

# Performance Tradeoffs for Exact Support Recovery of Sparse Signals

Yuzhe Jin, Young-Han Kim, and Bhaskar D. Rao  
 Department of Electrical and Computer Engineering  
 University of California, San Diego  
 9500 Gilman Drive, La Jolla, CA 92093-0407, USA  
 Email: {yujin,yhk,brao}@ucsd.edu

**Abstract**—We study the tradeoffs between the number of measurements, the signal sparsity level, and the measurement noise level for exact support recovery of sparse signals via random noisy measurements. By drawing analogy between exact support recovery and communication over the Gaussian multiple access channel, and exploiting mathematical tools developed for the latter problem, we derive sharp asymptotic sufficient and necessary conditions for exact support recovery. Specifically, when the number of nonzero entries is held fixed, the exact asymptotics on the number of measurements for support recovery is developed. When the number of nonzero entries increases in certain manners, we obtain sufficient conditions tighter than existing results. The proposed information theoretic framework for analyzing the performance of support recovery is further demonstrated to be capable of dealing with a variety of sparse signal recovery models.

## I. INTRODUCTION

Consider the estimation of a sparse signal  $\mathbf{X} \in \mathbb{R}^m$  via linear measurements  $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{Z}$ , where  $\mathbf{A} \in \mathbb{R}^{n \times m}$  is referred to as the measurement matrix and  $\mathbf{Z} \in \mathbb{R}^n$  is the measurement noise. The goal is to reconstruct the signal  $\mathbf{X}$  from as few number of measurements as possible. This problem has recently received much attention and has many applications such as compressed sensing, biomagnetic inverse problems, image processing, outlier detection, bandlimited extrapolation and spectral estimation, channel estimation, and wireless communication [1]–[4]. Efficient algorithms for sparse signal recovery have been proposed to find or approximate the sparse solution  $\mathbf{X}$  in various settings, e.g., matching pursuit algorithms [5], [6], convex [7], [8] and nonconvex [9] optimization algorithms, Bayesian methods [10], etc.

In many applications, it is important to find the exact support of the sparse signal [11], [12]. Performance analyses have been developed to shed light on the abilities of practical algorithms for support recovery in both noiseless settings [13]–[15] and noisy settings [16], [17]. Meanwhile, information theoretic tools have proven useful in understanding the fundamental performance tradeoff, regardless of algorithm complexity, for exact support recovery of sparse signals in the presence of noise. For instance, Wainwright [18] considered the problem of exact support recovery using the maximum likelihood decoder. Sufficient and necessary conditions are developed for different sparsity levels. Using the same decoder, Fletcher et al. [19] further improved the necessary condition in certain settings. Wang et al. [20] also presented a set of necessary conditions for exact support recovery. Akçakaya and Tarokh

[21] analyzed the performance of a joint typicality decoder and found a set of sufficient and necessary conditions under different performance metrics including the one for exact support recovery.

In this paper, we present sharper asymptotic tradeoffs between the signal dimension  $m$ , the number of nonzero entries  $k$ , and the number of measurements  $n$  for exact support recovery in the noisy setting. When  $k$  is held fixed, we find that  $n = \frac{\log m}{c(\mathbf{X})}$  is sufficient and necessary. We provide a complete characterization of  $c(\mathbf{X})$  that depends on the values of the nonzero entries of  $\mathbf{X}$ . When  $k$  increases in certain manner as specified later, we obtain sufficient and necessary conditions for support recovery which improve upon existing results. Our main results are inspired by the analogy to communication over the additive white Gaussian noise multiple access channel (AWGN-MAC) [22], [23]. According to this connection, the columns of the measurement matrix form a common codebook for all senders. Codewords from each sender are individually multiplied by unknown channel gains, which correspond to nonzero entries of  $\mathbf{X}$ . Then, the noise corrupted linear combination of these codewords is observed. Thus, support recovery can be interpreted as decoding messages from multiple senders. With appropriate modifications, the techniques for deriving channel capacity can be leveraged to provide performance tradeoffs for support recovery. This analytical framework can further incorporate alternate related models of sparse signal recovery, such as random nonzero entries and multiple measurement vectors (MMV).

