

On Probability of Support Recovery for Orthogonal Matching Pursuit Using Mutual Coherence

Ehsan Miandji[†], *Student Member, IEEE*, Mohammad Emadi[†], *Member, IEEE*,
Jonas Unger, *Member, IEEE*, and Ehsan Afshari, *Senior Member, IEEE*

Abstract—In this paper we present a new coherence-based performance guarantee for the Orthogonal Matching Pursuit (OMP) algorithm. A lower bound for the probability of correctly identifying the support of a sparse signal with additive white Gaussian noise is derived. Compared to previous work, the new bound takes into account the signal parameters such as dynamic range, noise variance, and sparsity. Numerical simulations show significant improvements over previous work and a closer match to empirically obtained results of the OMP algorithm.

Index Terms—Compressed Sensing (CS), Sparse Recovery, Orthogonal Matching Pursuit (OMP), Mutual Coherence

I. INTRODUCTION

Let $\mathbf{s} \in \mathbb{R}^N$ be an unknown variable that we would like to estimate from the measurements

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a deterministic matrix and $\mathbf{w} \in \mathbb{R}^M$ is a noise vector, often assumed to be white Gaussian noise with mean zero and covariance $\sigma^2 \mathbf{I}$, where \mathbf{I} is the identity matrix. The matrix \mathbf{A} is called a *dictionary*. We consider the case when \mathbf{A} is *overcomplete*, i.e. $N > M$, hence uniqueness of the solution of (1) cannot be guaranteed. However, if most elements of \mathbf{s} are zero, we can limit the space of possible solutions, or even obtain a unique one, by solving

$$\hat{\mathbf{s}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \epsilon, \quad (2)$$

where ϵ is a constant related to \mathbf{w} . The location of nonzero entries in \mathbf{s} is known as the *support* set, which we denote by Λ . In some applications, e.g. estimating the direction of arrival in antenna arrays [1], correctly identifying the support is more important than accuracy of values in $\hat{\mathbf{s}}$. When the correct support is known, the solution of the least squares problem $\|\mathbf{y} - \mathbf{A}_\Lambda \mathbf{x}_\Lambda\|_2^2$ gives $\hat{\mathbf{s}}$, where \mathbf{A}_Λ is formed using the columns of \mathbf{A} indexed by Λ , see [2], [3].

Solving (2) is an NP-hard problem and several greedy algorithms have been proposed to compute an approximate solution of (2); a few examples include Matching Pursuit (MP) [4], Orthogonal Matching Pursuit (OMP) [5], Regularized-OMP

(ROMP) [6], and Compressive Sampling Matching Pursuit (CoSaMP) [7]. In contrast to greedy methods, convex relaxation algorithms [8]–[11] replace the ℓ_0 pseudo-norm in (2) with an ℓ_1 norm, leading to a convex optimization problem known as the Basis Pursuit (BP) problem [12]. While convex relaxation methods require weaker conditions for exact recovery [2], [13], they are computationally more expensive than greedy methods, specially when $N \gg M$ [7], [14], [15].

The most important aspect of a sparse recovery algorithm is the uniqueness of the obtained solution. Mutual Coherence (MC) [16], cumulative coherence [13], the spark [17], Exact Recovery Coefficient (ERC) [18], and Restricted Isometry Constant (RIC) [19] are metrics proposed to evaluate the suitability of a dictionary for exact recovery. Among these metrics, RIC, spark, and ERC achieve better performance guarantees; however, computing RIC and the spark is in general NP-hard and calculating ERC is a combinatorial problem. In contrast, MC can be efficiently computed and has shown to provide acceptable performance guarantees [2], [3], [20]–[22].

In this paper, we derive a new lower bound for the probability of correctly identifying the support of a sparse signal using the OMP algorithm. Our main motivation is that previous methods do not directly take into account signal parameters such as dynamic range, sparsity, and the noise characteristics in the computed probability. We will elaborate on this in section II, where we discuss the most recent theoretical analysis for OMP based on MC. The main result of the paper will be presented in section III, followed by numerical evaluation of the new performance guarantee in section IV.

