

EXACT SPARSE SIGNAL RECOVERY VIA ORTHOGONAL MATCHING PURSUIT WITH PRIOR INFORMATION

Jinming Wen *

College of Cyber Security
Jinan University, Guangzhou, China
E-mail: jinming.wen@mail.mcgill.ca

Wei Yu

Dept. of Electrical and Computer Engineering
University of Toronto, Toronto, Canada
E-mail: weiyu@comm.utoronto.ca

ABSTRACT

The orthogonal matching pursuit (OMP) algorithm is a commonly used algorithm for recovering K -sparse signals $\mathbf{x} \in \mathbb{R}^n$ from linear model $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a sensing matrix. A fundamental question in the performance analysis of OMP is the characterization of the probability that it can exactly recover \mathbf{x} for random matrix \mathbf{A} . Although in many practical applications, in addition to the sparsity, \mathbf{x} usually also has some additional property (for example, the nonzero entries of \mathbf{x} independently and identically follow the Gaussian distribution), none of existing analysis uses these properties to answer the above question. In this paper, we first show that the prior distribution information of \mathbf{x} can be used to provide an upper bound on $\|\mathbf{x}\|_1^2/\|\mathbf{x}\|_2^2$, and then explore the bound to develop a better lower bound on the probability of exact recovery with OMP in K iterations. Simulation tests are presented to illustrate the superiority of the new bound.

Index Terms— Exact sparse signal recovery, orthogonal matching pursuit (OMP), exact recovery probability

1. INTRODUCTION

In many applications, such as sparse activity detection [1], we need to reconstruct a K -sparse signal \mathbf{x} (i.e., \mathbf{x} has at most K nonzero entries) from linear measurements:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m \ll n$) is a random sensing matrix with independent and identically distributed (i.i.d.) Gaussian $\mathcal{N}(0, 1/m)$ entries and $\mathbf{y} \in \mathbb{R}^m$ is a given observation vector. Numerous sparse recovery algorithms have been developed to recover \mathbf{x} based on \mathbf{y} and \mathbf{A} [2–4]. Among them, greedy algorithms are very popular, especially when m, n and/or K are large, due to their low computational complexities. The orthogonal matching pursuit (OMP) algorithm [5], which is described in Algorithm 1, is a widely-used greedy algorithm due to its high efficiency and effectiveness [6].

A fundamental question in the analysis of OMP is the characterization of its exact recovery capability. To this end, numerous works studied the recovery performance of OMP (see, e.g., [7–13]). In particular, [14] develops a lower bound on the probability of exact recovery of K -sparse \mathbf{x} with K iterations of OMP. To better understand its recover capability, it is natural to ask whether this lower bound can be improved.

In many practical applications, in addition to sparsity, \mathbf{x} also has some other properties. For example, in wireless communication problems involving the Rayleigh channel model, the nonzero entries of \mathbf{x} independently and identically follow the standard Gaussian distribution $\mathcal{N}(0, 1)$ [1]. In speech communication [15] and audio source separation [16], \mathbf{x} has exponentially decaying property, i.e., \mathbf{x} is a K -sparse α -strongly-decaying signal. Intuitively, a larger variation in the magnitudes of the nonzero entries of \mathbf{x} would typically lead to better exact recovery performance of OMP in K iterations.

This paper aims to develop a theoretical framework to capture the dependence of the exact recovery performance of OMP on the disparity in the magnitudes of the nonzero entries of \mathbf{x} . Toward this end, we define the following measure of the disparity, in term of a function $\phi(t)$, such that

$$\|\mathbf{x}_S\|_1^2 \leq \phi(|S|)\|\mathbf{x}_S\|_2^2 \quad (2)$$

for any set $S \subseteq \Omega$, where Ω is the support of \mathbf{x} , $|S|$ denotes the number of elements of S and $\phi(t)$ is a nondecreasing function of $t > 0$ with $0 < \phi(t) \leq t$. Note that by the Cauchy-Schwarz inequality, (2) with $\phi(t) = t$ holds for any K -sparse signal \mathbf{x} . Furthermore, (2) with $\phi(t)$ much smaller than t holds for α -strongly-decaying signals and random signals (more details will be provided in Sec. 2).