The rest of the paper is organized as follows. Section II formulates the problem of support recovery. Section III summarizes the main results and makes comparisons with other parallel work. The outline of the proofs of our main results is presented in Section IV. Further extensions to different model assumptions are discussed in Section V.

Throughout this paper, a set is a collection of unique objects. Let  $\mathbb{R}^m$  denote the  $m$ -dimensional real Euclidean space. Let  $[k]$  denote the set  $\{1, 2, \dots, k\}$ ,  $|S|$  denote the cardinality of set  $S$ , and  $\|\mathbf{x}\|$  denote the  $\ell_2$ -norm of a vector  $\mathbf{x}$ . The expression  $f(x) = o(g(x))$  denotes  $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$ ,  $f(x) = O(g(x))$  denotes  $|f(x)| \leq \alpha|g(x)|$  as  $x \rightarrow \infty$  for some constant  $\alpha > 0$ ,  $f(x) = \Theta(g(x))$  denotes  $f(x) = O(g(x))$  and  $g(x) = O(f(x))$ ,  $f(x) = \Omega(g(x))$  denotes  $g(x) = O(f(x))$ , and  $f(x) = \omega(g(x))$  denotes  $g(x) = o(f(x))$ .

## II. FORMAL DEFINITION OF THE PROBLEM

Let  $\mathbf{w} = [w_1, \dots, w_k]^\top \in \mathbb{R}^k$ , where  $w_i \neq 0$  for all  $i$ . Let  $\mathbf{S} = [S_1, \dots, S_k]^\top \in [m]^k$  be such that  $S_1, \dots, S_k$  are chosen uniformly at random from  $[m]$  without replacement. In particular,  $\{S_1, \dots, S_k\}$  is uniformly distributed over all size- $k$  subsets of  $[m]$ . Then, the signal of interest  $\mathbf{X} = \mathbf{X}(\mathbf{w}, \mathbf{S}) \in \mathbb{R}^m$  is generated as

$$X_s = \begin{cases} w_j & \text{if } s = S_j, \\ 0 & \text{if } s \notin \{S_1, \dots, S_k\}. \end{cases} \quad (1)$$

Thus, the support of  $\mathbf{X}$  is  $\text{supp}(\mathbf{X}) = \{S_1, \dots, S_k\}$ . According to the signal model (1),  $|\text{supp}(\mathbf{X})| = k$ . Throughout this paper, we assume  $k$  is known. The signal is said to be sparse when  $k \ll m$ .

We measure  $\mathbf{X}$  through the linear operation

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{Z} \quad (2)$$

where  $\mathbf{A} \in \mathbb{R}^{n \times m}$  is the measurement matrix,  $\mathbf{Z} \in \mathbb{R}^n$  is the measurement noise, and  $\mathbf{Y} \in \mathbb{R}^n$  is the noisy measurement. We further assume that the elements of matrix  $\mathbf{A}$  are independently generated according to  $\mathcal{N}(0, \sigma_a^2)$ , and the noise  $Z_i$  are independently and identically distributed (i.i.d.) according to  $\mathcal{N}(0, \sigma_z^2)$ .

Upon receiving the noisy measurement  $\mathbf{Y}$ , the goal is to recover the support of the sparse signal  $\mathbf{X}$ . Formally, a support recovery map is defined as

$$d : \mathbb{R}^n \mapsto 2^{\{1, 2, \dots, m\}}. \quad (3)$$

Given the signal model (1), the measurement model (2), and the support recovery map (3), the performance metric is defined to be the average probability of error in support recovery, i.e.,  $\mathbb{P}\{d(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}, \mathbf{S}))\}$ , for each (unknown) signal value vector  $\mathbf{w} \in \mathbb{R}^k$ . Note that this probability is taken over the random signal support vector  $\mathbf{S}$ , the measurement matrix  $\mathbf{A}$ , and the noise  $\mathbf{Z}$ .

## III. MAIN RESULTS AND THEIR IMPLICATIONS

### A. Fixed Number of Nonzero Entries

To discover the precise impact of the values of the nonzero entries on support recovery, we consider the support recovery of a sequence of sparse signals generated with the same  $\mathbf{w}$ , which means  $k$  is fixed. Define the auxiliary quantity

$$c(\mathbf{w}) \triangleq \min_{\mathcal{T} \subseteq [k]} \left[ \frac{1}{2|\mathcal{T}|} \log \left( 1 + \frac{\sigma_a^2}{\sigma_z^2} \sum_{j \in \mathcal{T}} w_j^2 \right) \right].$$

The following theorems summarize the performance tradeoff.

