$

$$\geq t \|r_s\|_2 .$$

Theorem IV.5. Denote by $a_{\text{min}}$ the nonzero coefficient with least magnitude, and $a_{\text{ompt}}$ the recovered coefficient vector of $f$ in $\Phi$ by OMPT after $k$ iterations. If

$$\delta_k + \sqrt{k}\nu_k < 1$$

and the noise level obeys

$$\epsilon < \frac{\sqrt{1 - \delta_k(1 - \delta_k - \sqrt{k}\nu_k)}}{(\sqrt{k} + 1)\sqrt{1 - \delta_k} + (1 - \delta_k) + \sqrt{k}\nu_k} |a_{\text{min}}|,$$

then there exists threshold $t$ satisfying conditions (27) and (28). Moreover, we have

a) $a_{\text{ompt}}$ has the correct sparsity pattern

$$\text{supp}(a_{\text{ompt}}) = \text{supp}(a);$$

b) $a_{\text{ompt}}$ approximates the ideal noiseless representation

$$\|a_{\text{ompt}} - a\|_2^2 \leq \frac{\epsilon^2}{1 - \delta_k} .$$
Proof: First we show that $a_{\text{ompt}}$ has the correct support.

We start with the first iteration. Combining conditions (29) and (30), Lemma IV.3 and Lemma IV.4, it is easy to see that OMPT is able to select and only select an atom $\phi_i$ with $i \in \Lambda$. Condition (29) and (30) guarantees the existance of threshold $t$ satisfying Lemma IV.3 and Lemma IV.4. Lemma IV.3 guarantees that OMPT will not choose any atom $\phi_i$ for $i \in [d] \setminus \Lambda$. While Lemma IV.4 guarantees that OMPT is able to choose atoms $\phi_i$ with $i \in \Lambda$.

Next we argue that by repeatedly applying Lemma IV.3 and Lemma IV.4, we are able to correctly recover the support of $a$. In fact, in each iteration, we have the same situation as in the first iteration. In addition, the orthogonal projection step guarantees that the procedure will not repeat the atoms already chosen in previous iterations. Thus, all the correct support of the noiseless coefficient vector $a$ can be recovered precisely after $k$ iterations.

The proof of the error bound (31) is the same as the proof for WOMP in Theorem III.7.

Theorem IV.5 basically says, if the minimal nonzero coefficient of the ideal noiseless signal is significant enough compared to the noise level, then the correct support of the coefficient vector can be recovered exactly, and moreover, the error can be bounded by (31).

Notice that by applying Lemma II.2, it is easy to obtain similar results for the restricted isometry constant $\delta_k$, the coherence index $\nu_k$, and the mutual coherence $M$ of the dictionary $\Phi$ respectively.

V. Conclusion

In this paper, we have introduced a new coherence index, the 2-coherence, which generalizes the mutual coherence and the so called 1-coherence. We established the connections among the mutual coherence, the 2-coherence, and the restricted isometry constant. Based on these relations, we analyzed the performance of WOMP as well as OMP for their recovery ability of sparse representations in both ideal noiseless and noisy cases. In particular, for the noiseless case, we showed an improved bound over the best known results on the restricted isometry constant for successful recovery using OMP. We have also initialized a study of a thresholding type algorithm, named OMPT, which replaces the expensive greedy step in OMP by a thresholding step, making it more feasible in practice. We showed that this simplified algorithm has exactly the same recovery performance as OMP for CS reconstruction in both noiseless and noisy cases.

REFERENCES

[1] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. on Inf. Theory, vol. 52, no. 2, pp. 489–509, 2006.
[2] E. J. Candes and T. Tao, “Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?” IEEE Trans. on Inf. Theory, vol. 52, no. 12, pp. 5406–5425, 2006.
[3] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inform. Theory, vol. 52, no. 4, pp. 1289–1306, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2006.871582
[4] C. Shannon, “Communication in the presence of noise,” Proceedings of the Institute of Radio Engineers (IRE), vol. 37, pp. 10–21, 1949.
[5] E. J. Candes and T. Tao, “Decoding by linear programming,” IEEE Trans. on Inf. Theory, vol. 51, no. 12, pp. 4203–4215, 2005.
[6] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathematique, vol. 346, no. 9-10, pp. 589 – 592, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1631073X08000964

[7] S. Foucart and M.-J. Lai, “Sparsest solutions of underdetermined linear systems via -minimization for,” Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 395 – 407, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520308000882

[8] S. Foucart, “A note on guaranteed sparse recovery via -minimization,” Applied and Computational Harmonic Analysis, vol. 29, no. 1, pp. 97 – 103, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520309001158

[9] T. Cai, L. Wang, and G. Xu, “Shifting inequality and recovery of sparse signals,” Signal Processing, IEEE Transactions on, vol. 58, no. 3, pp. 1300–1308, 2010.