In this paper, we develop a lower bound on the probability of the exact recovery for K -sparse signals \mathbf{x} that satisfy (2), using K -iterations of OMP, as a function of $\phi(t)$. Since the bound depends on the function $\phi(t)$, we develop closed-form expressions of $\phi(t)$ for general K -sparse signals, K -sparse α -strongly-decaying signals, and K -sparse signals whose nonzero entries independently and identically follow the $\mathcal{N}(0, 1)$ distribution, leading to exact lower bounds for these three classes of sparse signals.

*This work was partially supported by NSFC (No. 11871248) and “the Fundamental Research Funds for the Central Universities” (No. 21618329).

Algorithm 1 The OMP Algorithm [5]

Input: \mathbf{y} , \mathbf{A} , and stopping rule.Initialize: $k = 0$, $\mathbf{r}^0 = \mathbf{y}$, $\mathcal{S}_0 = \emptyset$.

until the stopping rule is met

1: $k = k + 1$,2: $s^k = \arg \max_{1 \leq i \leq n} |\langle \mathbf{r}^{k-1}, \mathbf{A}_i \rangle|$,3: $\mathcal{S}_k = \mathcal{S}_{k-1} \cup \{s^k\}$,4: $\hat{\mathbf{x}}_{\mathcal{S}_k} = \arg \min_{\mathbf{x} \in \mathbb{R}^{|\mathcal{S}_k|}} \|\mathbf{y} - \mathbf{A}_{\mathcal{S}_k} \mathbf{x}\|_2$,5: $\mathbf{r}^k = \mathbf{y} - \mathbf{A}_{\mathcal{S}_k} \hat{\mathbf{x}}_{\mathcal{S}_k}$.Output: $\hat{\mathbf{x}} = \arg \min_{\mathbf{x}: \text{supp}(\mathbf{x}) = \mathcal{S}_k} \|\mathbf{y} - \mathbf{A} \mathbf{x}\|_2$.

2. MAIN RESULTS

In the following, we provide a lower bound on the probability that OMP can exactly recover any K -sparse signal \mathbf{x} satisfying (2) in K iterations for random sensing matrix \mathbf{A} .

Theorem 1 *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries, and \mathbf{x} be a K -sparse signal that satisfies (2) for some particular $\phi(t)$. Define the event \mathbb{S} as*

$$\mathbb{S} = \{\text{OMP can exactly recover } \mathbf{x} \text{ in } (1) \text{ in } K \text{ iterations}\}. \quad (3)$$

Denote interval $\mathcal{I} = \left(0, 1 - \sqrt{\frac{K}{m}} - \sqrt{\frac{2\phi(K)}{m\pi}}\right]$, then

$$\mathbb{P}(\mathbb{S}) \geq \max_{\epsilon \in \mathcal{I}} (1 - e^{-\frac{\epsilon^2 m}{2}}) \prod_{k=1}^K \left(1 - \frac{e^{-\frac{\eta^2 m}{2\phi(k)}}}{\sqrt{\frac{\pi m}{2\phi(k)}} \eta}\right)^{(n-K)}, \quad (4)$$

where $\eta = 1 - \sqrt{K/m} - \epsilon$.

The proof of Theorem 1 can be found in the journal version of this paper. In the following, we give some remarks.

Theorem 1 is important from both theoretical and practical applications points of view. Theoretically, Theorem 1 characterizes the recovery performance of OMP. In practical applications, we can use (4) to give a lower bound on $\mathbb{P}(\mathbb{S})$. If the lower bound is large, saying close to 1, then we are confident to use the OMP algorithm to do the reconstruction. If the lower bound is small, saying much smaller than 1, then another more effective recovery algorithm (such as the basis pursuit [2]) may need to be used.

As far as we know, Theorem 1 gives the first lower bound on $\mathbb{P}(\mathbb{S})$ by using the extra information (i.e., inequality (2)) of the K -sparse signal \mathbf{x} . Note that [14, Theorem 6] also gives a lower bound on $\mathbb{P}(\mathbb{S})$, but it only uses the K -sparsity property of \mathbf{x} . Since Theorem 1 uses not only the sparsity of \mathbf{x} but also its additional property (2) to derive the lower bound, it provides a sharper lower bound on $\mathbb{P}(\mathbb{S})$ than [14, Theorem

6]. More details on the comparison of the two lower bounds are presented in Sec. 3.