*Theorem 1 (Sufficient condition):* If

$$\limsup_{m \rightarrow \infty} \frac{\log m}{n_m} < c(\mathbf{w}) \quad (4)$$

then there exists a sequence of support recovery maps  $\{d^{(m)}\}_{m=k}^\infty$ ,  $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{\{1, 2, \dots, m\}}$ , such that

$$\lim_{m \rightarrow \infty} \mathbb{P}\{d^{(m)}(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}, \mathbf{S}))\} = 0.$$

*Theorem 2 (Necessary condition):* If

$$\limsup_{m \rightarrow \infty} \frac{\log m}{n_m} > c(\mathbf{w}) \quad (5)$$

then for any sequence of support recovery maps  $\{d^{(m)}\}_{m=k}^\infty$ ,  $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{\{1, 2, \dots, m\}}$ , we have

$$\liminf_{m \rightarrow \infty} \mathbb{P}\{d^{(m)}(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}, \mathbf{S}))\} > 0.$$

Theorems 1 and 2 together indicate that  $n = \frac{1}{c(\mathbf{w})} \log m$  is sufficient and necessary for exact support recovery. The leading constant  $\frac{1}{c(\mathbf{w})}$  is explicitly characterized, capturing the role of the magnitudes of the nonzero entries.

### B. Growing Number of Nonzero Entries

Next, we consider the support recovery for the case where  $k$ , the number of nonzero entries, grows with  $m$ , the dimension of the signal. For ease of exposition, we assume that the magnitude of a nonzero entry is bounded from both below and above. This could be a reasonable assumption in practice due to the physical constraints underlying the signal generation and measurement procedure.

We first present a sufficient condition.

*Theorem 3:* Let  $\{\mathbf{w}^{(m)}\}_{m=1}^\infty$  be a sequence of vectors satisfying  $\mathbf{w}^{(m)} \in \mathbb{R}^{k_m}$  and  $0 < w_{\min} \leq |w_j^{(m)}| \leq w_{\max} < \infty$  for all  $j \in [k_m], m \geq 1$ . If

$$\limsup_{m \rightarrow \infty} \frac{1}{n_m} \max_{j \in [k_m]} \left[ \frac{6k_m \log k_m + 2j \log m}{\log(jw_{\min}^2 \sigma_a^2 / \sigma_z^2 + 1)} \right] < 1 \quad (6)$$

then there exists a sequence of support recovery maps  $\{d^{(m)}\}_{m=1}^\infty$ ,  $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{\{1, 2, \dots, m\}}$ , such that

$$\lim_{m \rightarrow \infty} \mathbb{P}\{d^{(m)}(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}^{(m)}, \mathbf{S}))\} = 0.$$

To understand Theorem 3, we note that (6) suggests that the sufficient number of measurements for support recovery be of the order  $n = \max[\Omega(k \log k), \Omega(\frac{k}{\log k} \log m)]$ . The following table lists in detail the sufficient orders of  $n$  paired with different relations between  $m$  and  $k$ .<sup>1</sup>

Relation b/w $m$ and $k$		Sufficient $n$
$m = \omega(k)$	$m = k^{\Omega(\log k)}$	$n = \Omega(\frac{k \log m}{\log k})$
	$e^{\omega(\log k)} \leq m \leq k^{o(\log k)}$ <sup>2</sup>	$n = \Omega(k \log k)$ <sup>3</sup>
	$\omega(k) \leq m \leq e^{\Theta(\log k)}$ <sup>4</sup>	$n = \Omega(k \log m)$
$m = \Theta(k)$		$n = \Omega(k \log m)$

In the existing literature, Wainwright [18] and Akçakaya et al. [21] both derived sufficient conditions for exact support recovery. Under the same assumption of Theorem 3, the sufficient conditions presented in these papers, respectively, are summarized in the following table:

<sup>1</sup>In this work, we mainly focus on the order of  $n$  in terms of  $m$  and  $k$  with other parameters ( $w_{\min}, w_{\max}, \sigma_a^2$ , and  $\sigma_z^2$ ) held fixed.