II. MOTIVATION

The mutual coherence of a dictionary \mathbf{A} , denoted $\mu_{\max}(\mathbf{A})$, is the maximum absolute cross correlation of its columns [16]:

$$\mu_{i,j}(\mathbf{A}) = \langle \mathbf{A}_i, \mathbf{A}_j \rangle, \quad (3)$$

$$\mu_{\max}(\mathbf{A}) = \max_{1 \leq i \neq j \leq N} |\mu_{i,j}(\mathbf{A})|, \quad (4)$$

where we have assumed, as with the rest of the paper, that $\|\mathbf{A}_i\|_2 = 1$, $i \in \{1, \dots, N\}$. Apart from MC and sparsity,

$$s_{\min} = \min(|s_i|), \quad \text{and} \quad s_{\max} = \max(|s_i|), \quad \forall i \in \Lambda, \quad (5)$$

which define the dynamic range of the signal, also affect the performance of OMP. The following theorem establishes an important coherence-based performance guarantee for OMP.

Theorem 1 (Ben-Haim et al. [3]). *Let $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\|\mathbf{s}\|_0 = \tau$ and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. If*

$$s_{\min} - (2\tau - 1)\mu_{\max}s_{\min} \geq 2\beta, \quad (6)$$

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

E. Miandji and J. Unger are with the Department of Science and Technology, Linköping University, Sweden (e-mail: {ehsan.miandji, jonas.unger}@liu.se). M. Emadi is with Qualcomm Technologies Inc., San Jose, CA USA (e-mail: memadi@qti.qualcomm.com). And E. Afshari is with the Department of Electrical Engineering and Computer Science, University of Michigan, MI USA (e-mail: afshari@umich.edu).

[†] Equal contributor

where $\beta \triangleq \sigma\sqrt{2(1+\alpha)\log N}$ is defined for some constant $\alpha > 0$, then with probability at least

$$1 - \frac{1}{N^\alpha \sqrt{\pi(1+\alpha)\log N}}, \quad (7)$$

OMP identifies the true support, denoted Λ .

The proof involves analyzing the probability event $\Pr\{|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta\}$, for some constant $\beta > 0$ and for all $j = 1, \dots, N$ (see [3] for details). They show that with the lower bound probability of (7), the inequality $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$ holds. It is then shown that if $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$ and (6) hold, then OMP identifies the correct support in each iteration. Moreover, it is assumed that the elements of the sparse vector \mathbf{s} are deterministic variables. Hence a strong condition such as (6) is required to determine if the support of \mathbf{s} can be recovered.

Our analysis removes the condition stated in (6) and introduces a probabilistic bound that depends on N , τ , μ_{\max} , s_{\max} , s_{\min} , and the signal noise. Hence we derive a probability bound that directly takes into account signal parameters and MC. Moreover, unlike [3], we assume that the nonzero elements of \mathbf{s} are centered independent random variables with arbitrary distributions. This enables the derivation of a more accurate bound for the probability of exact support recovery.

III. OMP CONVERGENCE ANALYSIS

In this section we present and prove the main result of the paper. Numerical results will be presented in section IV.

Theorem 2. *Let $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{w}$, where $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\tau = \|\mathbf{s}\|_0$ and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Moreover, assume that the nonzero elements of \mathbf{s} are independent centered random variables with arbitrary distributions. Let $\lambda = \Pr\{|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta\}$, for some constant $\beta \geq 0$ and $\forall j \in \{1, \dots, N\}$. If $s_{\min}/2 \geq \beta$, then OMP identifies the true support with lower bound probability*

$$\lambda \left(1 - 2N \exp\left(\frac{-N(s_{\min}/2 - \beta)^2}{2\tau^2\gamma^2 + 2N\gamma(s_{\min}/2 - \beta)/3}\right) \right), \quad (8)$$

where $\gamma = \mu_{\max}s_{\max}$. Moreover, λ is lower bounded by

$$1 - N \sqrt{\frac{2}{\pi}} \frac{\sigma}{\beta} e^{-\beta^2/2\sigma^2}. \quad (9)$$

Before presenting the proof, let us compare Theorems 1 and 2 analytically. It is important to note that (9) is indeed equivalent to (7). The apparent difference is only attributed to the use of α or β from the definition $\beta \triangleq \sigma\sqrt{2(1+\alpha)\log N}$. For instance, using the aforementioned definition of β on (9) leads to (7). As a result, the second term of (8) can be interpreted as a probabilistic representation of the condition imposed by (6) in Theorem 1. Moreover, because (9) is equal to (7) and the second term of (8) is in the range $[0, 1]$, therefore (8) is always smaller or equal to (7). However, as it will be seen in section IV, since the condition of Theorem 1 in (6) is not satisfied in many scenarios, our results match the empirical results more closely. Evidently, the condition $s_{\min}/2 \geq \beta$ in Theorem 2 is more relaxed compared to (6). Our numerical results in Section IV also verify this fact.

