Signal Recovery from Random Measurements via Extended Orthogonal Matching Pursuit

Sujit Kumar Sahoo, *Member, IEEE* and Anamitra Makur, *Senior Member, IEEE*

***Abstract*—Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP) are two well-known recovery algorithms in compressed sensing. To recover a d-dimensional m-sparse signal with high probability, OMP needs $O(m \ln d)$ number of measurements, whereas BP needs only $O\left(m \ln \frac{d}{m}\right)$ number of measurements. In contrary, OMP is a practically more appealing algorithm due to its superior execution speed. In this piece of work, we have proposed a scheme that brings the required number of measurements for OMP closer to BP. We have termed this scheme as OMP$_\alpha$, which runs OMP for $(m + \lfloor \alpha m \rfloor)$-iterations instead of m-iterations, by choosing a value of $\alpha \in [0, 1]$. It is shown that OMP$_\alpha$ guarantees a high probability signal recovery with $O\left(m \ln \frac{d}{\lfloor \alpha m \rfloor + 1}\right)$ number of measurements. Another limitation of OMP unlike BP is that it requires the knowledge of m. In order to overcome this limitation, we have extended the idea of OMP$_\alpha$ to illustrate another recovery scheme called OMP$_\infty$, which runs OMP until the signal residue vanishes. It is shown that OMP$_\infty$ can achieve a close to ℓ_0-norm recovery without any knowledge of m like BP.**

I. Introduction

Compressed sensing (CS) means acquiring/measuring the sparse signals from a limited number of linear projections at a subNyquist rate. It is a growing field of interest for researchers [1]. Through N linear projections $\boldsymbol{v} \in \mathbb{R}^N$, CS measures a d-dimensional real valued sparse signal $\boldsymbol{s} \in \mathbb{R}^d$, where $d \gg N$. CS stacks N projection vectors to form a measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{N \times d}$, and that makes

$$\boldsymbol{v} = \boldsymbol{\Phi s}.$$

The core idea of CS relies on the fact that the measured signal $\boldsymbol{s}$ is sparse, i.e. $\|\boldsymbol{s}\|_0 \ll d$. CS can also be extended to those signals which are sparse (compressible) in some basis or frame.

There are two basic problems in CS. The first one is to find a $\boldsymbol{\Phi}$ that ensures every m-sparse signal (i.e. $\|\boldsymbol{s}\|_0 = m$) has unique measurements, so that their unique reconstruction is possible. The following theorem gives an example of an admissible $\boldsymbol{\Phi}$.

Theorem 1 (Theorem 1 of [2])**.** *Let $N \geq \mathrm{C}_1 m \ln \frac{d}{m}$, and $\boldsymbol{\Phi}$ has $N \times d$ Gaussian i.i.d entries. The following statement is true with probability exceeding $1 - e^{-\mathrm{c}_1 N}$. It is possible to reconstruct every m-sparse signals $\boldsymbol{s} \in \mathbb{R}^d$ from the data $\boldsymbol{v} = \boldsymbol{\Phi s}$.*

Here C_1 and c_1 are positive constants. In order to bring generality, $\boldsymbol{\Phi}$ is usually quantified using the Restricted Isometry Property (RIP) [3]. Any matrix $\boldsymbol{\Phi}$ satisfies RIP of order m, if there exists a constant $0 \leq \delta_m < 1$ for which the following statement holds $\forall \|\boldsymbol{s}\|_0 \leq m$.

$$(1 + \delta_m)\|\boldsymbol{s}\|_2^2 \geq \|\boldsymbol{\Phi s}\|_2^2 \geq (1 - \delta_m)\|\boldsymbol{s}\|_2^2 \qquad \text{(RIP)}$$

In other words, any combination of m or less columns from $\boldsymbol{\Phi}$ will form a well conditioned submatrix. Therefore, $\boldsymbol{\Phi}$ needs to satisfying a RIP of order $2m$ to have the unique measurements for an m-sparse signal. Theorem 1 implies, $N = O\left(m \ln \frac{d}{m}\right)$ ensures RIP of order $2m$ for Gaussian measurement matrix.

The second problem in CS is to find a suitable algorithm that can exactly recover any sparse signal from its unique measurements,

$$(\ell_0) \qquad \hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_0 \text{ such that } \boldsymbol{v} = \boldsymbol{\Phi s}.$$

Our paper focuses on the second problem, where typically two major questions need to be answered.

1) Knowing that the measured signal $\boldsymbol{s}$ is sparse, i.e. $\|\boldsymbol{s}\|_0 \ll d$, can an algorithm reconstruct it exactly?

2) How many measurements are necessary for the algorithm to work?

Solution to (ℓ_0) is combinatorial in nature, where we have to find a sparse solution to an undetermined linear system of equations. Instead of exhaustively searching for the sparsest solution among all possible solutions, there exist many efficient techniques to solve (ℓ_0). Two broad classes of such techniques are iterative greedy pursuit [4], [5], [6], and convex relaxation [7], [8]. The convex relaxation technique has gained more importance in comparison to greedy pursuit because of two reasons. First, its theoretical recovery performance is much better than greedy pursuits. Second, it does not need the knowledge of sparsity m unlike greedy pursuit.

The greedy pursuits iteratively identify the nonzero indices of $\boldsymbol{s}$. One of the fundamental greedy pursuit techniques is Orthogonal Matching Pursuit (OMP) [5]. It minimizes the ℓ_2 norm of the residue by selecting one atom in each iteration, where atoms refer to the columns of the measurement matrix $\boldsymbol{\Phi}$ (i.e. $\varphi_j \in \mathbb{R}^N$). Therefore, for CS recovery of m-sparse signals, the usual scheme is to run OMP for m iterations. Some impressive theoretical guarantees for the OMP scheme have been established in [9], [10], [11]. The best among them shows, OMP can recover m-sparse signals exactly with high probability, when $N = O(m \ln d)$ [2].

The authors are with School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: sujit@pmail.ntu.edu.sg, eamakur@ntu.edu.sg).

This work is partially supported by Research Funding MOE2014-T2-1-079.

In contrast, the convex relaxation technique works by relaxing the problem (ℓ_0) as follows.

$$(\ell_1) \qquad \hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_1 \text{ such that } \boldsymbol{v} = \boldsymbol{\Phi s},$$

which is well know as Basis Pursuit (BP). It has been demonstrated that BP can recover any m-sparse signal if the measurements are unique [12], [13], [14]. This implies, BP only requires $N = O\left(m \ln \frac{d}{m}\right)$ for signal recovery in the case of Gaussian measurement matrices (Theorem 1).

However, in practice, BP is a computationally demanding technique, which requires $O\left(N^2 d^{3/2}\right)$ number of floating point operations [15]. In contrast, OMP requires $O(mNd)$ number of floating point operations [16]. The greedy pursuits are faster, and can be useful for large scale CS problems. Therefore, many variants of OMP have been proposed in recent years to achieve the benchmark performance of BP, e.g. regularized OMP [17], stagewise OMP [18], backtracking based adaptive OMP [19], etc. However, a well known behavior of standard OMP still remains unexplored. Experiments suggest OMP can produce superior result by going beyond m-iterations [20, chapter 8, footnote 6]. Some theoretical works on uniform signal recovery using OMP advocate going beyond m-iterations [21], [22], [23]. Using RIP, [21] analytically shows that any m-sparse signal can exactly be recovered in $O\left(m^{1.2}\right)$ iterations of OMP, if $N = O\left(m^{1.6} \log d\right)$. Along the same line of analysis, [22] shows that any m-sparse signal can exactly be recovered in $30m$ iterations of OMP, if $N = O(m \log d)$. An improvement is claimed in [23], which shows any m-sparse signal can exactly be recovered in $\lceil 2.8m \rceil$ iterations of OMP, if $N = O(m \log d)$. All these papers converge to a common conclusion that a minimum of $N = O(m \log d)$ is needed for CS recovery using OMP.

The aim of this article is to reduce the required N by OMP from $O(m \ln d)$ to $O\left(m \ln \frac{d}{m}\right)$ with few additional iterations beyond m. Therefore, we propose to run OMP for $m + \lfloor \alpha m \rfloor$ iterations, where $\alpha \in [0, 1]$. We refer this extended run of OMP as OMP$_\alpha$, and we analyze its recovery performance for noise free measurements. The result of our analysis is the following theorem.

Theorem 2 (OMP with Admissible Measurements)**.** *Fix $\alpha \in [0, 1]$, and choose $N \geq \mathrm{C}_0 m \ln \frac{d}{\lfloor \alpha m \rfloor + 1}$. Suppose that* $\mathbf{s}$ *is an arbitrary m-sparse signal in $\mathbb{R}^d$, and draw a random $N \times d$ admissible measurement matrix $\boldsymbol{\Phi}$ independent from the signal. Given the data $\boldsymbol{v} = \boldsymbol{\Phi s}$, OMP can reconstruct the signal with probability exceeding $1 - e^{-\mathrm{c}_0 \frac{N}{m}(\lfloor \alpha m \rfloor + 1)}$ in at most $m + \lfloor \alpha m \rfloor$ iterations.*

Here C_0 and c_0 are positive constants, and the admissible measurement matrix is defined in section IV.A. The proof of the theorem is in Section IV of the paper. Going further with the analysis of OMP$_\alpha$, we have illustrated another recovery scheme called OMP$_\infty$, which runs OMP until the signal residue vanishes. It is shown that OMP$_\infty$ can achieve a close to ℓ_0-norm recovery without any knowledge of m like BP. OMP$_\infty$ can be considered as the first OMP scheme for CS recovery that works without the knowledge of sparsity m.

