Limits on Support Recovery of Sparse Signals via Multiple-Access Communication Techniques

Yuzhe Jin, *Student Member, IEEE*, Young-Han Kim, *Member, IEEE*, and Bhaskar D. Rao, *Fellow, IEEE*

***Abstract*—In this paper, we consider the problem of exact support recovery of sparse signals via noisy linear measurements. The main focus is finding the sufficient and necessary condition on the number of measurements for support recovery to be reliable. By drawing an analogy between the problem of support recovery and the problem of channel coding over the Gaussian multiple-access channel (MAC), and exploiting mathematical tools developed for the latter problem, we obtain an information-theoretic framework for analyzing the performance limits of support recovery. Specifically, when the number of nonzero entries of the sparse signal is held fixed, the exact asymptotics on the number of measurements sufficient and necessary for support recovery is characterized. In addition, we show that the proposed methodology can deal with a variety of models of sparse signal recovery, hence demonstrating its potential as an effective analytical tool.**

***Index Terms*—Compressed sensing, Gaussian multiple-access channel (MAC), noisy linear measurement, performance tradeoff, sparse signal, support recovery.**

I. Introduction

CONSIDER the problem of estimating a sparse signal $\mathbf{X} \in \mathbb{R}^m$ in high dimension via noisy linear measurements $\mathbf{Y} = A\mathbf{X} + \mathbf{Z}$, where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix and $\mathbf{Z}$ is the measurement noise. A sparse signal informally refers to a signal whose representation in a certain basis contains a large proportion of zero elements. In this paper, we mainly consider signals that are sparse with respect to the canonical basis of the Euclidean space. The goal is to estimate the sparse signal $\mathbf{X}$ by making as few measurements as possible. This problem has received much attention from many research principles, motivated by a wide spectrum of applications such as compressed sensing [1], [2], biomagnetic inverse problems [3], [4], image processing [5], [6], bandlimited extrapolation and spectral estimation [7], robust regression and outlier detection [8], speech processing [9], channel estimation [10], [11], echo cancellation [12], [13], and wireless communication [10], [14].

Manuscript received March 03, 2010; revised March 11, 2011; accepted June 09, 2011. Date of current version December 07, 2011. This work was supported in part by the National Science Foundation under Grants CCF-0830612 and CCF-0747111. The material in this paper was presented in part at the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing and the 2008 IEEE International Symposium on Information Theory. A short version of this paper was presented at the 2010 IEEE International Symposium on Information Theory.

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093-0407 USA (e-mail: yuzhe.jin@gmail.com; yhk@ucsd.edu; brao@ece.ucsd.edu).

Communicated by J. Romberg, Associate Editor for Signal Processing.
Digital Object Identifier 10.1109/TIT.2011.2170116

Computationally efficient algorithms for sparse signal recovery have been proposed to find or approximate the sparse signal $\mathbf{X}$ in various settings. A partial list includes matching pursuit [15], orthogonal matching pursuit [16], LASSO [17], basis pursuit [18], FOCUSS [3], sparse Bayesian learning [19], finite rate of innovation [20], CoSaMP [21], and subspace pursuit [22]. At the same time, many exciting mathematical tools have been developed to analyze the performance of these algorithms. In particular, Donoho [1], Donoho *et al.* [23], Candès and Tao [24], and Candès *et al.* [25] presented sufficient conditions for ℓ_1-norm minimization algorithms, including basis pursuit, to successfully recover the sparse signals with respect to certain performance metrics. Tropp [26], Tropp and Gilbert [27], and Donoho *et al.* [28] studied greedy sequential selection methods such as matching pursuit and its variants. In these papers, the structural properties of the measurement matrix A, including coherence metrics [15], [23], [26], [29] and spectral properties [1], [24], are used as the major ingredient of the performance analysis. By using random measurement matrices, these results translate to relatively simple tradeoffs between the dimension of the signal $\mathbf{X}$, the number of nonzero entries in $\mathbf{X}$, and the number of measurements to ensure asymptotically successful reconstruction of the sparse signal. When the measurement noise is present, i.e., $\mathbf{Z} \neq 0$, the performance of the sparse signal recovery algorithms has been measured by the Euclidean distance between the true signal and the estimate [23], [25].

In many applications, however, finding the exact support of the signal is important even in the noisy setting. For example, in applications of medical imaging, magnetoencephalography (MEG) and electroencephalography (EEG) are common approaches for collecting noninvasive measurements of external electromagnetic signals [30]. A relatively fine spatial resolution is required to localize the neural electrical activities from a huge number of potential locations [31]. In the domain of cognitive radio, spectrum sensing plays an important role in identifying available spectrum for communication, where estimating the number of active subbands and their locations becomes a nontrivial task [32]. In multiple-user communication systems such as a code-division multiple-access (CDMA) system, the problem of neighbor discovery requires identification of active nodes from all potential nodes in a network based on a linear superposition of the signature waveforms of the active nodes [14]. In all these problems, finding the support of the sparse signal is more important than approximating the signal vector in the Euclidean distance. Hence, it is important to understand performance issues in the exact support recovery of sparse signals with noisy measurements. Information-theoretic tools have proven successful in this direction. Wainwright [33], [34] considered the problem of exact support recovery using the optimal maximum-likelihood decoder. Necessary and sufficient conditions are established for different scalings between the sparsity level and signal dimension. Using the same decoder, Rad [35] derived sharp upper bounds on the error probability of exact support recovery. Meanwhile, Fletcher *et al.* [36], [37] improved the necessary condition with the same decoder. Wang *et al.* [38], [39] also presented a set of necessary conditions for exact support recovery. Akçakaya and Tarokh [40] analyzed the performance of a joint typicality decoder and applied it to find a set of necessary and sufficient conditions under different performance metrics including the one for exact support recovery. In addition, a series of papers have leveraged many information-theoretic tools, including rate-distortion theory [41], [42], expander graphs [43], belief propagation and list decoding [44], and low-density parity-check codes [45], to design novel algorithms for sparse signal recovery and to analyze their performances.

In this paper, we develop sharper asymptotic tradeoffs between the signal dimension m, the number of nonzero entries k, and the number of measurements n for reliable support recovery in the noisy setting. Especially, when k is fixed, we show that $n = (\log m)/c(\mathbf{X})$ is sufficient and necessary. We give a complete characterization of $c(\mathbf{X})$ that depends on the values of all nonzero entries of $\mathbf{X}$. This result provides a clear insight into the role of nonzero entries in support recovery, which improves upon many existing results where only the minimum nonzero magnitude entered the performance tradeoffs. When k increases in certain manners as specified later, we obtain sufficient and necessary conditions for perfect support recovery which can be tight in the order.

Our main results are inspired by the analogy to communication over the Gaussian multiple-access channel (MAC) [46], [47]. According to this connection, the columns of the measurement matrix form a common codebook for all senders. Codewords from the senders 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.