The following lemma will provide us with the necessary tool for the proof of Theorem 2. The proof of the lemma is postponed to the Appendix.

Lemma 1. *Define $\Gamma_j = |\langle \mathbf{A}_j, \mathbf{A}\mathbf{s} + \mathbf{w} \rangle|$, for any $j \in \{1, \dots, N\}$, where $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ and $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$. Then for some constant $\xi \geq 0$, and assuming $\xi \geq \beta$, we have*

$$\Pr\{\Gamma_j \geq \xi\} \leq 2 \exp\left(\frac{-(\xi - \beta)^2}{2(N\nu + c(\xi - \beta)/3)}\right), \quad (10)$$

where

$$|\mu_{j,n}\mathbf{s}_n| \leq c, \quad E\{\mu_{j,n}^2\mathbf{s}_n^2\} \leq \nu, \quad \forall n \in \{1, \dots, N\} \quad (11)$$

We can now state the proof of Theorem 2.

Proof of Theorem 2. It was shown in [3] that OMP identifies the true support Λ if

$$\min_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_\Lambda \mathbf{s}_\Lambda + \mathbf{w} \rangle| \geq \max_{k \notin \Lambda} |\langle \mathbf{A}_k, \mathbf{A}_\Lambda \mathbf{s}_\Lambda + \mathbf{w} \rangle|. \quad (12)$$

The term on the left-hand side of (12) can be rewritten as

$$\begin{aligned} & \min_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_\Lambda \mathbf{s}_\Lambda + \mathbf{w} \rangle| \\ &= \min_{j \in \Lambda} |\mathbf{s}_j + \langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \end{aligned} \quad (13)$$

$$\geq \min_{j \in \Lambda} |\mathbf{s}_j| - \max_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle|. \quad (14)$$

From (12) and (14), we can see that the OMP algorithm identifies the true support if

$$\begin{cases} \max_{k \notin \Lambda} \{\Gamma_k\} < \min_{j \in \Lambda} \frac{|\mathbf{s}_j|}{2}, \\ \max_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| < \min_{j \in \Lambda} \frac{|\mathbf{s}_j|}{2}. \end{cases} \quad (15)$$

Using (15), we can define the probability of error as

$$\begin{aligned} \Pr\{\text{error}\} &\leq \Pr\left\{\max_{j \in \Lambda} |\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \geq \frac{s_{\min}}{2}\right\} \\ &\quad + \Pr\left\{\max_{k \notin \Lambda} \{\Gamma_k\} \geq \frac{s_{\min}}{2}\right\} \end{aligned} \quad (16)$$

$$\begin{aligned} &\leq \sum_{j \in \Lambda} \Pr\left\{|\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \geq \frac{s_{\min}}{2}\right\} \\ &\quad + \sum_{k \notin \Lambda} \Pr\left\{\Gamma_k \geq \frac{s_{\min}}{2}\right\}. \end{aligned} \quad (17)$$

For the first term on the right-hand side of (17), excluding the summation over the indices in Λ , from Lemma 1 we have

$$\begin{aligned} &\Pr_{j \in \Lambda} \left\{|\langle \mathbf{A}_j, \mathbf{A}_{\Lambda \setminus \{j\}} \mathbf{s}_{\Lambda \setminus \{j\}} + \mathbf{w} \rangle| \geq \frac{s_{\min}}{2}\right\} \\ &\leq 2 \exp\left(\underbrace{\frac{-\rho^2}{2((\tau-1)\nu + c\rho/3)}}_{P_1}\right), \end{aligned} \quad (18)$$

where $\rho = s_{\min}/2 - \beta$ is defined for notational brevity. Note that the dictionary \mathbf{A} in (18) is supported on $\Lambda \setminus \{j\}$, i.e. all the indices in the true support excluding j . Therefore the term $(\tau - 1)$, instead of N , appears in the denominator of (18). Similarly, for the second term of (17) we have

$$\Pr_{k \notin \Lambda} \left\{\Gamma_k \geq \frac{s_{\min}}{2}\right\} \leq 2 \exp\left(\underbrace{\frac{-\rho^2}{2(\tau\nu + c\rho/3)}}_{P_2}\right). \quad (19)$$