<sup>2</sup>For example,  $m = k^{\log \log k}$ .

<sup>3</sup>In this case,  $\log m = \omega(\log k)$ , and hence  $\Omega(k \log m) \subset \Omega(k \log k)$ . Thus,  $n = \Omega(k \log k)$  is a better sufficient condition than  $n = \Omega(k \log m)$ .

<sup>4</sup>For example,  $m = k^2$ .

Relation b/w $m$ and $k$	Wainwright [18]	Akçakaya et al. [21]
$m = \omega(k)$	$n = \Omega(k \log \frac{m}{k})$	$n = \Omega(k \log(m - k))$
$m = \Theta(k)$	$n = \Omega(m)$	$n = \Omega(m)$

To compare the results, we first examine the case for  $m = \omega(k)$  (i.e., sublinear sparsity). Note that in the regime where  $m = e^{\omega(\log k)}$ , our sufficient condition on  $n$  includes lower order growth rate, hence is better, than existing results. In the regime where  $\omega(k) \leq m \leq e^{\Theta(\log k)}$ , there exists a certain scenario, e.g.  $k = \frac{m}{\log m}$ , in which our sufficient condition is of the same order as in [21] but higher than in [18]. In the case of  $m = \Theta(k)$  (i.e., linear sparsity), we see that our sufficient condition is stricter, implying its inferiority to existing results in this regime.

Next, we present a necessary condition.

*Theorem 4:* Let  $\{\mathbf{w}^{(m)}\}_{m=1}^{\infty}$  be a sequence of vectors satisfying  $\mathbf{w}^{(m)} \in \mathbb{R}^{k_m}$  and  $0 < w_{\min} \leq |w_j^{(m)}| \leq w_{\max} < \infty$  for all  $j \in [k_m], m \geq 1$ . If

$$\limsup_{m \rightarrow \infty} \frac{2k_m \log(m/k_m)}{n_m \log(2k_m w_{\max}^2 \sigma_a^2 / \sigma_z^2 + 1)} > 1 \quad (7)$$

then for any sequence of support recovery maps  $\{d^{(m)}\}_{m=1}^{\infty}$ ,  $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{\{1, 2, \dots, m\}}$ , we have

$$\liminf_{m \rightarrow \infty} \mathbb{P}\{d^{(m)}(\mathbf{Y}) \neq \text{supp}(\mathbf{X}(\mathbf{w}^{(m)}, \mathbf{S}))\} > 0.$$

Under the same assumption<sup>5</sup> of Theorem 4, we summarize the necessary conditions developed in previous papers below:

Relation b/w $m$ and $k$	$m = \omega(k)$	$m = \Theta(k)$
Wainwright [18]	$n = \Omega(\log m)$	$n = \Omega(\log k)$
Wang et al. [20]	$n = \Omega(\frac{k \log(m/k)}{\log k})$	$n = \Omega(m)$
Akçakaya et al. [21] <sup>6</sup>	$n = \Omega(\frac{k \log(m/k)}{\log k})$	$n = \Omega(\frac{m}{\log k})$
Theorem 4	$n = \Omega(\frac{k \log(m/k)}{\log k})$	$n = \Omega(\frac{m}{\log k})$

We remark that when  $m = \omega(k)$ ,  $n = \Omega(\frac{k \log(m/k)}{\log k})$  is the best known necessary condition. When  $m = \Theta(k)$ ,  $n = \Omega(m)$  is necessary, which follows simply from the elementary constraint  $n \geq k$  that the number of measurements has to be no smaller than the number of nonzero entries for support recovery to be possible.

### C. Further Observations

Note that for the sublinear sparsity with  $m = k^{\Omega(\log k)}$ ,  $\log \frac{m}{k}$  and  $\log m$  are of the same order and hence our sufficient and necessary conditions both indicate  $n = \Omega(\frac{k}{\log k} \log m)$ .

<sup>5</sup>It should be noted that the necessary conditions derived in [18], [20], and [21] were originally under slightly different assumptions. They can be accommodated to fit our assumption. For the purpose of comparison, only the asymptotic order of  $n$  is relevant.