Outline of the paper: The standard scheme of OMP for CS recovery is discussed in section II. The proposed scheme of extended OMP (OMP$_\alpha$) for CS recovery is elaborated in section III. The recovery performance of OMP$_\alpha$ is analyzed in section IV. The scheme of sparsity unaware OMP (OMP$_\infty$) for CS recovery is demonstrated in section V. The theoretical CS recovery performances of OMP, OMP$_\alpha$, and OMP$_\infty$ are experimentally validated in section VI. The paper ends with a brief discussion in section VII.

Notations: Let $\boldsymbol{\Phi}$, $\boldsymbol{s}$ are as defined, and $I \subset \{1, 2, \ldots, d\}$. The matrix $\boldsymbol{\Phi}_I \in \mathbb{R}^{N \times |I|}$ consists of the columns of $\boldsymbol{\Phi}$ with indices $i \in I$, and $\boldsymbol{s}_I \in \mathbb{R}^{|I|}$ consists of the components of $\boldsymbol{s}$ indexed by $i \in I$. We also denote $(.)^T$ and $(.)^\dagger$ for transpose and Moore-Penrose pseudo-inverse respectively.

II. OMP FOR CS RECOVERY

In the problem of CS recovery using OMP, it is known a priori that the measured signal $\boldsymbol{s}$ is m-sparse, which means $\boldsymbol{s}$ has non-zero entries only at m unknown indices. Let's define the unknown support of $\boldsymbol{s}$ as I, and $\|\boldsymbol{s}\|_0 = |I| = m$. We refer to the atoms φ_j corresponding to these indices $j \in I$ as correct atoms, and rest $\varphi_j : j \notin I$ as wrong atoms. OMP identifies I by selecting one candidate index in each iteration. The detailed steps are describe in the following algorithm.

Algorithm 1 (OMP for CS Recovery)**.**
Input:

- *measurement matrix* $\boldsymbol{\Phi} \in \mathbb{R}^{N \times d}$
- *measurement* $\boldsymbol{v} \in \mathbb{R}^N$
- *maximum iterations* $t_{\max} = m$

Output:

- *signal estimation* $\hat{\boldsymbol{s}}$
- *index set* Λ_t *containing elements from* $\{1, \ldots, d\}$
- *approximation* $\boldsymbol{a}_t \in \mathbb{R}^N$ *of measurement* $\boldsymbol{v}$
- *residual* $\boldsymbol{r}_t \in \mathbb{R}^N$

Procedure:

1) Initialize: *residual* $\boldsymbol{r}_0 = \boldsymbol{v}$*, index set* $\Lambda_0 = \emptyset$ *and iteration counter* $t = 0$*;*
2) *Increment* $t = t + 1$*;*
3) *choose the atom* $\lambda_t = \arg\max_{j=1,\ldots,d} |\langle \varphi_j, \boldsymbol{r}_{t-1} \rangle|$*;*
4) *Update* $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$*;*
5) *Update* $\boldsymbol{x}_t = \boldsymbol{\Phi}_{\Lambda_t}^\dagger \boldsymbol{v}$*;*
6) *Update* $\boldsymbol{a}_t = \boldsymbol{\Phi}_{\Lambda_t} \boldsymbol{x}_t, \quad \boldsymbol{r}_t = \boldsymbol{v} - \boldsymbol{a}_t$*;*
7) *Go to Step.2 if* $t < t_{\max} = m$*, else terminate;*
8) *The estimation* $\hat{\boldsymbol{s}}$ *for the signal* $\boldsymbol{s}$ *has nonzero elements at* Λ_t *and rest zeros, i.e.* $\hat{\boldsymbol{s}}_{\Lambda_t} = \boldsymbol{x}_t$*.*

OMP begins by initializing the residual to the input measurement vector $\boldsymbol{r}_0 = \boldsymbol{v}$, selected index set to empty set $\Lambda_0 = \emptyset$ and initial approximation to a null vector $\boldsymbol{a}_0 = 0$. At iteration t, OMP chooses a new index λ_t by finding the best atom matching with the residual, $\lambda_t = \arg\max_{j=1,\ldots,d} |\langle \varphi_j, \boldsymbol{r}_{t-1} \rangle|$, and updates the selected index set $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$. Then, OMP obtains the best t-term approximation $\boldsymbol{a}_t$ by a least-squares (LS) minimization. That is,

$$\boldsymbol{x}_t = \arg\min_x \|\boldsymbol{v} - \boldsymbol{\Phi}_{\Lambda_t} x\|,$$

which has a close form solution $\boldsymbol{x}_t = \boldsymbol{\Phi}_{\Lambda_t}^{\dagger}\boldsymbol{v}$, where $\boldsymbol{\Phi}_{\Lambda_t}^{\dagger} = \left(\boldsymbol{\Phi}_{\Lambda_t}^T\boldsymbol{\Phi}_{\Lambda_t}\right)^{-1}\boldsymbol{\Phi}_{\Lambda_t}^T$. LS procedure in OMP [5] brings a significant improvement in comparison to its parent algorithm, the matching pursuit (MP) [4].

In OMP, the residue $\boldsymbol{r}_{t-1}$ is always orthogonal to all the selected atoms $\boldsymbol{\Phi}_{\Lambda_{t-1}}$. That means the non-zero correlation $\langle\varphi_j, \boldsymbol{r}_{t-1}\rangle \neq 0$ will only occur for those atoms, which are not linear combinations of atoms in $\boldsymbol{\Phi}_{\Lambda_{t-1}}$. Thus at iteration t, OMP will select an atom φ_{λ_t} which is linearly independent from the previously selected atoms $\boldsymbol{\Phi}_{\Lambda_{t-1}}$, i.e. $\lambda_t \in \{j : \varphi_j \neq \boldsymbol{\Phi}_{\Lambda_{t-1}}x\}$. Therefore, the obvious choice for m-sparse signal recovery is to identify m correct atoms in m iterations of OMP [2]. The following proposition provides the recovery scenarios.

Proposition 1. *Take an arbitrary m-sparse signal $\boldsymbol{s}$ in $\mathbb{R}^d$, and let $\boldsymbol{\Phi}$ be any $N \times d$ measurement ensemble with the property that any $2m$ atoms are linearly independent. Given the data vector $\boldsymbol{v} = \boldsymbol{\Phi s}$,*

- *OMP for $t_{\max} < m$ will result in $\boldsymbol{r}_{t_{\max}} \neq 0$;*
- *OMP for $t_{\max} = m$ will result in $\boldsymbol{r}_{t_{\max}} \neq 0$, if $\hat{\boldsymbol{s}} \neq \boldsymbol{s}$;*
- *OMP for $t_{\max} = m$ will result in $\boldsymbol{r}_{t_{\max}} = 0$, if $\hat{\boldsymbol{s}} = \boldsymbol{s}$.*

Proof. It can easily be proved by contradiction. If signal residue vanishes i.e. $\boldsymbol{r}_{t_{\max}} = 0$ after any $t_{\max}$ iterations, that means we have a $t_{\max}$-sparse solution $\boldsymbol{v} = \boldsymbol{\Phi}\hat{\boldsymbol{s}}$. As there exists a generating m-sparse solution $\boldsymbol{s}$, it can be stated as $\boldsymbol{\Phi}(\hat{\boldsymbol{s}} - \boldsymbol{s}) = 0$, where the signal $(\hat{\boldsymbol{s}} - \boldsymbol{s})$ can have a maximum of $t_{\max} + m$ nonzero coefficients i.e. $\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_0 \leq t_{\max} + m$. For $t_{\max} \leq m$ it becomes contradictory, if $\boldsymbol{\Phi}$ has a property that any $2m$ columns of it are linearly independent. Hence it is proved that for such $\boldsymbol{\Phi}$, the signal residue of OMP will not vanish for $t_{\max} < m$, or $t_{\max} = m$ and $\hat{\boldsymbol{s}} \neq \boldsymbol{s}$. □

Note 1. *Proposition 7 of [2] considers random $\boldsymbol{\Phi}$ case with $t_{\max} = m$, whereas proposition 1 is a more general version.*

- Note that since RIP of order $2m$ ensures that any $2m$ columns of $\boldsymbol{\Phi}$ are linearly independent, any $\boldsymbol{\Phi}$ satisfying RIP of order $2m$ will satisfy the above proposition.
- Note that since Gaussian or Bernoulli measurement ensemble of any $2m$ columns are linearly independent with probability close to one for $N \geq 2m$ [24], [25], any $\boldsymbol{\Phi}$ made out of these random ensemble will satisfy the above proposition with a very high probability.