Theorem 1 can theoretically explain that OMP has better recovery ability in recovering sparse signals with larger variation of the magnitudes of their nonzero entries. Specifically, it is not hard to see that the right-hand side of (4) becomes larger as $\phi(t)$ (or equivalently $\frac{\|\mathbf{x}_{\mathcal{S}}\|_1}{\|\mathbf{x}_{\mathcal{S}}\|_2}$ (see (2))) becomes smaller. By the Cauchy-Schwarz inequality, $\frac{\|\mathbf{x}_{\mathcal{S}}\|_1}{\|\mathbf{x}_{\mathcal{S}}\|_2}$ achieves the maximal value $\sqrt{|\mathcal{S}|}$ when the magnitudes of all the entries of $\mathbf{x}_{\mathcal{S}}$ are the same. Hence, the probability of exact recovery of K -sparse \mathbf{x} , whose non-zero entries have identical magnitudes, has the smallest lower bound. On the other hand, if the variation of the magnitudes of the nonzero entries of \mathbf{x} are large, then $\frac{\|\mathbf{x}_{\mathcal{S}}\|_1}{\|\mathbf{x}_{\mathcal{S}}\|_2}$ is small, and hence the right-hand side of (4) is large. Therefore, generally speaking, the probability of the exact recovery of this kind of K -sparse signals \mathbf{x} is large.

As (4) depends on $\phi(t)$, to lower bound $\mathbb{P}(\mathbb{S})$, we need to know $\phi(t)$. In the following, we give closed-form expressions of $\phi(t)$ for three cases. We begin with the first case where we only know that \mathbf{x} is K -sparse. By the Cauchy-Schwarz inequality, one can see that (2) holds if $\phi(t) = t$. Hence, by Theorem 1, we have

Corollary 1 *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries and \mathbf{x} be an arbitrary K -sparse signal. Then (4) holds with $\phi(t) = t$.*

Note that [14, Theorem 6] shows that

$$\begin{aligned} \mathbb{P}(\mathbb{S}) &\geq \max_{\epsilon \in (0, \sqrt{m/K-1})} (1 - e^{-\frac{\epsilon^2 m}{2}}) \left(1 - e^{-\frac{(\sqrt{m/K-1}-\epsilon)^2}{2}}\right)^{K(n-K)} \end{aligned} \quad (5)$$

where the event \mathbb{S} is defined in (3). Since the lower bounds on $\mathbb{P}(\mathbb{S})$ given by Corollary 1 and (5) are complicated, it is difficult to theoretically compare them. However, from the simulation results in Sec. 3, one can see that the new bound given by Corollary 1 is much sharper than that given by (5).

Next, we give a lower bound on $\mathbb{P}(\mathbb{S})$ for recovering α -strongly-decaying signal. First, we state the precise definition of α -strongly-decaying signals as follows:

Definition 1 ([10]) *Without loss of generality, let all the entries of K -sparse \mathbf{x} be ordered as*

$$|x_1| \geq |x_2| \geq \dots \geq |x_K| \geq 0, \quad x_j = 0, \text{ for } K+1 \leq j \leq n.$$

Then \mathbf{x} is called as a K -sparse α -strongly-decaying signal ($\alpha > 1$) if $|x_i| \geq \alpha |x_{i+1}|$, $1 \leq i \leq K-1$.

The following lemma provides a closed-form expression of $\phi(t)$ for K -sparse α -strongly-decaying signals.

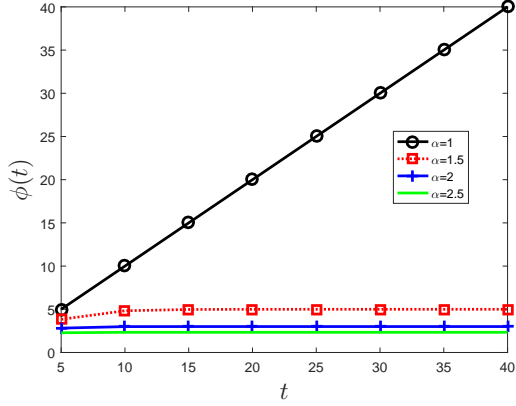


Fig. 1. $\phi(t)$ versus α with $\alpha = 1, 1.5, 2, 2.5$

Lemma 1 Let \mathbf{x} be a K -sparse α -strongly-decaying signal, then (2) holds with

$$\phi(t) = \frac{(\alpha^t - 1)(\alpha + 1)}{(\alpha^t + 1)(\alpha - 1)}, \quad t > 0. \quad (6)$$

The proof of Lemma 1 can be found in the journal version of this paper.