Substituting (18) and (19) into (17) yields

$$\Pr\{\text{error}\} \leq \tau P_1 + (N - \tau)P_2 \leq NP_2, \quad (20)$$

where the last inequality follows since $P_2 > P_1$.

Moreover, for the upper bounds c and ν in (11) we have

$$|\mu_{j,n} s_n| \leq \mu_{\max} s_{\max}, \quad (21)$$

$$\mathbb{E}\{\mu_{j,n}^2 s_n^2\} \leq \frac{1}{N} \sum_{n=1}^N \mu_{\max}^2 \mathbb{E}\{s_n^2\} \leq \frac{\tau}{N} s_{\max}^2 \mu_{\max}^2, \quad (22)$$

Combining (21) and (22) with (20), the following is obtained

$$\Pr\{\text{error}\} \leq 2N \exp\left(\frac{-N\rho^2}{2\tau^2\gamma^2 + 2N\gamma\rho/3}\right), \quad (23)$$

where we have defined $\gamma = \mu_{\max} s_{\max}$ for notational brevity.

So far we have assumed that $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$, $\forall j$. Therefore, the probability of success is the joint probability of $\Pr\{|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta\}$ and the inverse of (23). For the former, a lower bound was formulated in [3] as follows

$$\Pr\{|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta\} \geq 1 - \underbrace{\sqrt{\frac{2}{\pi}} \frac{\sigma}{\beta} e^{-\beta^2/2\sigma^2}}_{P_3}. \quad (24)$$

Since $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$ should hold $\forall j \in \{1, \dots, N\}$, we have

$$\Pr_{j=1, \dots, N}\{|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta\} \geq (1 - P_3)^N \geq 1 - NP_3. \quad (25)$$

Inverting the probability event in (23) and multiplying by the lower bound in (25) yields (8), which completes our proof. \square

IV. NUMERICAL RESULTS

In this section we compare numerical results of Theorem 1 (Ben-Haim et al. [3]), and Theorem 2 (proposed herein) with the empirical results of OMP. Indeed we only consider probability of successful recovery of the support. An upper bound for the MSE of the oracle estimator has been previously established, see e.g. Theorem 5.1 in [2] or Lemma 4 in [3]. The oracle estimator knows the support of the signal, *a priori*.

All the empirical results are obtained by performing the OMP algorithm 5000 times using a random sparse signal with additive white Gaussian noise in each trial. The probability of success is computed as the ratio of successful trials to the total number of trials; note that a trial is successful if $\Lambda = \hat{\Lambda}$, where $\hat{\Lambda}$ is the support of $\hat{\mathbf{s}}$ obtained from OMP by solving (2). Moreover, the number of trials was empirically set such that the probability of success for the OMP algorithm was stable across different parameters. For comparison, we use the dictionary of [3] defined as $\mathbf{A} = [\mathbf{I}, \mathbf{H}]$, where \mathbf{I} is an identity matrix and \mathbf{H} is a Hadamard matrix, hence we have $N = 2M$.

The sparse signal in each trial, denoted \mathbf{s} in (1), is constructed as follows: The support of the sparse signal, $\Lambda = \text{supp}(\mathbf{s})$, is constructed by uniform random permutation of the set $\{1, \dots, N\}$ and taking the first τ indices. The nonzero elements located at Λ are drawn randomly from a uniform distribution on the interval $[s_{\min}, s_{\max}]$, multiplied randomly by $+1$ or -1 . Once the sparse signal is constructed, the input of the OMP algorithm, \mathbf{y} , is obtained by evaluating (1).