RIP of order $2m$ requires $N = O\left(m \ln \frac{d}{m}\right)$ in the case of random measurement matrices. It could be inferred from Proposition 1 that a RIP of order $2m$ is necessary for a unique solution $\boldsymbol{s}$ at $t_{\max} = m$. However, it can not guarantee that OMP will obtain a solution at $t_{\max} = m$. In order for that to happen with high probability, OMP needs $N = O(m \ln d) > O\left(m \ln \frac{d}{m}\right)$ measurements. This is because, in addition to RIP of order $2m$, the probability of selecting m correct atoms in m iterations decides the required N for OMP.

III. Extended OMP (OMP$_\alpha$) for CS Recovery

Identifying a m-sparse signal in only m selections is a sheer restriction to OMP, which has motivated many backtracking based greedy algorithms. These algorithms work with the main strategy of selecting more atoms and then tracking back to m atoms. However, we are more interested in the fundamental behavior of OMP when it selects more atoms.

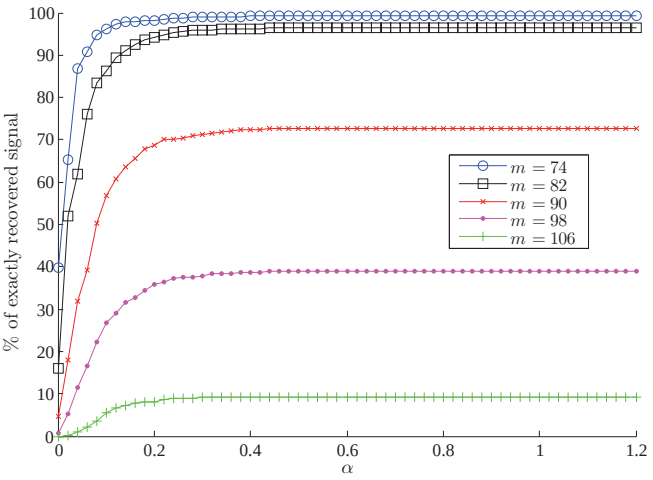


Fig. 1. The percentage of signal recovered in 1000 trials with increasing α, for various m-sparse signals in dimension $d = 1024$, from their $N = 256$ random measurements.

It can be observed that, when OMP has failed to pick m correct atoms out of $\boldsymbol{\Phi}_I$ in m iterations, it has not reached a solution and $\boldsymbol{r}_m \neq 0$. However, if we extend the iteration beyond m, then the chances of selecting m correct atoms will increase. Even though there are no published experimental results, this scenario is well known to the researchers working on greedy pursuits [20, chapter 8, footnote 6]. Recent attempts of OMP with extended run can be found in [21], [22], [23]. As discussed in the introduction, all of them have converged to a common conclusion that a minimum of $N = O(m \log d)$ is needed for CS recovery using OMP.

However, we are aiming at recovery of a m-sparse signal from $O\left(m \ln \frac{d}{m}\right)$ noise free measurements, by running few additional iterations of OMP beyond m. We have linearly extended the run of OMP beyond m iterations, and analyze it along the line of [2]. We propose to run OMP for $t_{\max} = m + \lfloor \alpha m \rfloor$ iterations, which is referred as OMP$_\alpha$ here onwards, where $\alpha \in [0, 1]$. This extended run may increase the computational cost of OMP only by a factor $1 + \alpha$, but it will still be of order $O(mNd)$. The proposed extended OMP algorithm is the following.

Algorithm 2 (OMP$_\alpha$ for CS Recovery)**.** *This algorithm is same as Algorithm 1 (OMP for CS recovery) except the following change in step 7:*

7) *Go to Step.2 if $t < m + \lfloor \alpha m \rfloor$, else terminate;*

By allowing an additional selection of $\lfloor \alpha m \rfloor$ atoms, we have increased the chance of acquiring m correct atoms. Thus, the conventional use of OMP for CS recovery can be viewed as a limiting case of OMP$_\alpha$ where $\alpha = 0$. By using its orthogonality property, and RIP of the sensing matrix, the following proposition shows how OMP$_\alpha$ can identify the m correct atoms from the $m + \lfloor \alpha m \rfloor$ selections.

Proposition 2. *Take an arbitrary m-sparse signal $\boldsymbol{s} \in \mathbb{R}^d$,*

and let $\mathbf{\Phi}$ be an $N \times d$ measurement ensemble satisfying RIP of order $m + \lfloor \alpha m \rfloor$. Given the data vector $\boldsymbol{v} = \mathbf{\Phi}\boldsymbol{s}$;

(S) *OMP$_\alpha$ will successfully identify any m-sparse signal $\boldsymbol{s}$, and $\boldsymbol{r}_{m+\lfloor \alpha m \rfloor} = 0$, if $I \subseteq \Lambda_{m+\lfloor \alpha m \rfloor}$,*

(F) *OMP$_\alpha$ will fail to identify any m-sparse signal $\boldsymbol{s}$, irrespective of $\boldsymbol{r}_{m+\lfloor \alpha m \rfloor}$, if $I \nsubseteq \Lambda_{m+\lfloor \alpha m \rfloor}$.*

Proof. At t^{th} iteration, OMP$_\alpha$ will find a t-term least square approximation $\hat{\boldsymbol{s}}_{\Lambda_t} = \mathbf{\Phi}_{\Lambda_t}^{\dagger}\boldsymbol{v}$. The best least square approximation for any linear system is the exact solution, leading to $\boldsymbol{a}_t = \mathbf{\Phi}\hat{\boldsymbol{s}} = \boldsymbol{v} \implies \boldsymbol{r}_t = 0$, which can only be possible if $\boldsymbol{v}$ lies in the column space $\mathbf{R}(\mathbf{\Phi}_{\Lambda_t})$. Since $I \subseteq \Lambda_{m+\lfloor \alpha m \rfloor}$ and $\boldsymbol{v} \in \mathbf{R}(\mathbf{\Phi}_I)$, which implies $\boldsymbol{v} \in \mathbf{R}(\mathbf{\Phi}_{\Lambda_{m+\lfloor \alpha m \rfloor}})$, the obtained $(m + \lfloor \alpha m \rfloor)$-term solution is exact, i.e. $\boldsymbol{v} = \mathbf{\Phi}\hat{\boldsymbol{s}}$. However, this makes $\mathbf{\Phi}(\hat{\boldsymbol{s}} - \boldsymbol{s}) = 0$, which implies that $\mathbf{\Phi}$ may contain $m+\lfloor \alpha m \rfloor$ or less number of linearly dependent atoms, because $\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_0 \leq m + \lfloor \alpha m \rfloor$. It becomes contradictory since $\mathbf{\Phi}$ satisfies RIP of order $m + \lfloor \alpha m \rfloor$. Therefore, $\mathbf{\Phi}(\hat{\boldsymbol{s}} - \boldsymbol{s}) = 0$ implies $\hat{\boldsymbol{s}} = \boldsymbol{s}$, and OMP$_\alpha$ successfully identifies the s-sparse signal.

Conversely, $I \nsubseteq \Lambda_{m+\lfloor \alpha m \rfloor} \implies \mathbf{R}(\mathbf{\Phi}_I) \nsubseteq \mathbf{R}(\mathbf{\Phi}_{\Lambda_{m+\lfloor \alpha m \rfloor}})$, then $\hat{\boldsymbol{s}}_{\Lambda_{m+\lfloor \alpha m \rfloor}}$ will either produce a $(m + \lfloor \alpha m \rfloor)$-term least square solution leading to signal residue $\boldsymbol{r}_{m+\lfloor \alpha m \rfloor} = 0$, or a $(m + \lfloor \alpha m \rfloor)$-term least square approximation with signal residue $\boldsymbol{r}_{m+\lfloor \alpha m \rfloor} \neq 0$. In either case OMP$_\alpha$ has failed to identify the exact m-term solution using columns of $\mathbf{\Phi}_I$. □

The event (S) stands for successful recovery in Proposition 2, which is a super set to the event of success in standard OMP. It is intuitive that the occurrence of event (S) has a higher probability for $\alpha > 0$ than for $\alpha = 0$. In order to see the behavior of event (S), we have shown an empirical observation of probability vs α in Fig. 1, which shows the increase in probability of recovery with α. In the next section Theorem 2 is proved by deriving the probability of event (F).

IV. Analysis of OMP$_\alpha$

In this section, we want to find the condition on N to recover a m-sparse signal with high probability using OMP$_\alpha$. In other words, we want to arrive at Theorem 2. The uniqueness of the measurement is a prerequisite for any recovery algorithm. Therefore $\mathbf{\Phi}$ is required to satisfy RIP of order $2m$ for $m \in (0, d/2)$. Since OMP$_\alpha$ requires RIP of order $m + \lfloor \alpha m \rfloor$ to function as a recovery algorithm for a m-sparse signal, we restrict α to the range $[0, 1]$. It is because α may be as large as 1 without requiring higher order of RIP than the unique measurement condition. In the following subsection, we define the Admissible Measurements which is the prerequisite for Theorem 2. In the succeeding subsection, we brief the approach to the proof of Theorem 2. Then we detail the events of OMP$_\alpha$, proof of the Theorem 2, and OMP as a special case in the subsequent subsections.