Despite these similarities between the problem of support recovery and that of MAC communication, there are also important differences between them, namely, the common codebook problem and the unknown channel gain problem, which make a straightforward translation of known results nontrivial. We customize tools from multiple-user information theory (e.g., distance decoding and Fano's inequality) to tackle the support recovery problem. Moreover, the analytical framework in this paper can be extended to different models of sparse signal recovery, such as non-Gaussian measurement noise, sources with random activity levels, and multiple measurement vectors (MMVs).

Some analogies between sparse signal recovery (in a broad sense) and channel coding have been observed from various perspectives in parallel work [41], [48, Sec. IV-D], [38, Sec. II-A], [40, Sec. III-A], [28, Sec. 11.2]. We first note that our approach is different from the analytical perspective in [41] where the Gaussian channel capacity and rate-distortion analysis were employed to established design constraints, and is also different from the point-to-point Gaussian channel coding perspective in [48, Sec. IV-D] and [38, Sec. II-A]. In [40, Sec. III-A], the sparse signal recovery problem was related to communication over a single-user multiple-input–single-output (MISO) channel, which was then employed to obtain a necessary condition under the assumption that the channel gains were known at the receiver. Unlike these approaches, we connect the problem of sparse signal recovery explicitly to a MAC communication problem where no coordination exists among senders. The advantage of this approach is evident in our main result that establishes matching sufficient and necessary conditions for reliable support recovery. To be fair, we note that the similarity between sparse signal recovery and multiple-user detection was described in [28, Sec. 11.2], but only at an intuitive level. Here we clarify the connection between the two problems and extend the analytical tool set for multiple-user communication, which is useful particularly in establishing the sufficient condition for support recovery.

The rest of the paper is organized as follows. We formally state the support recovery problem in Section II. To motivate the main results of the paper and their proof techniques, we discuss in Section III the similarities and differences between the support recovery problem and the multiple-access communication problem. Our main results are presented in Section IV, together with comparisons to existing results in the literature. The proofs of the main theorems are presented in Appendixes I–IV, respectively. Section V further extends the results to different signal models and measurement procedures. Section VI concludes the paper with further discussions.

Throughout this paper, a set is a collection of unique objects. Let $\mathbb{R}^m$ denote the m-dimensional real Euclidean space. Let $\mathbb{N} = \{1, 2, 3, \ldots\}$ denote the set of natural numbers. Let $[k]$ denote the set $\{1, 2, \ldots, k\}$. The notation $|S|$ denotes the cardinality of set S, $\|\mathbf{x}\|$ denotes the ℓ_2-norm of a vector $\mathbf{x}$, and $\|A\|_F$ denotes the Frobenius norm of a matrix A. The expression $f(x) = o(g(x))$ denotes $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$, $f(x) = O(g(x))$ denotes $|f(x)| \le \alpha|g(x)|$ as $x \to \infty$ 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. Problem Formulation

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

$$X_s = \begin{cases} w_j, & \text{if } s = S_j \\ 0, & \text{if } s \notin \{S_1, \ldots, S_k\}. \end{cases} \tag{1}$$

Thus, the support of $\mathbf{X}$ is $\mathrm{supp}(\mathbf{X}) = \{S_1, \ldots, S_k\}$. According to the signal model (1), $|\mathrm{supp}(\mathbf{X})| = k$. 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

$$\mathbf{Y} = A\mathbf{X} + \mathbf{Z} \tag{2}$$

where $A \in \mathbb{R}^{n\times m}$ 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 the measurement matrix A are independently generated according to $\mathcal{N}(0, \sigma_a^2)$, and the noise Z_i is independently and identically distributed (i.i.d.) according to the Gaussian distribution $\mathcal{N}(0, \sigma_z^2)$. We assume σ_z^2 is known.

Upon observing the noisy measurement $\mathbf{Y}$, one wishes to recover the support of the sparse signal $\mathbf{X}$. A support recovery map is defined as

$$d : \mathbb{R}^n \longmapsto 2^{[m]}. \tag{3}$$

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.,

$$\mathsf{P}\{d(\mathbf{Y}) \neq \operatorname{supp}(\mathbf{X}(\mathbf{w}, \mathbf{S}))\}$$

for each (unknown) signal value vector $\mathbf{w} \in \mathbb{R}^k$. Note that the probability here is taken over the random signal support vector $\mathbf{S}$, the measurement matrix A, and the noise $\mathbf{Z}$.

III. An Information-Theoretic Perspective on Sparse Signal Recovery

In this section, we will introduce an interpretation of the problem of sparse signal recovery via a communication problem over the Gaussian MAC. The similarities and differences between the two problems will be elucidated, hence progressively unraveling the intuition and facilitating technical preparation for the main results and their proof techniques.

A. Brief Review of the Gaussian MAC

We start by reviewing the background on the k-sender MAC. Suppose the senders wish to transmit information to a common receiver. Each sender i has access to a codebook $\mathcal{C}^{(i)} = \{\mathbf{c}_1^{(i)}, \mathbf{c}_2^{(i)}, \ldots, \mathbf{c}_{m^{(i)}}^{(i)}\}$, where $\mathbf{c}_j^{(i)} \in \mathbb{R}^n$ is a codeword and $m^{(i)}$ is the number of codewords in $\mathcal{C}^{(i)}$. The rate for the sender i is $R^{(i)} = (\log m^{(i)})/n$. To transmit information, each sender chooses a codeword from its codebook, and all senders transmit their codewords simultaneously over a Gaussian MAC [49]

$$Y_l = h_1 X_{1,l} + h_2 X_{2,l} + \cdots + h_k X_{k,l} + Z_l, \qquad l = 1, 2, \ldots, n \tag{4}$$

where $X_{i,l}$ denotes the input symbol from sender i to the channel at transmission time l, h_i denotes the channel gain associated with sender i, Z_l is the additive noise, i.i.d. $\mathcal{N}(0, \sigma_z^2)$, and Y_l is the channel output.

Upon receiving $Y_1, \ldots, Y_n$, the receiver needs to determine the codewords transmitted by each sender. Since the senders interfere with each other, there is an inherent tradeoff among their operating rates. The notion of capacity region is introduced to capture this tradeoff by characterizing all possible rate tuples $(R^{(1)}, R^{(2)}, \ldots, R^{(k)})$ at which reliable communication can be achieved with diminishing probability of decoding error. By assuming each sender obeys the power constraint $\|\mathbf{c}_j^{(i)}\|^2/n \le \sigma_c^2$ for all $j \in [m^{(i)}]$ and all $i \in \mathbb{N}_k$, the capacity region of a Gaussian MAC with known channel gains [49] is

$$\left\{ (R^{(1)}, \ldots, R^{(k)}) : \sum_{i\in\mathcal{T}} R^{(i)} \le \frac{1}{2}\log\left(1 + \frac{\sigma_c^2}{\sigma_z^2}\sum_{i\in\mathcal{T}} h_i^2\right), \forall\, \mathcal{T} \subseteq [k] \right\}. \tag{5}$$

B. Connecting Sparse Signal Recovery to the Gaussian MAC

In the measurement model (2), one can remove the columns in A which are nulled out by zero entries in $\mathbf{X}$ and obtain the following effective form of the measurement procedure:

$$\mathbf{Y} = X_{S_1}\mathbf{a}_{S_1} + \cdots + X_{S_k}\mathbf{a}_{S_k} + \mathbf{Z}. \tag{6}$$

By contrasting (6) to the Gaussian MAC (4), we can draw the following key connections that relate the two problems [46].