In order to facilitate the comparison of Theorems 1 and 2, we need to fix the value of β . To do this, we empirically

calculate β as $\max_{\mathbf{w}} \max_j |\langle \mathbf{A}_j, \mathbf{w} \rangle|$, where the maximum over \mathbf{w} is computed using 10^4 vectors $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, as assumed by both theorems. Given β , we can calculate α for Theorem 1 from the definition $\beta \triangleq \sigma \sqrt{2(1 + \alpha) \log N}$. Indeed, a lower value of β leads to better results for both theorems, see (8) and (6). As a result, here we consider the worst-case scenario. When (6) is not satisfied for Theorem 1, we set the probability of success to zero. We use the same procedure for the condition of Theorem 2; i.e. the probability of success is set to zero when $s_{\min}/2 < \beta$.

Numerical results are summarized in Figure 1. We analyze the effect of sparsity on the probability of successful support recovery in plots 1a, 1b, and 1c. Three signal dimensionalities and three noise variances: $\sigma_1^2 = 10^{-6}$, $\sigma_2^2 = 2.5 \times 10^{-5}$, and $\sigma_3^2 = 10^{-4}$, are considered. For all these cases we set $s_{\min} = 0.5$ and $s_{\max} = 1$. In Fig. 1a we see that Theorem 1 achieves a higher probability for σ_3 and small values of τ , while Theorem 2 leads to more accurate results for larger values of τ . Additionally, for σ_1 and σ_2 , Theorem 2 is much closer to empirical results. Most importantly, the shape of the probability curves for Theorem 2 matches the empirical curves. In contrast, Theorem 1 produces a step function due to the fact that condition (6) is not satisfied for a large range of values for τ , even though the success probability in (7) is close to one for different values of σ . The condition of Theorem 2 is satisfied across all the parameters for figures 1a-1c.

We discussed in section III that (8) is always smaller than (7) due to the second term of (8). We expect this term to become more accurate as the signal dimensionality grows since it is exponential in N ; moreover, β and μ_{\max} become smaller as N grows. This is confirmed in figures 1b and 1c. As we increase N , the gap between theorems 1 and 2 increases, confirming that the second term of (8) is becoming more accurate compared to (6). The empirical probability is close to one for all the values of τ plotted in figures 1b and 1c.

The effect of s_{\min} on the probability of success is demonstrated in figures 1d, 1e, and 1f. For each plot, we consider $\tau_1 = 16$, $\tau_2 = 32$, and $\tau_3 = 64$, while setting $\sigma^2 = 10^{-4}$. The empirical results show a probability of success close to one across the parameters considered. In Fig. 1d we see a significant difference between Theorems 1 and 2. The condition of Theorem 1 is not satisfied for any value of s_{\min} and τ . In contrast, Theorem 2 shows high probabilities for all three values of τ . The dynamic range (DR) of the signal can be defined as s_{\max}^2/s_{\min}^2 . As we increase the signal dimensionality (N), Theorem 2 reports larger probability for larger values of DR and all three values of τ . On the other hand, the condition of Theorem 1 fails for τ_2 and τ_3 , even when we have $M = 4096$. For τ_1 , Theorem 1 can produce valid results for a slightly higher DR.

Lastly, in plots 1g, 1h, and 1i, we analyze the effect of noise variance on the probability of success for $\tau_1 = 16$, $\tau_2 = 32$, and $\tau_3 = 64$. In Fig. 1g, where $M = 1024$, both theorems fail to produce valid results for $\tau_3 = 64$. However, Theorem 2 reports acceptable results for τ_1 and τ_2 , while the condition of Theorem 1 is not satisfied. As the signal dimensionality grows, see Fig. 1h and 1i, Theorem 2 becomes more tolerant of higher