A. Admissible Measurements

The properties of admissible measurement matrices are based on Gaussian/Bernoulli sensing matrices. We will use the properties of these admissible matrices to compute OMP$_\alpha$'s probabilities of failure.

Matrices $\mathbf{\Phi} \in \mathbb{R}^{N \times d}$ with entries $\mathbf{\Phi}(i, j)$ as i.i.d. Gaussian random variable $\mathbb{N}(0, \frac{1}{\sqrt{N}})$ or i.i.d. Bernoulli random variable with sample space$\{\frac{1}{\sqrt{N}}, -\frac{1}{\sqrt{N}}\}$ are considered to be good choices for measurement matrix. These matrices are known to satisfy RIP of order $2m$ [3]. Apart from this, four other useful properties of $\mathbf{\Phi}$ are the following.

(P_0) *Independence*: Columns of $\mathbf{\Phi}$ are statistically independent.

(P_1) *Normalized*: $\forall j \quad \mathbb{E}[\|\varphi_j\|_2^2] = 1$

(P_2) *Correlation*: Let $\mathbf{u}$ be a vector whose ℓ_2 norm $\|\boldsymbol{u}\|_2 = 1$, and φ be a column of $\mathbf{\Phi}$ independent of $\mathbf{u}$. Then, for any $\varepsilon > 0$, the probability

$$\mathbb{P}\left\{|\langle \varphi, \boldsymbol{u} \rangle| \geq \varepsilon\right\} \leq 2e^{-\mathrm{c}_2 \varepsilon^2 N}.$$

The above inequality can easily be verified from the tail bound of any probability distribution (Gaussian and Bernoulli).

(P_3) *Bounded singular value*: For a given $N \times m$ submatrix $\mathbf{\Phi}_I$ from $\mathbf{\Phi}$, the singular values $\sigma(\mathbf{\Phi}_I)$ satisfy,

$$\mathbb{P}\left\{\sigma(\mathbf{\Phi}_I) \geq (1 - \delta_m)\right\} \geq 1 - e^{-\mathrm{c}_1 N}$$

where $0 \leq \delta_m < 1$. This is equivalent to stating that for any vector $\mathbf{x}$,

$$\mathbb{P}\left\{\|\mathbf{\Phi}_I \mathbf{x}\|_2^2 \geq (1 - \delta_m)\|\mathbf{x}\|_2^2\right\} \geq 1 - e^{-\mathrm{c}_1 N},$$

which is based on the RIP property of Gaussian and Bernoulli measurement matrices.[1]

B. Approach to the proof of Theorem 2

The requirement of N for OMP$_\alpha$ is decided by the probability of achieving the condition (S) of Proposition 1, which is further to RIP of order $2m$. Basically, the condition (S) stands for the success scenario of OMP$_\alpha$. Therefore, we need to derive the expression for probability of success in terms of N, m and d. Instead of analyzing the set of all possible events of success (E_{succ}) as in [2], we analyze the set of all possible events of failure (E_{fail}).

The success probability can be expressed as $\mathbb{P}(E_{\text{succ}}) = \mathbb{P}(E_{\text{succ}}, \Sigma) + \mathbb{P}(E_{\text{succ}}, \Sigma^c)$, where the conditional event Σ means, $\mathbf{\Phi}$ satisfies RIP of order $2m$. For Gaussian and Bernoulli measurement matrices

$$\mathbb{P}(\Sigma) \geq 1 - e^{-\mathrm{c}_1 N},$$

when $N \geq \mathrm{C}_1 m \ln \frac{d}{m}$ [3]. It can be stated that

$$\begin{aligned}\mathbb{P}(E_{\text{succ}}) \geq \mathbb{P}(E_{\text{succ}}, \Sigma) &= \mathbb{P}(\Sigma)\,(E_{\text{succ}}|\Sigma) \\ &= \mathbb{P}(\Sigma)\,(1 - \mathbb{P}(E_{\text{fail}}|\Sigma)).\end{aligned} \tag{1}$$

Thus, a lesser value of $\mathbb{P}(E_{\text{fail}}|\Sigma)$ means a better chance of success. RIP of order m is sufficient for the analysis of $\mathbb{P}(E_{\text{fail}}|\Sigma)$. However, Σ is taken as RIP of order $2m$ to have the uniqueness of the measurement. This is also essential for OMP$_\alpha$ to function (see Proposition 2).

[1] Theorem 1 can be considered as an example of such RIP obeying measurement matrix.

C. Events of OMP$_\alpha$

OMP$_\alpha$ works by selecting the candidate atoms φ_j one after another based on their correlation with the residue $\boldsymbol{r}_{t-1}$. Let's partition the measurement matrix into two sets of atoms, i.e. $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_I, \boldsymbol{\Phi}_{I^c}]$, where $\boldsymbol{\Phi}_I \stackrel{\text{def}}{=} \{\varphi_j : j \in I\}$ is the set of correct atoms, and $\boldsymbol{\Phi}_{I^c} \stackrel{\text{def}}{=} \{\varphi_j : j \in I^c\}$ is set of the remaining atoms (also termed as wrong atoms). Using correlation of the partitioned $\boldsymbol{\Phi}$ it can be classified whether OMP$_\alpha$ will reliably select a correct atom from $\boldsymbol{\Phi}_I$ or a wrong atom from $\boldsymbol{\Phi}_{I^c}$.

Correct atom: $\iff \max_{j\in I^c} |\langle \varphi_j, \boldsymbol{r}_{t-1}\rangle| < \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty$.

Wrong atom: $\iff \exists_{j\in I^c} |\langle \varphi_j, \boldsymbol{r}_{t-1}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty$.

It is important to note that when we have $|\langle \varphi_j, \boldsymbol{r}_{t-1}\rangle| = \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty$, selections of both wrong and correct atoms are possible. In order to keep the analysis simple, we have classified this tie scenario as selection of wrong atoms.

In order to analyze the events, let's specify the outcome of a run of OMP$_\alpha$ as $\Lambda_{m+\lfloor \alpha m\rfloor} = \{\lambda_1, \lambda_2, \ldots, \lambda_{m+\lfloor \alpha m\rfloor}\}$, where $\lambda_t \in \{1, 2, \ldots, d\}$ denotes the index of the atom chosen in iteration t. Since the exact sequence of appearance of these atoms are not important in determining the success or failure, we only consider the set of indices $\{\lambda_t\}$. Let's define the set of correct selections as $J_{\mathrm{C}} = \{\lambda_t : \lambda_t \in I\}$, which means for these iterations $\max_{j\in I^c} |\langle \varphi_j, \boldsymbol{r}_{t-1}\rangle| < \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty$. Let's also define $J_{\mathrm{W}} = \{\lambda_t : \lambda_t \in I^c\}$, which in turn means that $\max_{j\in I^c} |\langle \varphi_j, \boldsymbol{r}_{t-1}\rangle| = |\langle \varphi_{\lambda_t}, \boldsymbol{r}_{t-1}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty$ denoting selection of a wrong atom. Using these sets we can explain the *Success* (S) and *Failure* (F) of the OMP$_\alpha$ algorithm.

(S) After $m + \lfloor \alpha m\rfloor$ steps if we have $|J_{\mathrm{C}}| = m$ and $|J_{\mathrm{W}}| = \lfloor \alpha m\rfloor$, then certainly $I \subseteq \Lambda_{m+\lfloor \alpha m\rfloor}$. Note that $\alpha = 0$ implies success in conventional OMP, while $\frac{1}{m} < \alpha \leq 1$ implies success in OMP$_\alpha$.

(F) After $m+\lfloor \alpha m\rfloor$ steps if we have $|J_{\mathrm{C}}| < m$, $\lfloor \alpha m\rfloor + 1 \leq |J_{\mathrm{W}}| \leq \lfloor \alpha m\rfloor + m$, Then $I \not\subset \Lambda_{m+\lfloor \alpha m\rfloor}$ (excluding tie scenario) and OMP$_\alpha$ has failed.

With our conservative definition of failure as described earlier, the event of all possible failures is defined as

$$E_{\text{fail}} \stackrel{\text{def}}{=} \bigcup_{k=\lfloor \alpha m\rfloor+1}^{\lfloor \alpha m\rfloor+m} \left\{ \bigcup_{|J_{\mathrm{W}}|=k} J_{\mathrm{W}} \right\} \tag{2}$$

and the complementary event of success is defined as E_{succ}.

D. Proof of Theorem 2

Let's now estimate the failure probability from equation (2) using union bound,

$$\begin{aligned}\mathbb{P}\left(E_{\text{fail}}\right) &\leq \sum_{k=\lfloor \alpha m\rfloor+1}^{\lfloor \alpha m\rfloor+m} \mathbb{P}\left\{ \bigcup_{|J_{\mathrm{W}}|=k} J_{\mathrm{W}} \right\} \\ &\leq \sum_{k=\lfloor \alpha m\rfloor+1}^{\lfloor \alpha m\rfloor+m} \binom{d-m}{k} \mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \right\} \end{aligned} \tag{3}$$

where $\bigcup_{|J_{\mathrm{W}}|=k} J_{\mathrm{W}}$ denotes all possible J_{W} having size k, and $J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k}$ denotes one such J_{W}. Due to the property (P_0), $\mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \right\}$ is same for any J_{W} having size k, and does not depend on the specific atomic indices in it.