To show how large the $\phi(t)$ in (6) is, we plot it for different values of α in Fig.1, where for comparison, we also plot $\phi(t) = t$ (note that this is equivalent to the $\alpha = 1$ case as $\lim_{\alpha \rightarrow 1} \phi(t) = t$). Fig. 1 shows that $\phi(t)$ is much smaller than t for large t and/or α .

Theorem 1 and Lemma 1 implies the following corollary:

Corollary 2 Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries, and \mathbf{x} be a K -sparse α -strongly-decaying signal. Then (4) holds with $\phi(t)$ being defined in (6).

Since $\phi(t)$ in (6) is much smaller than t when t and/or α is large (see Fig. 1), the right-hand side of (4) with $\phi(t)$ being defined in (6) can be much larger than that with $\phi(t) = t$. This essentially implies that $\mathbb{P}(\mathbb{S})$ is larger for recovering α -strongly-decaying sparse signals than that for recovering flat sparse signals (i.e., the magnitudes of all the nonzero entries are identical). More details on this are given in Sec. 3.

Finally, we consider the recovery of random signals \mathbf{x} . Specifically, we assume that \mathbf{x} is K -sparse with $\mathbf{x}_\Omega \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ for certain σ . This kind of sparse signals arise from many applications, such as sparse activity users detection [1]. If $\mathbf{x}_\Omega \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, then $\mathbf{x}_\Omega / \sigma \sim \mathcal{N}(0, \mathbf{I})$. Since $\frac{\|\mathbf{x}_\Omega\|_1}{\|\mathbf{x}_\Omega\|_2} = \frac{\|\mathbf{x}_\Omega / \sigma\|_1}{\|\mathbf{x}_\Omega / \sigma\|_2}$, to find a function $\phi(t)$ such that (2) holds for K -sparse signal \mathbf{x} satisfying $\mathbf{x}_\Omega \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, we only need to find a $\phi(t)$ such that (2) holds for K -sparse \mathbf{x} satisfying $\mathbf{x}_\Omega \sim \mathcal{N}(0, \mathbf{I})$. Since \mathbf{x} is a random signal, it is impossible to find a $\phi(t)$ such that (2) always holds. But we can find a $\phi(t)$ such that (2) holds with high probability.

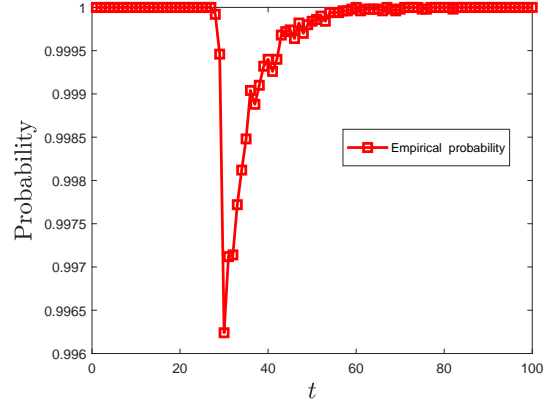


Fig. 2. The empirical probability of $\|\mathbf{x}\|_1^2 / \|\mathbf{x}\|_2^2 \leq \phi(t)$ over 50000 realizations of $\mathbf{x} \in \mathbb{R}^t \sim \mathcal{N}(0, \mathbf{I})$

If $\mathbf{x}_\Omega \sim \mathcal{N}(0, \mathbf{I})$, then the expected value of $\|\mathbf{x}\|_1^2$ divided by the expected value of $\|\mathbf{x}\|_2^2$ equals to $\sqrt{2/\pi}|\Omega|$. Therefore, we may try $\phi(t) = \sqrt{2/\pi}t$. However, from simulations, we found that (2) does not hold with high probability when $|\Omega|$ is small. Fortunately, $\phi(t)$ defined in (7) below is a suitable function to ensure (2) holds with high probability:

$$\phi(t) = \begin{cases} 0.8t & t \geq 30 \\ 24 & 25 \leq t \leq 29 \\ t & t \leq 24 \end{cases}. \quad (7)$$