1) *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.
2) $\mathbf{a}_j$ *as a codeword*: We treat the measurement matrix A as a codebook with each column $\mathbf{a}_j$, $j \in [m]$, as a codeword. Each element of $\mathbf{a}_{S_i}$ is fed one by one to the channel (4) as the input symbol X_i, resulting in n uses of the channel. The noise $\mathbf{Z}$ and measurement $\mathbf{Y}$ can be related to the channel noise Z and channel output Y in the same fashion.
3) X_{S_i} *as a channel gain*: The nonzero entry X_{S_i} in (6) plays the role of the channel gain h_i in (4). Essentially, we can interpret the vector representation (6) as n consecutive uses of the k-sender Gaussian MAC (4) with appropriate stacking of the inputs/outputs into vectors.
4) *Similarity between objectives*: In the problem of sparse signal recovery, the goal is to find the support $\{S_1, \ldots, S_k\}$ of the signal. In the problem of MAC communication, the receiver's goal is to determine the indices of codewords, i.e., $S_1, \ldots, S_k$, that are transmitted by the senders.

Based on the aforementioned aspects, the two problems share significant similarities which enable leveraging the information-theoretic methods for performance analysis of support recovery of sparse signals. However, as we will see next, there are domain specific differences between the support recovery problem and the channel coding problem that should be addressed accordingly to rigorously apply the information-theoretic approaches.

C. Key Differences

1) *Common codebook*: In MAC communication, each sender uses its own codebook. However, in sparse signal recovery, the "codebook" A is shared by all "senders." All senders choose their codewords from the same codebook and hence operate at the same rate. Different senders will not choose the same codeword, or they will collapse into one sender.
2) *Unknown channel gains*: In MAC communication, the capacity region (5) is valid assuming that the receiver knows

the channel gain h_i [50]. In contrast, for sparse signal recovery problem, X_{S_i} is actually unknown and needs to be estimated. Although coding techniques and capacity results are available for communication with channel uncertainty, a closer examination indicates that those results are not directly applicable to our problem. For instance, channel training with pilot symbols is a common practice to combat channel uncertainty [51]. However, it is not obvious how to incorporate the training procedure into the measurement model (2), and hence the related results are not directly applicable.

Once these differences are properly accounted for, the connection between the problems of sparse signal recovery and channel coding makes available a variety of information-theoretic tools for handling performance issues pertaining to the support recovery problem. Based on techniques that are rooted in channel capacity results, but suitably modified to deal with the differences, we will present the main results of this paper in the next section.

IV. 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 signal value vector $\mathbf{w}$. In particular, we assume that k is fixed. Define the auxiliary quantity

$$c(\mathbf{w}) \triangleq \min_{\mathcal{T}\subseteq[k]} \left[\frac{1}{2|\mathcal{T}|}\log\left(1+\frac{\sigma_a^2}{\sigma_z^2}\sum_{j\in\mathcal{T}} w_j^2\right)\right]. \tag{7}$$

For example, when $k = 2$

$$c(w_1,w_2) = \min\left[\frac{1}{2}\log\left(1+\frac{\sigma_a^2 w_1^2}{\sigma_z^2}\right), \frac{1}{2}\log\left(1+\frac{\sigma_a^2 w_2^2}{\sigma_z^2}\right), \frac{1}{4}\log\left(1+\frac{\sigma_a^2(w_1^2+w_2^2)}{\sigma_z^2}\right)\right].$$

We can see from Section III that this quantity is closely related to the two-sender MAC capacity with equal-rate constraint.

The following two theorems summarize our main results under this setup. The subscript in n_m denotes possible dependence between n and m. The proofs of the theorems are presented in Appendixes I and II, respectively.

Theorem 1: If

$$\limsup_{m\to\infty}\frac{\log m}{n_m} < c(\mathbf{w}) \tag{8}$$

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

$$\lim_{m\to\infty} \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq \operatorname{supp}(\mathbf{X}(\mathbf{w},\mathbf{S}))\} = 0. \tag{9}$$

Theorem 2: If

$$\limsup_{m\to\infty}\frac{\log m}{n_m} > c(\mathbf{w}) \tag{10}$$

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

$$\liminf_{m\to\infty} \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq \operatorname{supp}(\mathbf{X}(\mathbf{w},\mathbf{S}))\} > 0. \tag{11}$$

We provide the following observations. First, Theorems 1 and 2 together indicate that $n = (\log m)/(c(\mathbf{w})\pm\epsilon)$ is sufficient and necessary for exact support recovery. The constant $c(\mathbf{w})$ is explicitly characterized, capturing the role of all nonzero entries of a sparse signal in support recovery. Second, the proof of Theorem 2 for the necessary condition employs the assumption that the values of the nonzero entries are known. Immediately, it follows that even if the values of the nonzero entries are known, the sufficient condition for successfully recovering the support is still given by (8). This observation indicates that the unknown channel gain problem indeed does not pose a serious obstacle in support recovery for the case of fixed k. Further, the benefit of exploiting the connection between sparse signal recovery and multiple-access communication is also supported by the theorems. Resorting to channel capacity results enables us to explicitly extract the constant $c(\mathbf{w})$ and obtain the tight sufficient and necessary conditions.

B. Growing Number of Nonzero Entries

Next, we consider the support recovery for the case where the number of nonzero entries k grows with the dimension of the signal m. We assume that the magnitude of a nonzero entry is bounded from both below and above.

First, we present a sufficient condition for exact support recovery. The proof is given in Appendix III.

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

$$\limsup_{m\to\infty}\frac{1}{n_m}\max_{j\in[k_m]}\left[\frac{6k_m\log k_m + 2j\log\frac{me}{j}}{\log\left(\frac{jw_{\min}^2\sigma_a^2}{\sigma_z^2}+1\right)}\right] < 1 \tag{12}$$

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

$$\lim_{m\to\infty} \mathsf{P}\{d^{(m)}(A\mathbf{X}(\mathbf{w}^{(m)},\mathbf{S})+\mathbf{Z}) \neq \operatorname{supp}(\mathbf{X}(\mathbf{w}^{(m)},\mathbf{S}))\} = 0.$$

Note that, according to our proof technique, the upper bound $w_{\max}$ is not needed for performing support recovery, and it does not appear in the sufficient condition above. In the proof, however, we use the assumption that the nonzero signal values are uniformly bounded from above to show that the probability of error tends to zero as $m \to \infty$. To better understand Theorem 3, we present the following implication of (12) that shows the tradeoffs between the order of n versus m and k.