$|J_{\mathrm{W}}| = k$ means, OMP$_\alpha$ has selected k *wrong atoms*, i.e. $\bigcap_{\lambda_t\in J_{\mathrm{W}}} |\langle \varphi_{\lambda_t}, \boldsymbol{r}_{t-1}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty$ irrespective of iteration of occurrence t. Property (P_0) states that φ_{λ_t} are independent, and we make a pessimistic assumption that each event of unreliable selection is independent of each other. Thus using (P_1) it can be stated that

$$\begin{aligned}\mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \right\} &= \mathbb{P}\left\{ \bigcap_{\lambda_t\in J_{\mathrm{W}}} |\langle \varphi_{\lambda_t}, \boldsymbol{r}_{t-1}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty \right\} \\ &\simeq \mathbb{P}^k\left\{ |\langle \varphi_{\lambda_t}, \boldsymbol{r}_{t-1}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty \right\} \\ &= \mathbb{P}^k\left\{ |\langle \varphi, \boldsymbol{r}_{t-1}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{r}_{t-1}\|_\infty \right\}\end{aligned}$$

since the probability on the right hand side is same for any $\varphi \in \boldsymbol{\Phi}_{I^c}$.

In order to simplify the derivation let's normalize the residue vector to $\boldsymbol{u} = \frac{\boldsymbol{r}_{t-1}}{\|\boldsymbol{r}_{t-1}\|_2}$, which makes $\|\boldsymbol{u}\|_2 = 1$. Normalizing $\boldsymbol{r}_{t-1}$ on both sides of the equation will not affect the probability estimation, thus

$$\mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \right\} = \mathbb{P}^k\left\{ |\langle \varphi, \boldsymbol{u}\rangle| \geq \|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{u}\|_\infty \right\}.$$

It is known that $\forall x \in \mathbb{R}^m$, $\|x\|_\infty \geq \frac{\|x\|_2}{\sqrt{m}}$. As $\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{u}$ is a m-dimensional vector, it is true that $\|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{u}\|_\infty \geq \frac{\|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{u}\|_2}{\sqrt{m}}$. Thus it can be stated that

$$\mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \right\} \leq \mathbb{P}^k\left\{ |\langle \varphi, \boldsymbol{u}\rangle| \geq \frac{\|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{u}\|_2}{\sqrt{m}} \right\}.$$

Since the left hand side event is a subset of the right hand side event, the upper bound on its probability will remain true for any given condition. By taking the conditional event as Σ and using property (P_3), we have $\|\boldsymbol{\Phi}_I^{\mathrm{T}} \boldsymbol{u}\|_2 \geq \sqrt{(1-\delta_m)}\|\boldsymbol{u}\|_2$, or

$$\mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \middle| \Sigma \right\} \leq \mathbb{P}^k\left\{ |\langle \varphi, \boldsymbol{u}\rangle| \geq \sqrt{\frac{(1-\delta_m)}{m}} \middle| \Sigma \right\}.$$

Thus by using the property (P_2) of sensing matrices, i.e. the Gaussian tail probability, it can be written that

$$\mathbb{P}\left\{ J_{\mathrm{W}}\big|_{|J_{\mathrm{W}}|=k} \middle| \Sigma \right\} \leq \left[2e^{-\mathrm{c}_2 \frac{(1-\delta_m)}{m} N} \right]^k. \tag{4}$$

Using this bound of the conditional failure probability of equation (4), the combination inequality $\binom{A}{B} \leq \left(\frac{eA}{B}\right)^B$, and equation (3), it can be written that

$$\begin{aligned}\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) &\leq \sum_{k=\lfloor \alpha m\rfloor+1}^{\lfloor \alpha m\rfloor+m} \left[\frac{e(d-m)}{k} . 2e^{-\mathrm{c}_2 \frac{(1-\delta_m)}{m} N} \right]^k \\ &= \sum_{k=\lfloor \alpha m\rfloor+1}^{\lfloor \alpha m\rfloor+m} e^{\left[\ln \frac{2e(d-m)}{k} - \mathrm{c}_2 \frac{(1-\delta_m)}{m} N\right] k}.\end{aligned}$$

Changing the variable $i = k - \lfloor \alpha m\rfloor$ and $\mathrm{c}_3 = \mathrm{c}_2(1-\delta_m)$,

$$\begin{aligned}\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) &\leq \sum_{i=1}^{m} e^{\left[\ln \frac{2e(d-m)}{\lfloor \alpha m\rfloor+i} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m\rfloor+i)} \\ &\leq m e^{\left[\ln \frac{2e(d-m)}{\lfloor \alpha m\rfloor+1} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m\rfloor+1)} \\ &= e^{\left[\ln \frac{2e(d-m)}{\lfloor \alpha m\rfloor+1} + \frac{\ln m}{\lfloor \alpha m\rfloor+1} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m\rfloor+1)}.\end{aligned} \tag{5}$$

It can be found from the appendix of this paper that $\frac{\ln m}{\lfloor \alpha m \rfloor + 1} \leq \ln \frac{2m}{\lfloor \alpha m \rfloor + 1}$, when $m \geq 1$ and $0 \leq \alpha \leq 1$. Thus, the above upper bound can be expressed as

$$\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \leq e^{\left[\ln \frac{4e(d-m)m}{(\lfloor \alpha m \rfloor + 1)^2} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m \rfloor + 1)}.$$

Using the fact $(d-m)m \leq d^2/4$, it can be stated that

$$\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \leq e^{\left[\ln \frac{ed^2}{(\lfloor \alpha m \rfloor + 1)^2} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m \rfloor + 1)}. \tag{6}$$

The dominant variable term absorbs the constant, hence it can be stated that $2 \ln \frac{d}{\lfloor \alpha m \rfloor + 1} + 1 \leq \mathrm{C}_4 \ln \frac{d}{\lfloor \alpha m \rfloor + 1}$. This gives

$$\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \leq e^{\left[\mathrm{C}_4 \ln \frac{d}{\lfloor \alpha m \rfloor + 1} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m \rfloor + 1)}.$$

Using (1) along with the upper bound for $\mathbb{P}\left(\Sigma\right)$, it can be said that OMP_α will succeed with probability

$$\mathbb{P}\left(E_{\text{succ}}\right) \geq 1 - e^{\left[\mathrm{C}_4 \ln \frac{d}{\lfloor \alpha m \rfloor + 1} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m \rfloor + 1)} - e^{-\mathrm{c}_1 N}.$$

The third term can be absorbed into the second term, by increasing C_4 and decreasing c_3 if necessary.

$$\mathbb{P}\left(E_{\text{succ}}\right) \geq 1 - e^{\left[\mathrm{C}_4 \ln \frac{d}{\lfloor \alpha m \rfloor + 1} - \mathrm{c}_3 \frac{N}{m}\right](\lfloor \alpha m \rfloor + 1)}$$

By taking $N \geq \mathrm{C}_0 m \ln \frac{d}{\lfloor \alpha m \rfloor + 1}$ for $\mathrm{C}_0 \geq \frac{\mathrm{C}_4}{\mathrm{c}_3}$, we can ensure that OMP_α will succeed with probability $\mathbb{P}\left(E_{\text{succ}}\right) \geq 1 - e^{-\mathrm{c}_0 \frac{N}{m}(\lfloor \alpha m \rfloor + 1)}$, where $\mathrm{c}_0 \geq \mathrm{c}_3 - \frac{\mathrm{C}_4}{\mathrm{C}_0}$. This proves Theorem 2.

The above result brings the number of measurements for BP and OMP to the same order, when $\alpha \to 1$. Our result is mostly inspired by Tropp and Gilbert's analysis of OMP for m-iterations [2], and it simplifies to their result when $\alpha = 0$. Similar to [2], our result also is valid for random independent atoms. In contrast, the result for BP shows uniform recovery of all sparse signals over a single set of random measurement vectors. Nevertheless, OMP remains a valuable tool along with its inherent advantages, which makes Theorem 2 more attractive.