[10] ———, “New bounds for restricted isometry constants,” Information Theory, IEEE Transactions on, vol. 56, no. 9, pp. 4388–4394, 2010.

[11] Q. Mo and S. Li, “New bounds on the restricted isometry constant,” Applied and Computational Harmonic Analysis, vol. 31, no. 3, pp. 460 – 468, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520311000625

[12] A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, “Approximation of functions over redundant dictionaries using coherence,” in Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms, ser. SODA ’03. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2003, pp. 243–252. [Online]. Available: http://dl.acm.org/citation.cfm?id=644108.644149

[13] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” USC, Tech. Rep. 04:06, 2004.

[14] J. Tropp, “Greed is good: algorithmic results for sparse approximation,” Information Theory, IEEE Transactions on, vol. 50, no. 10, pp. 2231–2242, 2004.

[15] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” IEEE Trans. Inform. Theory, vol. 52, no. 1, pp. 6–18, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2005.860430

[16] D. Needell and J. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” Applied and Computational Harmonic Analysis, vol. 26, no. 3, pp. 301 – 321, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520308000638

[17] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” Information Theory, IEEE Transactions on, vol. 55, no. 5, pp. 2230–2249, 2009.

[18] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” Selected Topics in Signal Processing, IEEE Journal of, vol. 4, no. 2, pp. 310–316, 2010.

[19] M. Davenport and M. Wakin, “Analysis of orthogonal matching pursuit using the restricted isometry property,” Information Theory, IEEE Transactions on, vol. 56, no. 9, pp. 4395–4401, 2010.

[20] S. Huang and J. Zhu, “Recovery of sparse signals using omp and its variants: convergence analysis based on rip,” Inverse Problems, vol. 27, no. 3, 2011.

[21] E. Liu and V. Temlyakov, “The orthogonal super greedy algorithm and applications in compressed sensing,” Information Theory, IEEE Transactions on, vol. 58, no. 4, pp. 2040–2047, 2012.

[22] Q. Mo and Y. Shen, “A remark on the restricted isometry property in orthogonal matching pursuit,” Information Theory, IEEE Transactions on, vol. 58, no. 6, pp. 3654–3656, 2012.

[23] H. Rauhut, “Compressive sensing and structured random matrices,” in Theoretical Foundations and Numerical Methods for Sparse Recovery, ser. Radon Series Comp. Appl. Math., M. Fornasier, Ed. deGruyter, 2010, vol. 9, pp. 1–92, corrected Version (April 4, 2011): Noncommutative Khintchine inequality for Rademacher chaos (Theorem 6.22) corrected, Proof of Theorem 9.3 corrected. Proof of Lemma 8.2 corrected.

[24] V. Temlyakov, “Weak greedy algorithms["this research was supported by national science foundation grant dms 9970326 and by ons grant n000149611003.” Advances in Computational Mathematics, vol. 12, no. 2-3, pp. 213–227, 2000.

[25] R. Maleh, “Improved RIP Analysis of Orthogonal Matching Pursuit,” ArXiv e-prints, Feb. 2011.

[26] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” Information Theory, IEEE Transactions on, vol. 47, no. 7, pp. 2845 –2862, nov 2001.

[27] M. Yang, “Greedy algorithms in approximation theory and compressed sensing,” Ph.D. dissertation, University of South Carolina, Columbia, SC, USA, 2011, aAI3469193.
[28] V. Temlyakov, “Relaxation in greedy approximation,” *Constructive Approximation*, vol. 28, no. 1, pp. 1–25, 2008. [Online]. Available: [http://dx.doi.org/10.1007/s00365-006-0652-5](http://dx.doi.org/10.1007/s00365-006-0652-5)

[29] V. N. Temlyakov, “Greedy approximation,” *Acta Numer.*, vol. 17, pp. 235–409, 2008. [Online]. Available: [http://dx.doi.org/10.1017/S0962492906380014](http://dx.doi.org/10.1017/S0962492906380014)

[30] V. Temlyakov, *Greedy Approximation*. Cambridge University Press, 2011.