Corollary 1: Under the assumption of Theorem 3

$$\lim_{m\to\infty} \mathsf{P}\{d^{(m)}(A\mathbf{X}(\mathbf{w}^{(m)},\mathbf{S})+\mathbf{Z}) \neq \operatorname{supp}(\mathbf{X}(\mathbf{w}^{(m)},\mathbf{S}))\} = 0$$

TABLE I
SUFFICIENT CONDITIONS FOR SUPPORT RECOVERY IN DIFFERENT SPARSITY REGIONS [WHEN $w_j = \Theta(1)$]

Relation between m and k		Sufficient n
$k = o(m)$	$k = O(e^{\sqrt{\log m}})$	$n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$
	$\omega(e^{\sqrt{\log m}}) \le k \le o(m)$	$n = \Omega(k \log m)$
$k = \Theta(m)$		$n = \Omega(k \log m)$

TABLE II
SUFFICIENT CONDITIONS FOR SUPPORT RECOVERY IN THE EXISTING LITERATURE [WHEN $w_j = \Theta(1)$]

	$k = o(m)$	$k = \Theta(m)$
Wainwright [34]	$n = \Omega(k \log \frac{m}{k})$	$n = \Omega(m)$
Akçakaya et al. [40]	$n = \Omega(k \log(m-k))$	$n = \Omega(m)$
Rad [35][1]	$n = \max\{\Omega(\frac{k}{\log k} \log \frac{m}{k}), \Omega(k)\}$	$n = \Omega(m)$

provided that

$$n = \max\left\{\Omega(k \log k), \Omega\left(\frac{k}{\log k} \log \frac{m}{k}\right)\right\}.$$

In particular, we have the following:

1) when $k = O(e^{\sqrt{\log m}})$, the sufficient number of measurements is $n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$;
2) when $\omega(e^{\sqrt{\log m}}) \le k \le \Theta(m)$, the sufficient number of measurements is $n = \Omega(k \log m)$.

Table I summarizes the sufficient conditions on n paired with different relations between k and m in Corollary 1.

In the existing literature, Wainwright [34], Akçakaya and Tarokh [40], and Rad [35] 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 Table II.[1]

To compare the results, we first examine the case of $k = o(m)$ (i.e., sublinear sparsity). Note that in the regime where $k = O(e^{\sqrt{\log m}})$, our sufficient condition on n is among the best existing results. In the remaining sublinear regime and in the linear regime, i.e., $\omega(e^{\sqrt{\log m}}) \le k \le \Theta(m)$, our results are not as tight as the best existing results. More discussions will be provided in Section IV-C.

Next, we present a necessary condition, the proof of which is given in Appendix IV.

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

$$\limsup_{m\to\infty} \frac{2k_m \log(m/k_m)}{n_m \log\left(\frac{k_m w_{\max}^2 \sigma_a^2}{\sigma_z^2} + 1\right)} > 1 \tag{13}$$

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

$$\liminf_{m\to\infty} \mathsf{P}\{d^{(m)}(A\mathbf{X}(\mathbf{w}^{(m)}, \mathbf{S}) + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X}(\mathbf{w}^{(m)}, \mathbf{S}))\} > 0.$$

To compare with existing results under the same assumption[2] of Theorem 4, we first note that when $k = \Theta(m)$ (linear sparsity), Theorem 4 indicates $n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$ as the necessary condition. Compared to the best known sufficient condition $n = \Omega(m)$ (see Table II), there is a nontrivial gap. When $k = o(m)$ (sublinear sparsity), we summarize the necessary conditions developed in previous papers in Table III.[3]

TABLE III
NECESSARY CONDITIONS FOR SUPPORT RECOVERY [WHEN $w_j = \Theta(1)$]

	$k = o(m)$
Wainwright [34]	$n = \Omega(\log m)$
Wang et al. [39]	$n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$
Akçakaya et al. [40] [3]	$n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$
Theorem 4	$n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$

In this case, $n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$ is the best known necessary condition.[4]

C. Further Discussions

We offer more insights into the analytical framework and proof techniques.

The sufficient conditions in this paper are derived based on the distance decoding technique which was used in channel decoding problem [52]. In order to perform the distance decoding, the channel gains need to be known or can be estimated. This is in contrary to the fact that the nonzero entries of a sparse signal are unknown, and therefore raises the unknown channel gain problem in Section III-C. To tackle this problem, we employ the following procedure in the proofs for sufficient conditions.

1) Find an estimate of $\|\mathbf{w}\|$, and denote it by $\hat{\rho}$.
2) Find a set $\mathcal{Q}$ of points which can be viewed as ϵ-covering of the k-dimensional hypersphere of radius $\hat{\rho}$. By construction of $\mathcal{Q}$, there exists a $\hat{\mathbf{W}} \in \mathcal{Q}$ such that $\|\hat{\mathbf{W}} - \mathbf{w}\| \le \epsilon$ with high probability.
3) Find $\{\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_k\} \subseteq [m]$ such that
$$\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k} \hat{W}_j \mathbf{A}_{\hat{s}_j}\right\|^2 \le \sigma_z^2 + \epsilon^2 \sigma_a^2 \tag{14}$$
for some $\hat{\mathbf{W}} \in \mathcal{Q}$. We declare $\{\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_k\}$ as the estimated support of the sparse signal. As a byproduct, the

[1] We use Theorem 5 in [35] in the table. The sufficient condition in Corollary 6.6 therein seems to be incorrect.

[2] The necessary conditions derived in [34], [39], and [40] were originally derived under slightly different assumptions. Here we adapted them to compare the asymptotic orders of n.

[3] This result is implied in [40], by identifying C_4' in Theorem 1.6 therein, and clarifying the order of n. The proof of Theorem 1.6 states that [below its (25)] asymptotically reliable support recovery is not possible if $n < [\log(1 + \|\mathbf{w}^{(m)}\|^2/\sigma_z^2)]^{-1} mH(k/m) - \log(m+1)$. Note that $mH(k/m) = \Theta(k \log(m/k))$. Hence, we consider $n = \Omega(\frac{k \log(m/k)}{\log k})$ an appropriate necessary condition resulting from the proof in [40].

[4] Note that when $w_{\max} = \infty$, we can show that $n \ge k$ is necessary for both linear and sublinear sparsity [39]. Hence, when $w_{\max} = \infty$, $n = \max\left\{\Omega(\frac{k}{\log k} \log \frac{m}{k}), \Omega(k)\right\}$ is the best known necessary condition.

elements of the corresponding $\hat{\mathbf{W}}$ can be viewed as estimates of the values of the nonzero entries.

The success of this support recovery procedure is closely related to the estimation quality of $\|\mathbf{w}\|$ and the cardinality of the set $\mathcal{Q}$. Accordingly, our methodology shows different strength in different regions of sparsity levels. First, in the case for fixed number of nonzero entries, consistent estimation of $\|\mathbf{w}\|$ can be obtained, and the cardinality of $\mathcal{Q}$ can be bounded from above. This provides the opportunity to discover the exact sufficient and necessary conditions for successful support recovery. Next, in the case with growing number of nonzero entries, the estimation quality of $\|\mathbf{w}\|$ and the cardinality of $\mathcal{Q}$ must be carefully controlled. To this end, the constraint $k = o(n)$, which is implied by Theorem 3, is needed for the estimation of $\|\mathbf{w}\|$ to be consistent, and $w_{\max}$ as the upper bound for the nonzero magnitudes is needed for controlling the cardinality of $\mathcal{Q}$. Note that for the sublinear sparsity with $k = O(e^{\sqrt{\log m}})$, our sufficient and necessary conditions both indicate $n = \Omega(\frac{k}{\log k} \log \frac{m}{k})$, and hence are tight in terms of order. As k increases with m at a faster rate, our sufficient and necessary conditions have gaps, which is a consequence of the difficulty in consistently estimating $\|\mathbf{w}\|$ and handling the large size of $\mathcal{Q}$.