E. Cross-validating OMP_α

Algorithm 1 (OMP for CS recovery) can be viewed as a limiting case of OMP_α, where the extended run factor $\alpha = 0$. Therefore, analysis of OMP_α at $\alpha = 0$ should converge to the analysis of OMP in [2]. In order to verify that let's evaluate $\mathbb{P}\left(E_{\text{fail}}|\Sigma\right)$ at $\alpha = 0$. It can be obtained by substituting $\alpha = 0$ in equation (5),

$$\begin{aligned}\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) &\leq e^{\left[\ln\{2e(d-m)\} + \ln m - \mathrm{c}_3 \frac{N}{m}\right]} \\ &\leq e^{\left[\ln\{2e(d-m)m\} - \mathrm{c}_3 \frac{N}{m}\right]}.\end{aligned}$$

Using the fact $(d-m)m \leq d^2/4$, it can be stated that

$$\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \leq e^{\left[\ln \frac{ed^2}{2} - \mathrm{c}_3 \frac{N}{m}\right]}.$$

The dominant variable term can absorb the constant, hence $2 \ln d + \ln \frac{e}{2} \leq \mathrm{C}_4 \ln d$. This gives

$$\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \leq e^{\left[\mathrm{C}_4 \ln d - \mathrm{c}_3 \frac{N}{m}\right]}.$$

Using (1) along with the upper bound for $\mathbb{P}\left(\Sigma\right)$, it can be said that OMP_α will succeed with probability

$$\mathbb{P}\left(E_{\text{succ}}\right) \geq 1 - e^{\left[\mathrm{C}_4 \ln d - \mathrm{c}_3 \frac{N}{m}\right]} - e^{-\mathrm{c}_1 N}.$$

The third term can be absorbed into the second term, by increasing C_4 and decreasing c_3 if necessary.

$$\mathbb{P}\left(E_{\text{succ}}\right) \geq 1 - e^{\left[\mathrm{C}_4 \ln d - \mathrm{c}_3 \frac{N}{m}\right]}$$

By taking $N \geq \mathrm{C}_0 m \ln d$ for $\mathrm{C}_0 \geq \frac{\mathrm{C}_4}{\mathrm{c}_3}$, we can ensure that OMP will succeed with probability $\mathbb{P}\left(E_{\text{succ}}\right) \geq 1 - e^{-\mathrm{c}_0 \frac{N}{m}}$, where $\mathrm{c}_0 \geq \mathrm{c}_3 - \frac{\mathrm{C}_4}{\mathrm{C}_0}$.

It serves as another validation of OMP_α, because the limiting result for $\alpha = 0$ coincides with the result of OMP in [2]. It again proves that OMP_α would require a reduced number of measurements for the same success probability.

F. OMP_α with less computation

While Algorithm 2 (OMP_α for CS Recovery) uses the looping condition as iteration $t < m + \lfloor \alpha m \rfloor$, some computations may be saved by using condition $t < m + \lfloor \alpha m \rfloor$ & $\boldsymbol{r}_t \neq 0$. The advantage is obvious under the success scenario (S), $I \subseteq \Lambda_t$, resulting in $\boldsymbol{r}_t = 0$. However, since other sparse solutions may exist, $\boldsymbol{r}_t$ may become 0 with $I \not\subseteq \Lambda_t$. With the $\boldsymbol{r}_t = 0$ based termination, the iterations will stop with this wrong solution. One may wonder if continuing the iterations might have allowed OMP_α to obtain the correct solution, or $I \subseteq \Lambda_{m+\lfloor \alpha m \rfloor}$. The following proposition shows that, once a wrong solution is found, we can never reach the correct solution by completing the $m + \lfloor \alpha m \rfloor$ iterations, since too many wrong atoms have already been selected.

Proposition 3. *Take an arbitrary m-sparse signal $\boldsymbol{s}$ in $\mathbb{R}^d$, let $\boldsymbol{\Phi}$ be an $N \times d$ measurement ensemble satisfying RIP of order $m + \lfloor \alpha m \rfloor$, and execute OMP_α with the data $\boldsymbol{v} = \boldsymbol{\Phi s}$. If OMP_α arrives at $\boldsymbol{r}_t = 0 : m < t < m + \lfloor \alpha m \rfloor$, and $I \not\subseteq \Lambda_t$, then it has already selected more than $\lfloor \alpha m \rfloor$ wrong atoms. Thus, by completing $m + \lfloor \alpha m \rfloor$ selections it will never achieve $I \subseteq \Lambda_t$.*

Proof. When the signal residue vanishes after t iterations (i.e. $\boldsymbol{r}_t = 0$), it means that we have obtained a t-sparse solution $\boldsymbol{v} = \boldsymbol{\Phi}\hat{\boldsymbol{s}}$. Let's assume that in this t-sparse solutions we have p such atoms which are not from $\boldsymbol{\Phi}_I$. As there exists a generating m-sparse solution $\boldsymbol{s}$ using atoms of $\boldsymbol{\Phi}_I$, it can be stated that $\boldsymbol{\Phi}(\hat{\boldsymbol{s}} - \boldsymbol{s}) = 0$, where the signal $(\hat{\boldsymbol{s}} - \boldsymbol{s})$ has $p + m$ nonzero coefficients i.e. $\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_0 = p + m$. It implies, $\boldsymbol{\Phi}$ contains $p + m$ linearly dependent atoms, which is only possible if $p > \lfloor \alpha m \rfloor$. It is because $\boldsymbol{\Phi}$ obeys RIP of order $m + \lfloor \alpha m \rfloor$. Hence it is proved that OMP_α has already selected more than $\lfloor \alpha m \rfloor$ wrong atoms. Thus, by completing $m + \lfloor \alpha m \rfloor$ selections it will never achieve $I \subseteq \Lambda_t$. □

We now argue that early termination does not affect the results in Theorem 2. The failure event analysis in equation (2) covers all possible events of wrong selection from $\lfloor \alpha m \rfloor + 1$ to $\lfloor \alpha m \rfloor + m$ wrong atoms. With early termination, one may argue that the algorithm selects only upto $\lfloor \alpha m \rfloor + m'$ wrong atoms, where $m' < m$. Even then, the proof of Theorem 2 will remain unaltered, because replacing $\frac{\ln m}{\lfloor \alpha m \rfloor + 1}$ with $\frac{\ln m'}{\lfloor \alpha m \rfloor + 1}$ in equation (5) will not affect the upper bound in equation (6).

V. Sparsity unaware OMP (OMP_∞) for CS recovery

The superior execution speed of OMP comes with two drawbacks in its present form of CS recovery. First, it needs more

number of measurements in comparison to BP for recovering the same signal. Second, it requires prior knowledge of the sparsity m, whereas no such information is needed for BP. Through the scheme of OMP$_\alpha$, we have brought down the gap between OMP and BP in terms of required N both in theory and practice. However, the requirement of prior knowledge of the sparsity in OMP still remains.

Since OMP$_\alpha$ runs for more than m iterations unlike OMP, if we get rid of the bound on the number of iterations in OMP$_\alpha$, prior knowledge of m is no longer required. The bound of $m + \alpha m$ iterations is only required to prove Theorem 2. More iterations may only improve the performance of OMP. Removing the iteration based termination condition from step 7 of Algorithm 2, we obtain the modified algorithm referred as OMP$_\infty$.

Algorithm 3 (OMP$_\infty$ for CS Recovery)**.** *This algorithm is the same as Algorithm 2 (OMP$_\alpha$ for CS recovery) except the following change in step 7:*

7) *Go to Step.2 if $\boldsymbol{r}_t \neq 0$, else terminate;*

Contrary to its name, the worst case scenario of OMP$_\infty$ is not infinite iterations but when it selects N linearly independent vectors that spans the whole $\mathbb{R}^N$ space to reach $\boldsymbol{r}_N = 0$. Thus, its worst case complexity is of order $O(N^2 d)$, which is still less than BP. Also, the probability bounds of Theorem 1 are equally applicable for OMP$_\infty$ based reconstruction for not only Gaussian but other admissible measurement matrices. Choosing $N \geq \mathrm{C}_1 m \ln \frac{d}{m}$, supposing that $\boldsymbol{s}$ is an arbitrary m-sparse signal in $\mathbb{R}^d$, drawing a random $N \times m$ admissible measurement matrix $\boldsymbol{\Phi}$ independent from the signal, and given the data $\boldsymbol{v} = \boldsymbol{\Phi s}$, OMP$_\infty$ can reconstruct the signal with probability exceeding $1 - e^{-\mathrm{c}_1 N}$, where C_1 and c_1 are positive constants. Due to its restriction of m iteration, the same is not true for earlier OMP based reconstruction.

In the following we present an analysis of OMP$_\infty$ performance. OMP$_\infty$ can be viewed as running $\lim_{\alpha \to \infty}$ OMP$_\alpha$. Consider an inadequate number of measurements N_0 for some sparsity m_0, and lets interpret the outcome with increasing α. It can be observed from equation (6) that the conditional failure probability $\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \approx 1$, till α reaches

$$\frac{1}{2}\left(\mathrm{c}_3 \frac{N_0}{m_0} - 1\right) > \ln \frac{d}{\lfloor \alpha m_0 \rfloor + 1}.$$

Thereafter, it will start decaying exponentially with α, which can be continuously approximated as

$$\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \leq e^{-\mathrm{c}_5\left(\alpha + \frac{1}{m_0}\right)N_0}.$$

Here $\mathrm{c}_5 = \mathrm{c}_3 - \frac{m_0}{N_0}\left(2 \ln \frac{d}{\lfloor \alpha m_0 \rfloor + 1} + 1\right)$. However, since $\mathbb{P}(E_{\text{succ}}, \Sigma^c) \to 0$ and may be ignored, the final probability of successful recovery of a sparse vector can be expressed as

$$\mathbb{P}\left(E_{\text{succ}}\right) \simeq \mathbb{P}\left(E_{\text{succ}}, \Sigma\right) = \mathbb{P}\left(\Sigma\right)\left(1 - \mathbb{P}\left(E_{\text{fail}}|\Sigma\right)\right).$$

While increasing α, we will achieve a point where $\mathbb{P}\left(E_{\text{fail}}|\Sigma\right) \to 0$, and the final success probability

$$\mathbb{P}\left(E_{\text{succ}}\right) \simeq \mathbb{P}\left(\Sigma\right).$$

It shows that $\mathbb{P}\left(E_{\text{succ}}\right)$ will increase with α and it will meet $\mathbb{P}\left(\Sigma\right)$ asymptotically. Fig.1 is a nice illustration of this behavior, where percentage of signal recovered represents $\mathbb{P}\left(E_{\text{succ}}\right)$.