<sup>6</sup>This set of results are implied in [21], by identifying  $C'_4$  and  $C_4$  in Thm. 1.6 and 1.3 therein, respectively, and clarifying the order of  $n$  in different cases. The proof of Thm. 1.3 (and 1.6) states that (below its (25)) asymptotically reliable support recovery is not possible if  $n < \lceil \log(1 + \|\mathbf{w}^{(m)}\|^2 / \sigma_z^2) \rceil^{-1} m H(k/m) - \log(m+1)$ . Note that  $m H(k/m) = \Theta(k \log(m/k))$ . Hence, we consider  $n = \Omega(\frac{k \log(m/k)}{\log k})$  an appropriate necessary condition resulting from the proof in [21].

This provides a sharp performance tradeoff for support recovery in this specific regime, which to our knowledge has not been observed in previous work (see, for example, the remarks in [18, III-A-1]), [19, III-Remark 2]). For the regime where  $\omega(k) \leq m \leq k^{o(\log k)}$ , the orders of  $n$  in any pair of sufficient and necessary conditions have a nontrivial difference, leaving an open question on further narrowing the gap in this remaining regime of sublinear sparsity.

In addition, it is worthwhile to note that our analytical framework could also be adapted to the case where  $w_{\min} = O(1/\sqrt{k})$ . This is a scenario extensively discussed in [18]–[20]. We will not pursue this direction in detail due to the existing development and space limitation.

## IV. MOTIVATION AND PROOF OUTLINE OF MAIN RESULTS

Our work is inspired by the connection between the problems of support recovery of sparse signals and channel coding in network information theory. We start with a discussion on this connection to assist understanding of the proof.

### A. Analogies and Differences to the Problem of Channel Coding

Our previous work [22], [23] established the connection between sparse signal recovery and the channel coding problem. We briefly reiterate the key issues here. First, by removing the columns of  $A$  corresponding to the zero entries in  $\mathbf{X}$ , the effective form of the measurement procedure (2) is given by

$$\mathbf{Y} = \mathbf{A}_{S_1} X_{S_1} + \mathbf{A}_{S_2} X_{S_2} + \dots + \mathbf{A}_{S_k} X_{S_k} + \mathbf{Z}. \quad (8)$$

Note that (8) can be actually interpreted as communication over a Gaussian multiple access channel (MAC) [24] by making the following analogies.

- i) **A nonzero entry as a sender:** We can view the existence of a nonzero entry position  $S_j$  as sender  $j$  that accesses the MAC.
- ii)  $X_{S_j}$  **as the channel gain:** The nonzero entry  $X_{S_j}$ , i.e.,  $w_j$ , plays the role of the channel gain.
- iii)  $\mathbf{A}_i$  **as the codeword:** We treat the measurement matrix  $A$  as a codebook with each column  $\mathbf{A}_i$ ,  $i \in [m]$ , as a codeword. Each element of  $\mathbf{A}_{S_j}$  is fed one by one to the channel as input symbols for the  $j$ th sender, resulting in  $n$  uses of the channel.
- iv) **Similarity of objectives:** In the problem of sparse signal recovery, we focus on finding the support  $\{S_1, \dots, S_k\}$  of the signal. In the problem of MAC communication, the receiver needs to determine the indices of codewords, i.e.,  $S_1, \dots, S_k$ , that are transmitted by senders.

As a natural consequence, the MAC capacity result could be leveraged to provide the performance tradeoff for support recovery of sparse signals. It should be noted that although the analogy between the problems of sparse signal recovery and channel coding has also been observed from various perspectives in parallel work [25], [26, IV-D], [20, II-A], [21, III-A], [27], [28, 11.2], our approach is different by making this connection precise. Especially, we realize that the following domain specific differences from the support recovery problem impose difficulties that must be addressed accordingly in order to rigorously apply the information theoretic approaches.

- i) **Common codebook:** In the MAC communication, each sender uses its own codebook. However, in sparse signal recovery, the “codebook”  $A$  is shared by all “senders”. Hence, all senders operate at the same rate. Different senders will not choose the same codeword, or they will collapse into one sender.
- ii) **Unknown channel gains:** For sparse signal recovery problem, the  $X_{S_i}$  values are unknown. This corresponds to the analogy of unknown channel gains. Although capacity results are available for communication under channel uncertainty, a closer examination indicates that they are not directly applicable to our problem. For instance, training using pilot symbols is common for combating channel uncertainties [29]; however, it is not obvious how to incorporate the training procedure into the measurement model (2).