Another interesting region which has been extensively discussed in previous work is the case where $w_{\min} = O(1/\sqrt{k})$ [34], [37], [38]. Although Theorem 4 can be extended to provide a necessary condition for this case, it does not offer improvement upon existing results. Theorem 3 may not be extended to this scenario, which indicates that our analytical technique for proving sufficient conditions is not suited for this scaling.

V. Extensions

The connection between the problems of support recovery and channel coding can be further explored to provide the performance tradeoff for different models of sparse signal recovery. Next, we discuss its potential to address several important variants.

A. Non-Gaussian Noise

Note that the rules for support recovery, mainly reflected in (20) and (26) in the proof of Theorem 1 in Appendix I, are similar to the method of nearest neighbor decoding in information theory. Following the argument in [52], one can show that by replacing the assumption in (2) on measurement noise $Z_i \sim \mathcal{N}(0, \sigma_z^2)$ by any non-Gaussian noise with $\mathrm{Var}(Z_i) = \sigma_z^2$, the previous sufficient conditions continue to hold.

B. Random Signal Activities

In Theorem 1, $\mathbf{w}$ is assumed to be a fixed vector of nonzero entries. We now relax this condition to allow random $\mathbf{W}$, which leads to sparse signals whose nonzero entries are randomly generated and located. For simplicity of exposition, assume that k is fixed. Interestingly, the model (2) with this new assumption can now be contrasted to a MAC with random channel gains

$$Y_l = H_1 X_{1,l} + H_2 X_{2,l} + \cdots + H_k X_{k,l} + Z_l, \quad l = 1, 2, \ldots, n. \quad (15)$$

The difference between (15) and (4) is that the channel gains H_i are random variables in this case. Specifically, in order to contrast the problem of support recovery of sparse signals, H_i should be considered as being realized once and then kept fixed during the entire channel use [46]. This channel model is usually termed as a slow fading channel [50].

The following theorem states the performance of support recovery of sparse signals under random signal activities.

Theorem 5: Suppose $\mathbf{W}$ has bounded support, and $\limsup_{m\to\infty} \frac{\log m}{n_m} = r$. Then, there exists a sequence of support recovery maps $\{d^{(m)}\}_{m=k}^{\infty}$, $d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{[m]}$, such that

$$\limsup_{m\to\infty} \mathsf{P}\{d^{(m)}(A^{(m)}\mathbf{X}(\mathbf{W},\mathbf{S}) + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X})\} \leq \mathsf{P}\{c(\mathbf{W}) \leq r\}$$

where $c(\mathbf{W})$ is defined as in (7).

Proof: Note that

$$\begin{aligned}
&\limsup_{m\to\infty} \mathsf{P}\{d^{(m)}(A^{(m)}\mathbf{X}(\mathbf{W},\mathbf{S}) + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X})\} \\
&= \limsup_{m\to\infty} \int_{\mathbf{w}} \mathsf{P}\{d^{(m)}(A^{(m)}\mathbf{X}(\mathbf{w},\mathbf{S}) + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X})\} \cdot dF(\mathbf{w}) \\
&= \limsup_{m\to\infty} \int_{\mathbf{w}:c(\mathbf{w})>r} \mathsf{P}\{d^{(m)}(A^{(m)}\mathbf{X} + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X})\} \cdot dF(\mathbf{w}) \\
&\quad + \limsup_{m\to\infty} \int_{\mathbf{w}:c(\mathbf{w})\leq r} \mathsf{P}\{d^{(m)}(A^{(m)}\mathbf{X} + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X})\} \cdot dF(\mathbf{w}) \\
&\leq \int_{\mathbf{w}:c(\mathbf{w})>r} \limsup_{m\to\infty} \mathsf{P}\{d^{(m)}(A^{(m)}\mathbf{X} + \mathbf{Z}) \neq \mathrm{supp}(\mathbf{X})\} \cdot dF(\mathbf{w}) + \int_{\mathbf{w}:c(\mathbf{w})\leq r} dF(\mathbf{w}) \quad (16) \\
&\leq \mathsf{P}\{c(\mathbf{W}) \leq r\} \quad (17)
\end{aligned}$$

where (16) follows from Fatou's lemma [53] and (17) follows by applying the proof of Theorem 1 to the integrand. □

Theorem 5 implies that generally, rather than having a diminishing error probability, we have to tolerate certain error probability which is upperbounded by $\mathsf{P}(c(\mathbf{W}) \leq r)$, when the nonzero values are randomly generated. Conversely, in order to design a system with probability of success at least $(1 - p)$, one can find r that satisfies $\mathsf{P}(c(\mathbf{W}) \leq r) \leq p$. Note that $\mathsf{P}\{c(\mathbf{W}) \leq r\}$ can be viewed as the outage probability of a slow fading MAC given the target rate r of each sender [50]. Thus, $\mathsf{P}\{c(\mathbf{W}) \leq r\}$ represents the probability that the channel gains are realized too poorly to support the target rate.

C. Multiple Measurement Vectors

Recently, increasing research effort has been focused on sparse signal recovery with MMVs [54]–[58]. In this problem, we wish to measure multiple sparse signals $\mathbf{X}_1(\mathbf{w}_1, \mathbf{S}), \mathbf{X}_2(\mathbf{w}_2, \mathbf{S}), \ldots,$ and $\mathbf{X}_t(\mathbf{w}_t, \mathbf{S})$ that possess a common sparsity profile, that is, the locations of nonzero entries are the same in each $\mathbf{X}_t$. We use the same measurement matrix $A \in \mathbb{R}^{n\times m}$ to perform

$$Y = AX + Z \tag{18}$$

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

Note that model (2) can be viewed as a special case of the MMV model (18) with $t = 1$. The methodology that has been developed in this paper has a potential to be extended to deal with the performance issues for the MMV model by noting the following connections to channel coding [46]. First, the same set of columns in A are scaled by entries in different $\mathbf{X}_j$, forming outputs as elements in 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 received symbols at receiver antenna j, and hence the MMV model indeed corresponds to a single-input–multiple-output (SIMO) MAC. Third, the aim is to recover the locations of nonzero rows of X upon receiving Y. This implies that, in the language of SIMO MAC communication, the receiver will decode the information sent by all senders through multiple receiver antennas. Via proper accommodation of the method developed in this paper, the capacity results for the SIMO MAC can be leveraged to shed light on the performance tradeoff of sparse signal recovery with MMV.

VI. Concluding Remarks

In this paper, we developed techniques rooted in multiple-user information theory to address the performance issues in the exact support recovery of sparse signals, and discovered necessary and sufficient conditions on the number of measurements. It is worthwhile to note that the interpretation of sparse signal recovery as MAC communication opens new avenues to different theoretic and algorithmic problems in sparse signal recovery. We conclude this paper by briefly discussing several interesting potential directions stemming from this interpretation.