Fig. 2 shows the probability of $\|\mathbf{x}\|_1^2 / \|\mathbf{x}\|_2^2 \leq \phi(t)$ for $t = 1, 2, \dots, 50$ over 50000 realizations, where $\phi(t)$ is defined in (7). From Fig. 2, one can see that (2) holds with $\phi(t)$ in (7) with probability larger than 0.996. Hence, we have the following observation:

Observation 1 Suppose that $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a random matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries, and \mathbf{x} is a K -sparse signal satisfying $\mathbf{x}_\Omega \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ for certain σ . Then (4) holds with $\phi(t)$ being defined in (7) with empirical probability larger than 0.996.

Fig. 2 indicates that (2) holds with $\phi(t)$ being defined in (7) with probability larger than 0.996. Since $\phi(t)$ being defined in (7) is much smaller than $\phi(t) = t$ for large t , Observation 1 essentially implies that $\mathbb{P}(\mathbb{S})$ is larger for recovering Gaussian sparse signals than that for recovering flat sparse signals. More details on this will be provided in Sec. 3.

3. SIMULATION TESTS

This section performs simulations to illustrate Theorem 1, Corollaries 1–2 and Observation 1 and compare them with [14, Theorem 6].

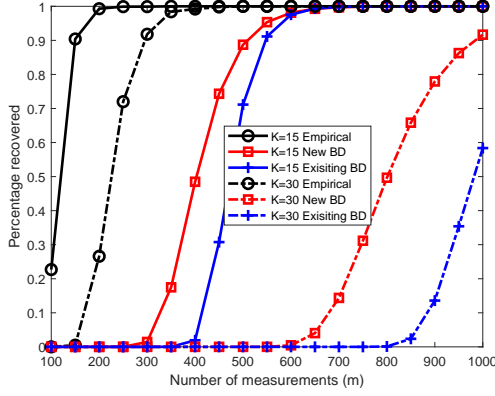


Fig. 3. Recovery of K -sparse flat signals

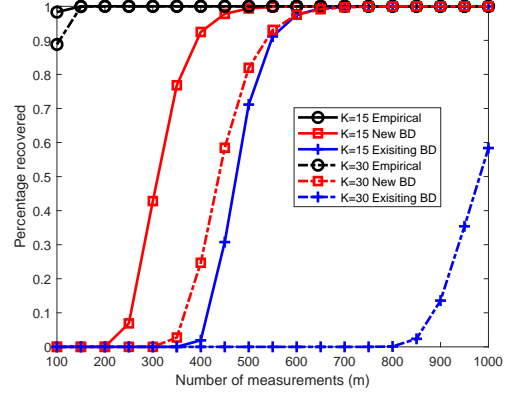


Fig. 5. Recovery of K -sparse 1.2-strongly-decaying signals

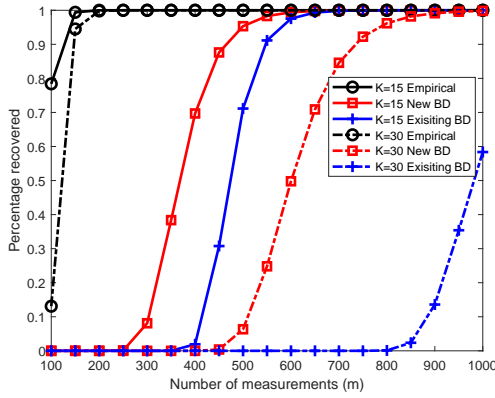


Fig. 4. Recovery of K -sparse 1.1-strongly-decaying signals

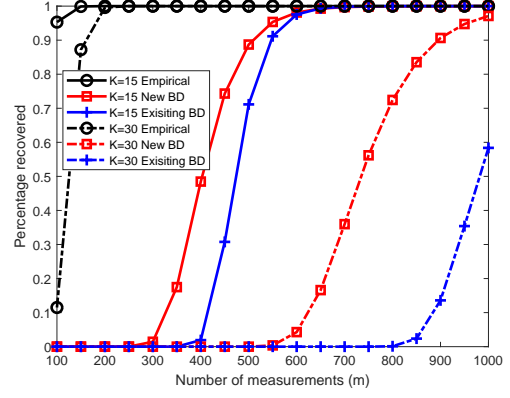


Fig. 6. Recovery of K -sparse Gaussian signals

We generated 1000 realizations of linear model (1). More specifically, for each fixed m , n and K , for each realization, we generated an $\mathbf{A} \in \mathbb{R}^{m \times n}$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries; we randomly selected K elements from the set $\{1, 2, \dots, n\}$ to form the support Ω of \mathbf{x} ; and then generated an $\mathbf{x} \in \mathbb{R}^n$ according to the following four cases: 1) $x_i = 1$ for $i \in \Omega$ and $x_i = 0$ for $i \notin \Omega$; 2) The i -th element of \mathbf{x}_Ω is 1.1^{K-i} for $i \in \Omega$ and $x_i = 0$ for $i \notin \Omega$; 3) The i -th element of \mathbf{x}_Ω is 1.2^{K-i} for $i \in \Omega$ and $x_i = 0$ for $i \notin \Omega$; 4) $\mathbf{x}_\Omega = \text{randn}(K, 1)$ and $\mathbf{x}_{\Omega^c} = \mathbf{0}$, where randn is a MATLAB built-in function. After generating \mathbf{A} and \mathbf{x} , we set $\mathbf{y} = \mathbf{A}\mathbf{x}$. Then, we use OMP to reconstruct \mathbf{x} , and denote the number of exact recovery of \mathbf{x} (note that \mathbf{x} is thought as exactly recovered if the 2-norm of the difference between the returned \mathbf{x} and generated \mathbf{x} is not larger than 10^{-10}) over 1000 as “Empirical”.