Next, we develop techniques that are rooted in network information theory, but suitably modified to account for these differences.

### B. Outline of Proofs for Theorem 1 and Theorem 3

Theorems 1 and 3 can be proved in a similar manner. We shall first focus on the proof of Theorem 1. The key steps are outlined as follows.

1) *Estimation of the Values of Nonzero Entries:* This is to estimate  $\mathbf{w}$ . We first form an estimate of  $\|\mathbf{w}\|$  as

$$\hat{W} \triangleq \sqrt{\|\mathbf{Y}\|^2/n - \sigma_z^2/\sigma_a^2}.$$

Fix an  $\epsilon > 0$ . Let  $\mathcal{Q} = \mathcal{Q}(\hat{W}, \epsilon)$  be a minimal set of points in  $\mathbb{R}^k$  satisfying the following properties:

- i)  $\mathcal{Q} \subseteq \mathcal{B}_k(\hat{W})$ , where  $\mathcal{B}_k(\hat{W})$  is the  $k$ -dimensional hypersphere of radius  $\hat{W}$ .
- ii) For any  $\mathbf{b} \in \mathcal{B}_k(\hat{W})$ , there exists  $\hat{\mathbf{W}} \in \mathcal{Q}$  such that  $\|\hat{\mathbf{W}} - \mathbf{b}\| \leq \frac{\epsilon}{2}$ .

It can be shown that with high probability there exists  $\hat{\mathbf{W}} \in \mathcal{Q}$  such that  $\|\hat{\mathbf{W}} - \mathbf{w}\| \leq \epsilon$ . Essentially,  $\mathcal{Q}$  contains a good estimate of  $\mathbf{w}$  while still maintaining a controllable cardinality.

2) *Method of Support Recovery:* This is inspired by the distance decoding technique [30] for MAC. We declare  $d(\mathbf{Y}) = \{\hat{s}_1, \hat{s}_2, \dots, \hat{s}_k\} \subseteq [m]$  is the recovered support of the signal, if it is the unique set of indices such that

$$\frac{1}{n} \left\| \mathbf{Y} - \sum_{j=1}^k \hat{W}_j \mathbf{A}_{\hat{s}_j} \right\|^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \quad (9)$$

for some  $\hat{\mathbf{W}} \in \mathcal{Q}$ . If there is none or more than one such set, pick an arbitrary set of  $k$  indices. Essentially, this recovery method tries all the points in  $\mathcal{Q}$  as candidate estimates of the values of nonzero entries.

3) *Analysis of the Error Probability:* There are two types of error events to consider.

- i) *Error event  $\mathcal{E}_1$ :* There does not exist a  $\hat{\mathbf{W}} \in \mathcal{Q}$  for the true support set to satisfy (9).
- ii) *Error event  $\mathcal{E}_2$ :* There exists a set of indices other than the true support such that with some  $\hat{\mathbf{W}} \in \mathcal{Q}$ , (9) is satisfied.

The remaining task is to find an appropriate condition involving the model parameters (i.e.,  $m, n, k, \mathbf{w}, \sigma_a^2$ , and  $\sigma_z^2$ ) to ensure that the overall error probability diminishes as  $m \rightarrow \infty$ .

First, it can be readily shown that, according to the weak law of large numbers [31], the probability that  $\mathcal{E}_1$  occurs tends to zero as  $m \rightarrow \infty$ . It does not impose any special condition on the model parameters.

Next, let us consider  $\mathcal{E}_2$ . Due to the symmetry of the problem setup, we assume without loss of generality that the true support set is  $[k]$ . Consider a set of  $k$  indices  $\{s_1, s_2, \dots, s_k\} \subseteq [m]$  where  $\{s_1, s_2, \dots, s_k\} \neq [k]$ . Let  $\mathcal{T} = [k] \setminus \{s_1, s_2, \dots, s_k\}$ . Then, motivated by the analysis for distance decoding technique [30], we can show that