1) Among the large collection of algorithms for sparse signal recovery, the sequential selection methods, including matching pursuit [15] and orthogonal matching pursuit (OMP) [16], determine one nonzero entry at a time, remove its contribution in the residual signal, and repeat this procedure until a certain stopping criterion is satisfied. In contrast, the class of convex relaxation methods, including basis pursuit [18] and LASSO [17], jointly estimate the nonzero entries. The sequential selection methods can be potentially viewed as successive interference cancellation (SIC) decoding [50] for MACs, whereas the convex relaxation methods can be viewed as joint decoding. It would be interesting to ask whether one can make these analogies more precise and use them to address performance issues of these methods. Similarities at an intuitive level between OMP and SIC have been discussed in [47] with performance results supported by empirical evidence. More insights are yet to be explored.
2) The design of channel codes and the development of decoding methods have been extensively studied in the contexts of information theory and wireless communication. Some of these ideas have been transformed into design principles for sparse signal recovery [43]–[45], [59], [60]. Thus far, however, the efforts in utilizing the codebook designs and decoding methods are mainly focused on the point-to-point channel model, which implies that the recovery methods iterate between first recovering one nonzero entry or a group of nonzero entries by treating the rest of them as noise and then removing the recovered nonzero entries from the residual signal. In this paper, we established the analogy between the sparse signal recovery and the multiple-access communication. It motivates us to envision opportunities beyond a point-to-point channel model. One important question is, for example, whether we can develop practical codes for joint decoding and reconstruction techniques to simultaneously recover all the nonzero entries.

Appendix I
Proof of Theorem 1

The proof of Theorem 1 employs the distance decoding technique [52]. Let $\mathbf{A}_j$ denote the jth column of A.

For simplicity of exposition, we describe the support recovery procedure for two distinct cases on the number of nonzero entries.

Case 1: $k = 1$: In this case, the signal of interest is $\mathbf{X} = \mathbf{X}(w_1, S_1)$. Consider the following support recovery procedures. Fix $\epsilon > 0$. First form an estimate $\hat{\rho}$ of $|w_1|$ as

$$\hat{\rho} \triangleq \sqrt{\frac{\left|\frac{1}{n_m}\|\mathbf{Y}\|^2 - \sigma_z^2\right|}{\sigma_a^2}}. \tag{19}$$

Declare that $\hat{s}_1 \in [m]$ is the estimated location for the nonzero entry, i.e., $d^{(m)}(\mathbf{Y}) = \{\hat{s}_1\}$, if it is the unique index such that

$$\frac{1}{n_m}\|\mathbf{Y} - (-1)^q \hat{\rho}\mathbf{A}_{\hat{s}_1}\|^2 \le \sigma_z^2 + \epsilon^2\sigma_a^2 \tag{20}$$

for either $q = 1$ or $q = 2$. If there is none or more than one, pick an arbitrary index.

We now analyze the average probability of error

$$\mathsf{P}(\mathcal{E}) = \mathsf{P}\{d^{(m)}(\mathbf{Y}) \ne \{S_1\}\}.$$

Due to the symmetry in the problem and the measurement matrix generation, we assume without loss of generality $S_1 = 1$, that is

$$\mathbf{Y} = w_1\mathbf{A}_1 + \mathbf{Z}$$

for some w_1. In the following analysis, we drop superscripts and subscripts on m for notational simplicity when no ambiguity arises. Define the events for $s \in [m]$

$$\mathcal{E}_s \triangleq \left\{\exists q \in \{1,2\} \text{ such that } \frac{1}{n}\|\mathbf{Y} - (-1)^q\hat{\rho}\mathbf{A}_s\|^2 \le \sigma_z^2 + \epsilon^2\sigma_a^2\right\}.$$

Then

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}\left(\mathcal{E}_1^c \cup \left(\cup_{s=2}^m \mathcal{E}_s\right)\right). \tag{21}$$

Let

$$\mathcal{E}_{\text{aux}} \triangleq \{\hat{\rho} - |w_1| \in (-\epsilon, \epsilon)\} \cap \left\{\frac{1}{n}\|\mathbf{Y}\|^2 - [w_1^2\sigma_a^2 + \sigma_z^2] \in (-\epsilon, \epsilon)\right\}.$$

Then, by the union of events bound and the fact that $\mathcal{A}^c \cup \mathcal{B} = \mathcal{A}^c \cup (\mathcal{B} \cap \mathcal{A})$

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_1^c) + \sum_{s=2}^m \mathsf{P}(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}). \tag{22}$$

We bound each term in (22). First, by the weak law of large numbers (LLN), $\lim_{m\to\infty} \mathsf{P}(\mathcal{E}_{\text{aux}}^c) = 0$. Next, we consider $\mathsf{P}(\mathcal{E}_1^c)$. If $w_1 > 0$

$$\begin{aligned}\frac{1}{n}\|\mathbf{Y} - \hat{\rho}\mathbf{A}_1\|^2 &= \frac{1}{n}\|w_1\mathbf{A}_1 + \mathbf{Z} - \hat{\rho}\mathbf{A}_1\|^2 \\ &= (w_1 - \hat{\rho})^2\frac{\|\mathbf{A}_1\|^2}{n} + 2(w_1 - \hat{\rho})\frac{\mathbf{A}_1^{\intercal}\mathbf{Z}}{n} + \frac{\|\mathbf{Z}\|^2}{n}. \end{aligned} \tag{23}$$

For any $\epsilon_1 > 0$, as $m \to \infty$, by the LLN

$$\mathsf{P}\left(\{w_1 - \hat{\rho} \in (-\epsilon_1, \epsilon_1)\} \cap \left\{\frac{\|\mathbf{A}_1\|^2}{n} - \sigma_a^2 \in (-\epsilon_1, \epsilon_1)\right\}\right) \to 1.$$

Hence, we have for the first term in (23)

$$\mathsf{P}\left((w_1 - \hat{\rho})^2\frac{\|\mathbf{A}_1\|^2}{n} \in [0, \epsilon_1^2\sigma_a^2 + \epsilon_1^3)\right) \to 1.$$

Following a similar reasoning using LLN, for the second term in (23)

$$\mathsf{P}\left((w_1 - \hat{\rho})\frac{\mathbf{A}_1^{\intercal}\mathbf{Z}}{n} \in (-\epsilon_1^2, \epsilon_1^2)\right) \to 1$$

and for the third term

$$\mathsf{P}\left(\frac{\|\mathbf{Z}\|^2}{n} \in (\sigma_z^2 - \epsilon_1, \sigma_z^2 + \epsilon_1)\right) \to 1.$$

Therefore, for any $\epsilon_1 > 0$

$$\lim_{m\to\infty} \mathsf{P}\left(\frac{1}{n}\|\mathbf{Y} - \hat{\rho}\mathbf{A}_1\|^2 \in (\sigma_z^2 - \epsilon_1, \sigma_z^2 + \epsilon_1)\right) = 1$$

which implies that

$$\lim_{m\to\infty} \mathsf{P}\left(\frac{1}{n}\|\mathbf{Y} - \hat{\rho}\mathbf{A}_1\|^2 \le \sigma_z^2 + \epsilon^2\sigma_a^2\right) = 1.$$

Similarly, if $w_1 < 0$

$$\lim_{m\to\infty} \mathsf{P}\left(\frac{1}{n}\|\mathbf{Y} + \hat{\rho}\mathbf{A}_1\|^2 \le \sigma_z^2 + \epsilon^2\sigma_a^2\right) = 1.$$

Hence, $\lim_{m\to\infty} \mathsf{P}(\mathcal{E}_1^c) = 0$.