In other words, success of OMP$_\infty$ depends on $\mathbb{P}\left(\Sigma\right)$, the probability that $\boldsymbol{\Phi}$ obeys a RIP of order 2m. In the case of Gaussian and Bernoulli random matrices, RIP of order $2m$ holds for entire range of $m \in (0, d/2)$ with high probability exceeding $1 - e^{-\mathrm{c}_1 N}$, if $N \geq \mathrm{C}_1 m \ln \frac{d}{m}$.

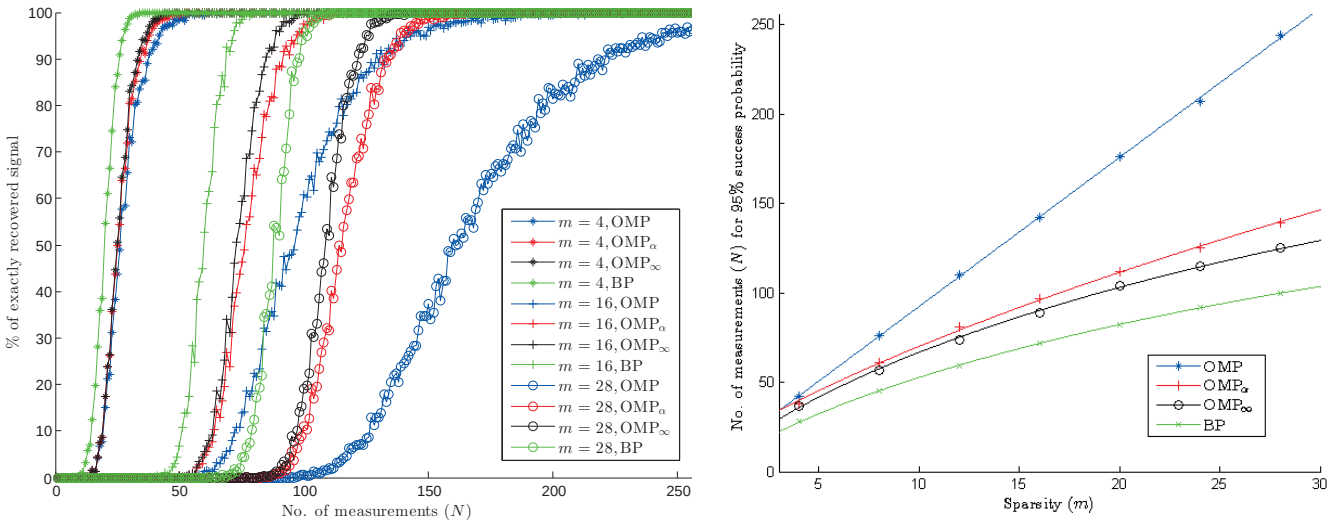
VI. Experiments

The proposed extension of OMP is validated in this section. It is experimentally illustrated that OMP$_\alpha$ has not only improved the performance of OMP but also it has been competitive to BP. As per Theorem 2, we validate the algorithm on random sensing matrices. The obtained results for Bernoulli ensemble are strikingly indifferent to Gaussian, thus we have only presented the results on Gaussian ensemble. The practical question is to determine how many measurements N are needed to recover an m-sparse signal in $\mathbb{R}^d$ with high probability. Thus the experimental set up is the following.

The probability of success is viewed as the percent of a m-sparse signal recovered successfully out of 1000 trials. Successful recovery implies that the distance between the original and recovered sparse signal is insignificant, i.e. $\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2 \leq 10^{-6}$. For each trial the m-sparse signal $\boldsymbol{s}$ is generated by setting nonzero values at m random locations of a d-dimensional null vector. The measurement matrix $\boldsymbol{\Phi}$ is constructed by generating $N \times d$ Gaussian random variables of parameters $(0, 1/\sqrt{N})$. The recovered signal $\hat{\boldsymbol{s}}$ is obtained performing BP, OMP, OMP$_\alpha$ and OMP$_\infty$ on the measurement $\boldsymbol{v} = \boldsymbol{\Phi s}$. Although it is possible to obtain different set of results in OMP$_\alpha$ by varying the extended run factor $0 < \alpha \leq 1$, the results presented here are for $\alpha = 0.25$.

The nonzero coefficients in $\boldsymbol{s}$ play an important role in the performance of matching based greedy algorithms from a practical point of view. The measurement matrix $\boldsymbol{\Phi}$ is obtained using zero mean random variables. Thus, when all the nonzero coefficients become equal, the measurement $\boldsymbol{v} = \boldsymbol{\Phi s}$ becomes the scaled sample mean of the random variables making it very close to zero i.e. $\boldsymbol{v} \to 0$. This scenario degrades the performance of the matching step of OMP depending on the precision of the device. Hence, all the results are obtained for this extreme scenario, when the sparse coefficients are set equal i.e. $\boldsymbol{s}_I = 1$ (same as the experimental setup in [2]). Signal dimension is taken as $d = 256$ and each m-sparse signal is recovered from the number of measurements starting with $N = 4$ to $N = 256$ in steps of 4. The percentage of successful trials is plotted against the number of measurements (N) in plot (A) of Fig.2.

With the same philosophy we might be interested to know, for a given sparsity level how many measurements will be needed to ensure a recovery with certain probability of success (for example 0.95 or 95%). Since the $\%$-success vs. N is monotonically increasing, the N at which we first achieved success rate of 95% can be obtained empirically. Plot (B) of Fig.2 shows the plot N vs. m for 95% success. In order to study the characteristic of N vs. m data points, a



(A) (B)

Fig. 2. (A) The percentage of 1000 input signals correctly recovered as a function of the sparsity level m for different numbers N of measurements in dimension $d = 256$. (B) The number N of measurements necessary to recover an m-sparse signal in dimension $d = 256$ at least 95% of the time.

curve fitting is done using Matlab toolbox. The results are tabulated in Table.I, which shows $O(m \ln d)$ nature of OMP and $O(m \ln \frac{d}{\alpha m+1})$ nature of OMP$_\alpha$, but $O(m \ln \frac{d}{m})$ nature of OMP$_\infty$ and BP.

In order to validate Theorem 2, we obtained the curve fitting result for OMP$_\alpha$ for $\alpha = 0, 1/16, 1/8, 1/4, 1/2$ in similar manner. However, we have increased the signal dimension to $d = 1024$ to acquire more integer points for better curve fitting. Fig. 3 shows a tight fitting of the curve $\mathrm{C}_0 m \ln \frac{d}{\alpha m+1} + \mathrm{C}_6$ with the obtained data points, and the values of C_0 and C_6 for various α are tabulated in Table. II.

TABLE I
LINEAR FITTING OF FIG. 2(B)

Algorithm	Expression
OMP	$1.504m \ln d + 9.0$
OMP$_\alpha$	$1.288m \ln (\frac{d}{0.25m+1}) + 14.87$
OMP$_\infty$	$1.962m \ln (\frac{d}{m}) + 3.134$
BP	$1.596m \ln (\frac{d}{m}) + 0.991$

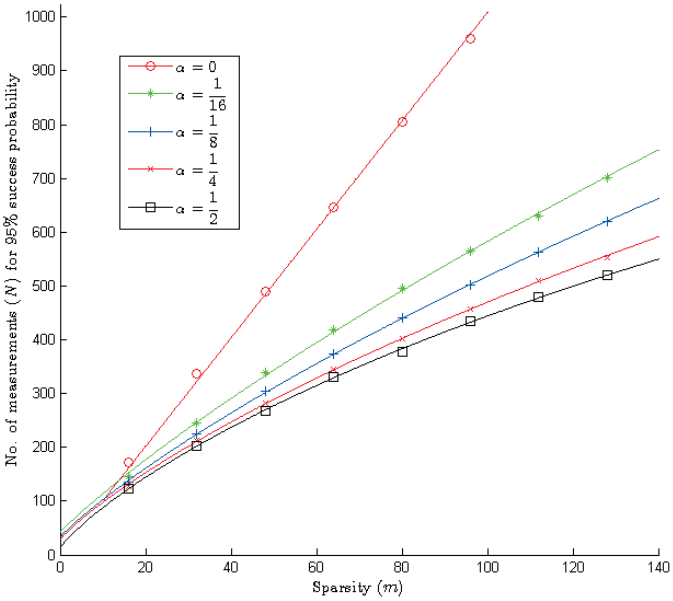


Fig. 3. The number N of measurements necessary to recover an m-sparse signal in dimension $d = 1024$ at least 95% of the time.