We respectively compute the right-hand side of (4) with $\phi(t) = t$, $\phi(t)$ defined by (6) with $\alpha = 1.1$ and $\alpha = 1.2$ and $\phi(t)$ defined by (7) for the four cases, and denote them as “New BD”. To compare Corollaries 1–2 and Observation 1 with [14, Theorem 6], we also compute the right-hand side of (5) and denote it as “Existing BD”. Since the lower bound

on $\mathbb{P}(\mathcal{S})$ given by [14, Theorem 6] uses the sparsity of \mathbf{x} only, “Existing BD” are the same for all the four cases.

Figs. 3-6 respectively display “Empirical”, “New BD” and “Existing BD” for $m = 100 : 50 : 1000$ and $n = 1024$ with $K = 15$ and $K = 30$ for \mathbf{x} from cases 1-4. Figs. 3-6 show that “New BD” are much larger than “Existing BD” for all the four cases which indicates that the lower bounds on $\mathbb{P}(\mathcal{S})$ given by Corollaries 1–2 and Observation 1 are much sharper than that given by [14, Theorem 6]. They also show that OMP has significantly better recovery performance in recovering α -strongly-decaying and Gaussian sparse signals than recovering flat sparse signals, and the recovery performance of the OMP algorithm for recovering α -strongly-decaying sparse signals becomes better as α gets larger.

4. CONCLUSIONS

In this paper, we developed lower bounds on the probability of exact recovery using K iterations of OMP for \mathbf{x} satisfying a condition that characterizes the variations in the magnitudes of the nonzero entries of \mathbf{x} .

5. REFERENCES

- [1] Z. Chen, F. Sahrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, April 2018.
- [2] E. J. Candés and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [5] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, Nov. 1993, pp. 40–44.
- [6] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [7] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sept 2011.
- [8] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [9] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, 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," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sept. 2010.
- [11] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, March 2012.
- [12] J. Wen, X. Zhu, and D. Li, "Improved bounds on the restricted isometry constant for orthogonal matching pursuit," *Electronics Letters*, vol. 49, pp. 1487–1489, 2013.
- [13] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1370–1382, March 2017.
- [14] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit: The Gaussian case," *Caltech, Pasadena, CA, 2007, ACM Tech. Rep. 2007-01*, 2007.
- [15] E. A. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, vol. 16, no. 9, pp. 770–773, Sept. 2009.
- [16] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, "From blind to guided audio source separation: How models and side information can improve the separation of sound," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 107–115, May 2014.