For the third term in (22), we need the following lemma, whose proof is presented at the end of this Appendix.

Lemma 1: Let $0 < \beta < \alpha$. Let $\{u_i\}_{i=1}^n$ be a real sequence satisfying

$$\frac{1}{n}\sum_{i=1}^n u_i^2 \in (\alpha - \beta, \alpha + \beta).$$

Let $\{V_i\}_{i=1}^n$ be an i.i.d. random sequence where $V_i \sim \mathcal{N}(0, \sigma_v^2)$. Then, for any $\gamma \in (0, \alpha - \beta)$

$$\mathsf{P}\left(\frac{1}{n}\sum_{i=1}^n (u_i - V_i)^2 \le \gamma\right) \le 2^{-\frac{n}{2}\log\left(\frac{\alpha-\beta}{\gamma}\right)}.$$

Continuing the proof of Theorem 1, we consider $\mathsf{P}(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}})$ for $s \ne 1$. Then

$$\begin{aligned}\mathsf{P}(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}) &\le \mathsf{P}(\mathcal{E}_s|\mathcal{E}_{\text{aux}}) \\ &= \int_{\mathbf{y}\in\mathcal{E}_{\text{aux}}} \mathsf{P}(\mathcal{E}_s|\{\mathbf{Y} = \mathbf{y}\} \cap \mathcal{E}_{\text{aux}}) f(\mathbf{y}|\mathcal{E}_{\text{aux}}) d\mathbf{y}.\end{aligned}$$

Since $\mathbf{A}_s$ is independent of $\mathbf{Y}$ and $\hat{\rho}$, it follows from the definition of $\mathcal{E}_{\text{aux}}$ and Lemma 1 (with $\alpha = w_1^2\sigma_a^2 + \sigma_z^2$ and $\gamma = \sigma_z^2 + \epsilon^2\sigma_a^2$) that

$$\begin{aligned}&\mathsf{P}\left(\frac{1}{n}\|\mathbf{Y} - (-1)^q\hat{\rho}\mathbf{A}_s\|^2 \le \sigma_z^2 + \epsilon^2\sigma_a^2 \,\middle|\, \{\mathbf{Y} = \mathbf{y}\} \cap \mathcal{E}_{\text{aux}}\right) \\ &\quad \le 2^{-\frac{n}{2}\log\left(\frac{w_1^2\sigma_a^2 + \sigma_z^2 - \epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}\right)}\end{aligned}$$

for $q = 1, 2$, if ϵ is sufficiently small. Thus

$$\mathsf{P}(\mathcal{E}_s|\{\mathbf{Y} = \mathbf{y}\} \cap \mathcal{E}_{\text{aux}}) \le 2 \cdot 2^{-\frac{n}{2}\log\left(\frac{w_1^2\sigma_a^2 + \sigma_z^2 - \epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}\right)}$$

and therefore

$$\sum_{s=2}^m \mathsf{P}(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}) \le 2m \cdot 2^{-\frac{n}{2}\log\left(\frac{w_1^2\sigma_a^2 + \sigma_z^2 - \epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}\right)}$$

which tends to zero as $m \to \infty$, if

$$\limsup_{m\to\infty} \frac{\log m}{n_m} < \frac{1}{2}\log\left(\frac{w_1^2\sigma_a^2 + \sigma_z^2 - \epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}\right). \tag{24}$$

Therefore, by (22), the probability of error $\mathsf{P}(\mathcal{E})$ tends to zero as $m \to \infty$, if (24) is satisfied. Finally, since $\epsilon > 0$ is chosen arbitrarily, we have the desired proof of Theorem 1.

Case 2: $k \ge 2$: In this case, the signal of interest is $\mathbf{X} = \mathbf{X}(\mathbf{w}, \mathbf{S})$, where $\mathbf{w} = [w_1, \ldots, w_k]^{\intercal}$ and $\mathbf{S} = [S_1, \ldots, S_k]^{\intercal}$. Consider the following support recovery procedures. Fix $\epsilon > 0$. First, form an estimate $\hat{\rho}$ of $\|\mathbf{w}\|$ as

$$\hat{\rho} \triangleq \sqrt{\frac{\left|\frac{1}{n}\|\mathbf{Y}\|^2 - \sigma_z^2\right|}{\sigma_a^2}}. \tag{25}$$

For $r, \zeta > 0$, let $\mathcal{Q} = \mathcal{Q}(r, \zeta)$ be a minimal set of points in $\mathbb{R}^k$ satisfying the following properties.

i) $\mathcal{Q} \subseteq \mathcal{B}_k(r)$, where $\mathcal{B}_k(r)$ is the k-dimensional hypersphere of radius r, i.e., $\mathcal{B}_k(r) \triangleq \{\mathbf{b} : \mathbf{b} \in \mathbb{R}^k, \|\mathbf{b}\| = r\}$,

ii) For any $\mathbf{b} \in \mathcal{B}_k(r)$, there exists $\hat{\mathbf{w}} \in \mathcal{Q}$ such that $\|\hat{\mathbf{w}} - \mathbf{b}\| \leq \frac{\zeta}{2}$.

The following properties are useful.

Lemma 2:

1) $\lim_{m\to\infty} \mathsf{P}(\exists \hat{\mathbf{W}} \in \mathcal{Q}(\hat{\rho}, \zeta)$ such that $\|\hat{\mathbf{W}} - \mathbf{w}\| < \zeta) = 1$.

2) $q(r, \zeta) \triangleq |\mathcal{Q}(r, \zeta)|$ is monotonically nondecreasing in r for fixed ζ.

Lemma 2–1) will be proved at the end of this Appendix, whereas Lemma 2–2) is obvious.

Given $\hat{\rho}$ and ϵ, fix $\mathcal{Q} = \mathcal{Q}(\hat{\rho}, \epsilon)$. Declare $d(\mathbf{Y}) = \{\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_k\} \subseteq [m]$ is the recovered support of the signal, if it is the unique set of indices such that

$$\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k} \hat{W}_j \mathbf{A}_{\hat{s}_j}\right\|^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \tag{26}$$

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.