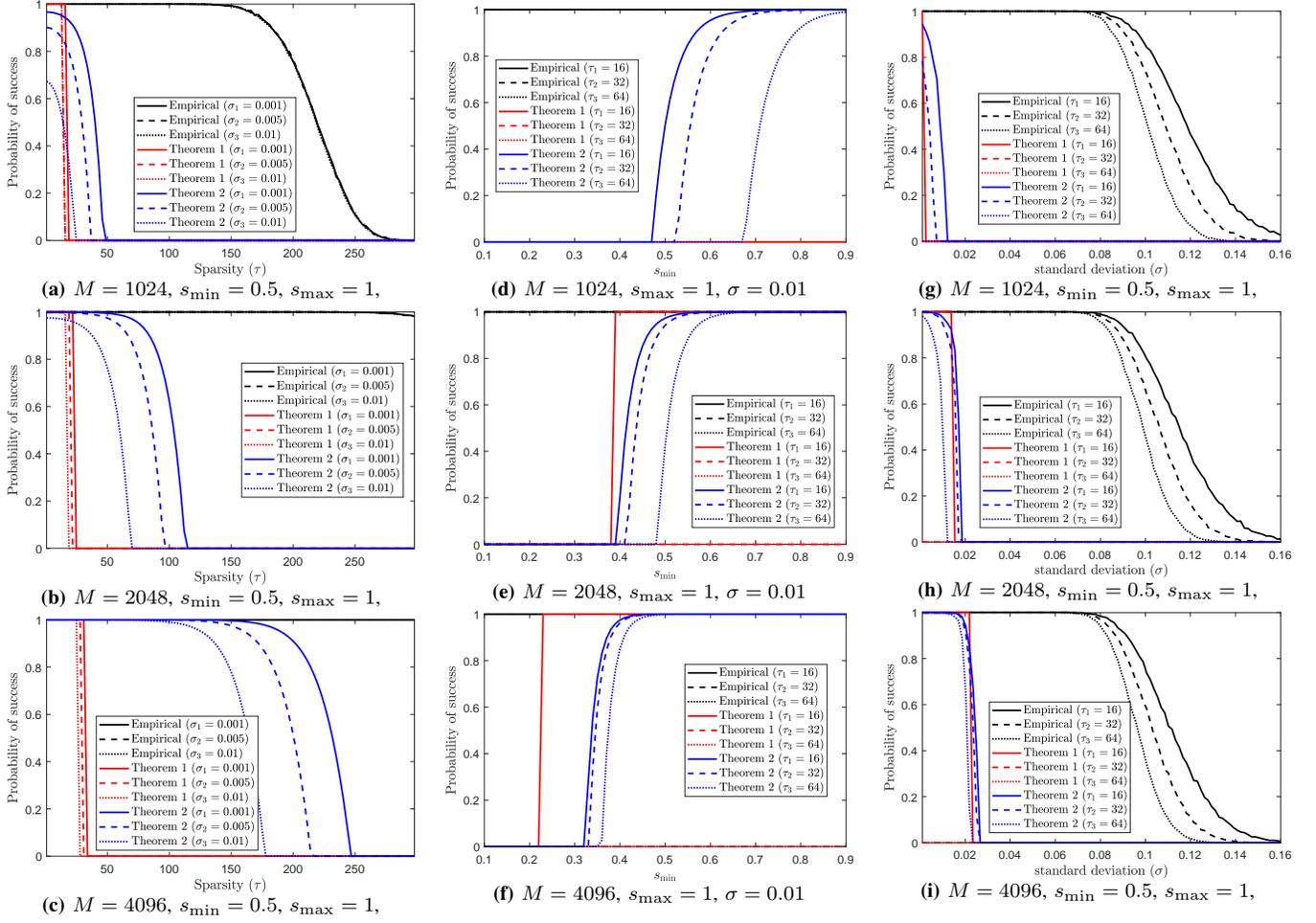


Fig. 1: In each column of plots we demonstrate the effect of one parameter on the probability of successful support recovery while fixing the other parameters. Rows represent different values of M . The mutual coherence of the dictionary varies based on M . For $M = 1024$, $M = 2048$, and $M = 4096$, we have $\mu_{\max} = 0.0313$, $\mu_{\max} = 0.0221$, and $\mu_{\max} = 0.0156$, respectively.

noise variances. The results for Theorem 1 also improves with increasing signal dimensionality, however only for τ_1 . This shows the robustness of Theorem 2 to larger values of sparsity.

V. CONCLUSIONS

We presented a new bound for the probability of correctly identifying the support of a noisy sparse signal using the OMP algorithm. Compared to the analysis of Ben-Haim et al. [3], our analysis replaces a sharp condition with a probabilistic bound. Comparisons to empirical results obtained by OMP show a much improved correlation than previous work.