$$\begin{aligned} & \mathbb{P}\{\exists \hat{\mathbf{W}} \in \mathcal{Q} \text{ s.t. } \{s_1, s_2, \dots, s_k\} \text{ satisfies (9)}\} \\ & \leq k! \cdot |\mathcal{Q}| \cdot 2^{-\frac{n}{2} \log \frac{(\sum_{j \in \mathcal{T}} w_j^2) \sigma_a^2 + \sigma_z^2 - \epsilon}{\sigma_z^2 + \epsilon^2 \sigma_a^2}} \end{aligned} \quad (10)$$

where the factor  $|\mathcal{Q}|$  can be upper bounded when the measurement is typical in certain sense. There are  $\binom{m-k}{|\mathcal{T}|}$  of such sets of indices whose corresponding  $\mathcal{E}_2$  can be governed by the same probability upper bound (10). By enumerating over  $\mathcal{T} \subseteq [k]$  and using the union of events bound, we can identify (4) as a sufficient condition to ensure the probability that  $\mathcal{E}_2$  occurs tends to zero as  $m \rightarrow \infty$ .

To prove Theorem 3 for the case of growing number of nonzero entries, we consider the following modifications. First, for a consistent estimate of  $\|\mathbf{w}^{(m)}\|$ , we need  $\lim_{m \rightarrow \infty} \frac{k_m}{n_m} = 0$ . Second, one can replace any  $w_i$  by  $w_{\min}$  in (10) to obtain a similar but general probability upper bound. Third, in the resultant general probability upper bound, the factor  $k! \cdot |\mathcal{Q}|$  increases as  $m \rightarrow \infty$ . Its growth rate can be shown as  $\log(k! \cdot |\mathcal{Q}|) = O(k \log k)$ . Taking into consideration all these aspects, one can show that (6) serves as a proper sufficient condition to ensure asymptotically successful support recovery.

### C. Outline of Proofs for Theorem 2 and Theorem 4

The proofs of Theorems 2 and 4 mimic the converse proof of the coding theorem for the AWGN-MAC [24]. Important aspects are summarized as follows.

First, we assume the nonzero vector  $\mathbf{w}^{(m)}$  is known. Since knowing the nonzero values does not make the support recovery more difficult, the necessary condition obtained under this assumption works for the original problem where the nonzero values are unknown.

Once we assume that the values of the nonzero entries are known, the proof techniques for the converse of the channel coding theorem become useful. At the core is Fano's inequality, which provides an upper bound for the entropy of an estimate. Under the assumption that the error probability diminishes as  $m \rightarrow \infty$ , Fano's inequality bridges the model parameters to yield a set of necessary conditions that the model has to satisfy. Theorem 2 indeed reflects this set of necessary conditions precisely. Instead, Theorem 4 selects only one of these necessary conditions for its clear interpretation.

## V. EXTENSIONS

The connection between the problems of support recovery and channel coding can be further explored to provide the

performance tradeoff for different scenarios. We briefly discuss its potential to several important cases.

#### A. Random Nonzero Values

Thus far, the nonzero vector  $\mathbf{w}^{(m)}$  is assumed to be deterministic but unknown for any  $m$ . It is also possible to consider the case where the elements of  $\mathbf{w}^{(m)}$  are generated according to a certain probability distribution. Recall that  $\mathbf{w}^{(m)}$  can be interpreted as the channel gains of a MAC. Under the new assumption, the channel gains will be random. They will be realized once and then kept fixed during the entire channel use. This channel model is usually termed as a slow fading channel [32]. Since there is a nontrivial probability that the channel gains are realized too poorly to support the target rate, an outage analysis is usually employed to evaluate the performance [32]. Correspondingly, some nonzero entries may be too weak to be reliably recovered. Based on this new analogy and the method of outage analysis, the performance analysis for support recovery becomes feasible.

#### B. Multiple Measurement Vectors (MMV)

Recently, increasing research effort has been focusing on the sparse signal recovery with multiple measurement vectors [33], where there are multiple signals  $\mathbf{X}_1(\mathbf{w}_1, \mathbf{S}), \mathbf{X}_2(\mathbf{w}_2, \mathbf{S}), \dots, \mathbf{X}_t(\mathbf{w}_t, \mathbf{S})$  that one would measure. Note that all  $\mathbf{X}_j$  possess the common sparsity profile. That is, the locations of nonzero entries are the same in each  $\mathbf{X}_j$ ,  $j \in [t]$ . Let  $X = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t] \in \mathbb{R}^{m \times t}$ . To measure  $X$ , one performs

$$Y = AX + Z \quad (11)$$

where  $Z = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_t] \in \mathbb{R}^{n \times t}$  is the measurement noise and  $Y = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_t] \in \mathbb{R}^{n \times t}$  is the noisy measurement.