TABLE II
LINEAR FITTING OF $\mathrm{C}_0 m \ln \frac{d}{\alpha m+1} + \mathrm{C}_6$ IN FIG. 3

α	0	1/16	1/8	1/4	1/2
C_0	1.418	1.089	1.119	1.199	1.434
C_6	17.73	43.17	33.73	29.25	13.84

VII. DISCUSSIONS

Greedy pursuit is advantageous in terms of computational cost, which interests researchers to improve its performance towards the benchmark of convex relaxation (BP). The proposed OMP$_\alpha$ uses the orthogonality property of OMP and the probabilistic linear independences of random ensemble to enhance its performance. Its required number of measurements for high probability signal recovery follows a logarithmic trend like BP, instead of following a linear trend as OMP. Further, the proposed OMP$_\infty$ shows an overwhelming improvement in OMP by bringing it close to BP in terms of both required order of measurements and no prior knowledge of sparsity. The theoretical guarantee of OMP$_\alpha$ along with the obtained empirical results make OMP$_\alpha$ a more compelling algorithm.

Convex relaxation has rich varieties of results including the cases when the measured signal is not exactly sparse or is contaminated by noise. The results presented for OMP$_\alpha$ and OMP$_\infty$ are focused on strictly sparse signals. How these proposed schemes of OMP behave recovering the measurements contaminated by noise is an interesting direction to pursue.

REFERENCES

[1] E. Candes and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21 –30, march 2008.

[2] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655 –4666, dec. 2007.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, 2008.

[4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397 –3415, dec 1993.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conference Record of the Asilomar Conference on Signals, Systems & Computers*, vol. 1, 1993, pp. 40–44.

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

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[8] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.

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

[10] 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, sept. 2010.

[11] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4973 –4976, sept. 2012.

[12] M. Rudelson and R. Vershynin, "Geometric approach to error-correcting codes and reconstruction of signals," *International Mathematics Research Notices*, vol. 2005, no. 64, pp. 4019–4041, 2005.

[13] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[14] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, Dec 2006.

[15] N. Yurii and N. Arkadii, *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[16] B. Åke, *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, 1996.

[17] 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, april 2010.

[18] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1094 –1121, feb. 2012.

[19] H. Huang and A. Makur, "Backtracking-based matching pursuit method for sparse signal reconstruction," *Signal Processing Letters, IEEE*, vol. 18, no. 7, pp. 391 –394, july 2011.

[20] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[21] E. D. Livshits, "On the efficiency of the orthogonal matching pursuit in compressed sensing," *Sbornik: Mathematics*, vol. 203, no. 2, p. 183, 2012.

[22] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6215–6221, Sept 2011.

[23] J. Wang and B. Shim, "How many iterations are needed for the exact recovery of sparse signals using orthogonal matching pursuit?" [Online]. Available: http://arxiv.org/pdf/1211.4293v2.pdf

[24] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[25] J. Kahn, J. Komls, and E. Szemerdi, "On the probability that a random ± 1-matrix is singular," *Journal of the American Mathematical Society*, vol. 8, no. 1, pp. pp. 223–240, 1995.

APPENDIX

For an appropriate c_7,

$$\frac{\ln m}{\lfloor \alpha m \rfloor + 1} \leq \ln \frac{\mathrm{c}_7 m}{\lfloor \alpha m \rfloor + 1}, \tag{7}$$

where sparsity $m \geq 1$ and $0 \leq \alpha \leq 1$.

A. For $m = 1$

Let's substitute the limiting value $m = 1$ in inequality (7).

$$0 \leq \ln \frac{\mathrm{c}_7}{\lfloor \alpha \rfloor + 1} \implies \mathrm{c}_7 \geq \lfloor \alpha \rfloor + 1.$$

As $\alpha \leq 1$, inequality (7) will be true for $\mathrm{c}_7 \geq 2$.

B. For $m \geq 2$

The inequality (7) can be rearranged as the following.

$$\begin{aligned}
&\ln \frac{\lfloor \alpha m \rfloor + 1}{\mathrm{c}_7} \leq \left(1 - \frac{1}{\lfloor \alpha m \rfloor + 1}\right) \ln m \\
\implies &\log_m \frac{\lfloor \alpha m \rfloor + 1}{\mathrm{c}_7} \leq \left(1 - \frac{1}{\lfloor \alpha m \rfloor + 1}\right) \\
\implies &\frac{\lfloor \alpha m \rfloor + 1}{\mathrm{c}_7} \leq \frac{m}{m^{\frac{1}{\lfloor \alpha m \rfloor + 1}}} \\
\implies &\mathrm{c}_7 \geq \frac{(\lfloor \alpha m \rfloor + 1) m^{\frac{1}{\lfloor \alpha m \rfloor + 1}}}{m}
\end{aligned} \tag{8}$$

Interestingly, the condition on c_7 is a function of α and m, $f(m, \alpha) = \frac{(\alpha m + 1) m^{\frac{1}{\alpha m + 1}}}{m}$. For any give m, if we set

$$\mathrm{c}_7 \geq \max_{0 \leq \alpha \leq 1} f(m, \alpha) \tag{9}$$

inequality (7) would be valid for all range of $\alpha \in [0, 1]$. It can be seen that

$$\begin{aligned}
\frac{\partial f(m,\alpha)}{\partial \alpha} &= m^{\frac{1}{\alpha m + 1}} \left[1 - \frac{\ln m}{\alpha m + 1}\right] \\
&< 0 \text{ for } \alpha < \frac{\ln m - 1}{m} \\
&= 0 \text{ at } \alpha = \frac{\ln m - 1}{m} \\
&> 0 \text{ for } \alpha > \frac{\ln m - 1}{m}.
\end{aligned}$$

This implies, $f(m, \alpha)$ decreases with α until $\alpha = \frac{\ln m - 1}{m}$, and then increases. However, $f(m, \alpha)$ is a monotonically increasing function of α for $m < e$, because $\ln m < 1$ makes $\frac{\partial f(m,\alpha)}{\partial \alpha} > 0$ unconditionally. Hence,

$$\mathrm{c}_7 \geq \max\{f(m, 0), f(m, 1)\} = f(m, 1) \tag{10}$$

since

$$f(m, 1) = \left(1 + \frac{1}{m}\right) m^{\frac{1}{m+1}} \geq 1 = f(m, 0).$$

If we set

$$\mathrm{c}_7 \geq \max_{2 \leq m} f(m, 1) \tag{11}$$

inequality (7) would be valid for all $m \geq 2$. The derivative

$$\frac{\partial f(m, 1)}{\partial m} = \frac{(m+1) m^{\frac{1}{m+1}}}{m} \left[\frac{-\ln m}{(m+1)^2}\right] < 0$$

shows that $f(m,1)$ is a decreasing function of m. Hence,

$$\mathrm{c}_7 \geq \max_{2\leq m} f(m,1) = f(2,1) = \frac{3}{2}2^{\frac{1}{3}}.$$

However, the previously obtained condition $\mathrm{c}_7 \geq 2$ for the case of $m = 1$, is higher than $\frac{3}{2}2^{\frac{1}{3}}$. Therefore, it is proved that at $\mathrm{c}_7 = 2$ the inequality (7) holds for the entire range of m and α.

Anamitra Makur received the B.Tech. degree from the Indian Institute of Technology, Kharagpur, and the M.S. and Ph.D. degrees from the California Institute of Technology, Pasadena. He was with the Indian Institute of Science, Bangalore, until 2002 in various capacities, the most recent being a Professor. He also held visiting positions at University of California at Santa Barbara, University of Kaiserslautern, Griffith University at Brisbane, and Keio University at Yokohama. Since 2002, he has been an Associate Professor at Nanyang Technological University, Singapore. His current research interests include compressed sensing, multirate signal processing, signal/image/video compression, and image processing. Dr. Makur is the recipient of several awards, notable being the 1998 Young Engineer award from the Indian National Academy of Engineering, and the best paper award in the IEEE APCCAS 2006 conference. He was also the co-author in the student paper award of the IEEE ICASSP 2006 conference. He serves or has served as the Associate Editor of several journals, including the IEEE Transactions on Signal Processing.

Sujit Kumar Sahoo received the B.Tech. degree in electrical engineering in 2006 from the National Institute of Technology, Rourkela, India, and the Ph.D. degrees in electrical and electronic engineering in 2014 from the Nanyang Technological University, Singapore. From October 2006 to December 2007, he was a Software Engineer at the Sasken Communication Technologies Ltd., Bangalore, India. Since January 2012, he has been a Researcher at Nanyang Technological University, Singapore. His current research interests include sparse representation, compressed sensing, multirate signal processing, signal/image/video compression, and image processing. He serves as reviewer of several journals, including the IEEE Transactions on Signal Processing.