APPENDIX

proof of Lemma 1. Expanding Γ_j , we can show that

$$\Gamma_j = \left| \sum_{m=1}^M \mathbf{A}_{m,j} \left(\sum_{n=1}^N \mathbf{A}_{m,n} \mathbf{s}_n + \mathbf{w}_m \right) \right| \quad (26)$$

$$= \left| \sum_{n=1}^N \left\{ \sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{A}_{m,n} \mathbf{s}_n + \frac{1}{N} \sum_{m=1}^M \mathbf{A}_{m,j} \mathbf{w}_m \right\} \right|. \quad (27)$$

$$= \left| \sum_{n=1}^N \left\{ \mu_{j,n} \mathbf{s}_n + \frac{1}{N} \langle \mathbf{A}_j, \mathbf{w} \rangle \right\} \right|. \quad (28)$$

We are interested in tail bounds for sum of random variables $\mu_{j,n} \mathbf{s}_n + N^{-1} \langle \mathbf{A}_j, \mathbf{w} \rangle$, for $n = 1, \dots, N$. Let us define $\mathbf{x}_n = \mu_{j,n} \mathbf{s}_n$. Using the assumption $|\langle \mathbf{A}_j, \mathbf{w} \rangle| \leq \beta$ we have

$$\begin{aligned} \Pr \{ \Gamma_j \geq \xi \} &\leq \Pr \left\{ \left| \sum_{n=1}^N \mathbf{x}_n \right| + \left| \frac{1}{N} \sum_{n=1}^N \langle \mathbf{A}_j, \mathbf{w} \rangle \right| \geq \xi \right\} \\ &\leq \Pr \left\{ \left| \sum_{n=1}^N \mathbf{x}_n \right| \geq \xi - \beta \right\}. \end{aligned} \quad (29)$$

Since $\{\mathbf{s}_n\}_{n=1}^N$, and hence $\{\mathbf{x}_n\}_{n=1}^N$, are centered independent real random variables, according to Bernstein's inequality [23], if $\mathbb{E} \{\mathbf{x}_n^2\} \leq \nu$, and $\Pr \{ |\mathbf{x}_n| < c \} = 1$, then for a positive constant δ we have

$$\begin{aligned} \Pr \left\{ \left| \sum_{n=1}^N \mathbf{x}_n \right| \geq \delta \right\} &\leq 2 \exp \left(\frac{-\delta^2}{2 \left(\sum_{n=1}^N \mathbb{E} \{ \mathbf{x}_n^2 \} + c\delta/3 \right)} \right) \\ &\leq 2 \exp \left(\frac{-\delta^2}{2(N\nu + c\delta/3)} \right). \end{aligned} \quad (30)$$

Setting $\delta = \xi - \beta$ in (30) completes the proof. \square

REFERENCES

- [1] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 3010–3022, Aug 2005.
- [2] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6–18, Jan 2006.
- [3] Z. Ben-Haim, Y. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5030–5043, Oct 2010.
- [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. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, Nov 1993, pp. 40–44 vol.1.
- [6] 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.
- [7] 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.
- [8] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2479–2493, July 2009.
- [9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [10] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec 2007.
- [11] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.
- [12] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [13] J. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1030–1051, March 2006.
- [14] S. H. Hsieh, C. S. Lu, and S. C. Pei, "Fast omp: Reformulating omp via iteratively refining ℓ_2 -norm solutions," in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, Aug 2012, pp. 189–192.
- [15] F. Marvasti, A. Amini, F. Haddadi, M. Soltanolkotabi, B. H. Khalaj, A. Aldroubi, S. Sanei, and J. Chambers, "A unified approach to sparse signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, p. 44, 2012.
- [16] 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.
- [17] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [18] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, Oct 2004.
- [19] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [20] P. L. Dragotti and Y. M. Lu, "On sparse representation in fourier and local bases," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7888–7899, Dec 2014.
- [21] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, "Exact recovery conditions for sparse representations with partial support information," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7509–7524, Nov 2013.
- [22] T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3516–3522, July 2010.
- [23] G. Bennett, "Probability inequalities for the sum of independent random variables," *Journal of the American Statistical Association*, vol. 57, no. 297, pp. 33–45, 1962.