Note that the model (2) can be viewed as a special case of the MMV model (11) with  $t = 1$ . The methodology that has been developed in this paper has the potential to be extended to deal with the performance issues with the MMV model by noting the following connections [22]. First, the same set of columns in  $A$  are scaled by entries in different  $\mathbf{X}_j$  in the generation of the different  $\mathbf{Y}_j$ . The nonzero entries of  $X$  can then be viewed as the coefficients that connect different pairs of inputs and outputs of a channel. Second, each measurement vector  $\mathbf{Y}_j$  can be viewed as the data received at the  $j$ th receiver. Hence, the MMV model corresponds to a multiple-input multiple-output (MIMO) channel model. Third, the goal is to recover the locations of nonzero rows of  $X$  upon receiving  $Y$ . This implies that, in the corresponding channel coding problem, the receivers will fully collaborate to decode the information sent by all senders. We hope that, via proper accommodation of the developed method in this paper, the capacity results for MIMO channels can be leveraged to shed light on the performance tradeoff of sparse signal recovery with MMV. Since the MIMO channel capacity grows as  $\min(k, t)$ , the support recovery can be significantly enhanced in the MMV problem.

#### ACKNOWLEDGMENT

This research was supported by NSF Grants CCF-0830612 and CCF-0747111.

#### REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candes, "Compressive sampling," *Proceedings of the Int. Congress of Mathematicians*, pp. 1433–1452, 2006.
- [3] K. Kreutz-Delgado and B. D. Rao, "Measures and algorithms for best basis selection," *Proc. ICASSP*, vol. 3, pp. 1881–1884, 1998.
- [4] Y. Jin and B. D. Rao, "Algorithms for robust linear regression by exploiting the connection to sparse signal recovery," *ICASSP*, 2010.
- [5] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [6] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *27th Annual Asilomar Conf. Sig. Sys. Comp.*, 1993.
- [7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIREV*, vol. 43, no. 1, pp. 129–159, 2001.
- [8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [9] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted norm minimization algorithm," *IEEE Trans. Sig. Proc.*, vol. 45, no. 3, pp. 600–616, 1997.
- [10] D. Wipf and B. D. Rao, "Sparse bayesian learning for basis selection," *IEEE Trans. Sig. Proc.*, vol. 52, no. 8, pp. 2153–2164, August 2004.
- [11] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Process. Mag.*, pp. 14–30, 2001.
- [12] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," *ICASSP*, pp. 1357–1360, 2007.
- [13] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal  $\ell_1$ -norm solution is also the sparsest solution," *Comm. Pure Appl. Math.*, vol. 59, pp. 797–829, 2004.
- [14] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [15] J. A. Tropp, "Greedy is good: Algorithmic results for sparse approximation," *IEEE Trans. Info. Theo.*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [16] D. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [17] M. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (lasso)," *IEEE Trans. Info. Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.
- [18] —, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Info. Theo.*, Dec 2009.
- [19] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5758–5772, Dec 2009.
- [20] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement," *Proc. of ISIT*, pp. 2197–2201, 2008.
- [21] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 492–504, 2010.
- [22] Y. Jin and B. D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," *ICASSP*, 2008.
- [23] —, "Performance limits of matching pursuit algorithms," *ISIT*, 2008.
- [24] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.
- [25] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing*, 2006.
- [26] J. Tropp, "Just relax: Convex programming methods for identifying sparse signal in noise," *IEEE Trans. Info. Theo.*, vol. 52, no. 3, 2006.
- [27] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," *submitted to IEEE Trans. Information Theory*.
- [28] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *preprint*, 2006.
- [29] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Info. Theory*, vol. 49, no. 4, 2003.
- [30] A. Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Trans. Info. Theory*, vol. 42, pp. 1520–1529, 1996.
- [31] S. I. Resnick, *A probability path*. Birkhauser Boston, 1999.
- [32] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [33] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Sig. Proc.*, vol. 53, no. 7, pp. 2477–2488, 2005.