Next, we analyze the average probability of error

$$\mathsf{P}(\mathcal{E}) = \mathsf{P}\{d^{(m)}(\mathbf{Y}) \neq \{S_1, \ldots, S_k\}\}.$$

As before, we assume without loss of generality that $S_j = j$ for $j = 1, 2, \ldots, k$, which gives

$$\mathbf{Y} = \sum_{j=1}^{k} w_j \mathbf{A}_j + \mathbf{Z}$$

for some $\mathbf{w}$. Define the event

$$\mathcal{E}_{s_1,s_2,\ldots,s_k} \triangleq \left\{ \begin{array}{l} \exists \hat{\mathbf{W}} \in \mathcal{Q} \text{ and } \{s_1', s_2', \ldots, s_k'\} = \{s_1, s_2, \ldots, s_k\} \\ \quad \text{such that } \frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k} \hat{W}_j \mathbf{A}_{s_j'}\right\|^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \end{array} \right\}.$$

Then

$$\begin{aligned}
\mathsf{P}(\mathcal{E}) &= \mathsf{P}\left(\mathcal{E}_{1,2,\ldots,k}^c \cup \left(\bigcup_{s_1<\cdots<s_k:\{s_1,\ldots,s_k\}\neq[k]} \mathcal{E}_{s_1,s_2,\ldots,s_k}\right)\right) \\
&\leq \mathsf{P}\Bigg(\mathcal{E}_{\text{aux}}^c \cup \mathcal{E}_{1,2,\ldots,k}^c \\
&\qquad \cup \left(\bigcup_{s_1<\cdots<s_k:\{s_1,\ldots,s_k\}\neq[k]} (\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}})\right)\Bigg) \\
&\leq \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) \\
&\quad + \sum_{s_1<\cdots<s_k:\{s_1,\ldots,s_k\}\neq[k]} \mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}})
\end{aligned} \tag{27}$$

where in this case

$$\begin{aligned}
\mathcal{E}_{\text{aux}} \triangleq\ & \{\hat{\rho} - \|\mathbf{w}\| \in (-\epsilon, \epsilon)\} \\
& \cap \left(\bigcap_{j=1}^{k} \left\{\frac{1}{n}\|\mathbf{A}_j\|^2 - \sigma_a^2 \in (-\epsilon, \epsilon)\right\}\right) \\
& \cap \left(\bigcap_{j=1}^{k} \bigcap_{l=j+1}^{k} \left\{\frac{1}{n}\mathbf{A}_j^\intercal \mathbf{A}_l \in (-\epsilon, \epsilon)\right\}\right) \\
& \cap \left(\bigcap_{j=1}^{k} \left\{\frac{1}{n}\mathbf{A}_j^\intercal \mathbf{Z} \in (-\epsilon, \epsilon)\right\}\right) \\
& \cap \left\{\frac{1}{n}\|\mathbf{Z}\|^2 - \sigma_z^2 \in (-\epsilon, \epsilon)\right\}.
\end{aligned}$$

We now bound the terms in (27). First, by the LLN, $\lim_{m\to\infty} \mathsf{P}(\mathcal{E}_{\text{aux}}^c) = 0$. Next, we consider $\mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c)$. Note that, for any $\hat{\mathbf{W}} \in \mathcal{Q}$

$$\begin{aligned}
&\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k} \hat{W}_j \mathbf{A}_j\right\|^2 \\
&= \frac{1}{n}\left\|\sum_{j=1}^{k} w_j \mathbf{A}_j + \mathbf{Z} - \sum_{j=1}^{k} \hat{W}_j \mathbf{A}_j\right\|^2 \\
&= \frac{1}{n}\sum_{j=1}^{k}\sum_{l=1}^{k} (w_j - \hat{W}_j)(w_l - \hat{W}_l)\mathbf{A}_j^\intercal \mathbf{A}_l \\
&\quad + \frac{2}{n}\sum_{j=1}^{k} (w_j - \hat{W}_j)\mathbf{A}_j^\intercal \mathbf{Z} + \frac{1}{n}\|\mathbf{Z}\|^2.
\end{aligned} \tag{28}$$

By applying the LLN to each term in (28), as similarly done in Case 1, and using Lemma 2–1), we have

$$\lim_{m\to\infty} \mathsf{P}\left(\exists \hat{\mathbf{W}} \in \mathcal{Q} \text{s.t. } \frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k} \hat{W}_j \mathbf{A}_j\right\|^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2\right) = 1$$

which implies that $\lim_{m\to\infty} \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) = 0$.

Next, we consider $\mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}})$ for $\{s_1, s_2, \ldots, s_k\} \neq [k]$. Note that

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}) \\
&\quad \leq \mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k} | \mathcal{E}_{\text{aux}}) \\
&\quad = \int\cdots\int_{\{\mathbf{a}_1,\ldots,\mathbf{a}_k,\mathbf{z}\}\in\mathcal{E}_{\text{aux}}} \mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k} | \{\mathbf{A}_1 = \mathbf{a}_1\} \cap \cdots \\
&\qquad\qquad \cap \{\mathbf{A}_k = \mathbf{a}_k\} \cap \{\mathbf{Z} = \mathbf{z}\} \cap \mathcal{E}_{\text{aux}}) \\
&\qquad \times f(\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{z} | \mathcal{E}_{\text{aux}}) d\mathbf{a}_1 \cdots d\mathbf{a}_k d\mathbf{z}.
\end{aligned} \tag{29}$$

For notational simplicity, define $\xi \triangleq \sigma_z^2 + \epsilon^2 \sigma_a^2$, $\mathcal{T} \triangleq \{s_1, s_2, \ldots, s_k\} \cap [k]$, $\mathcal{T}^c \triangleq \{s_1, s_2, \ldots, s_k\} \backslash \mathcal{T}$, and $\mathcal{E}_{\text{cond}} \triangleq \{\mathbf{A}_1 = \mathbf{a}_1\} \cap \cdots \cap \{\mathbf{A}_k = \mathbf{a}_k\} \cap \{\mathbf{Z} = \mathbf{z}\} \cap \mathcal{E}_{\text{aux}}$.

For any permutation $(s'_1, s'_2, \ldots, s'_k)$ of $\{s_1, s_2, \ldots, s_k\}$ and any $\hat{\mathbf{W}} \in \mathcal{Q}$

$$\begin{aligned}
&\mathsf{P}\left(\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k}\hat{W}_j\mathbf{A}_{s'_j}\right\|^2 \le \xi \middle| \mathcal{E}_{\text{cond}}\right) \\
&= \mathsf{P}\left(\frac{1}{n}\left\|\sum_{j=1}^{k} w_j\mathbf{A}_j + \mathbf{Z} - \sum_{j=1}^{k}\hat{W}_j\mathbf{A}_{s'_j}\right\|^2 \le \xi \middle| \mathcal{E}_{\text{cond}}\right) \\
&= \mathsf{P}\left(\frac{1}{n}\left\|\left[\sum_{j=1}^{k} w_j\mathbf{A}_j - \sum_{s'_j\in\mathcal{T}}\hat{W}_j\mathbf{A}_{s'_j} + \mathbf{Z}\right]\right.\right. \\
&\qquad\left.\left. - \sum_{s'_j\in\mathcal{T}^c}\hat{W}_j\mathbf{A}_{s'_j}\right\|^2 \le \xi \middle| \mathcal{E}_{\text{cond}}\right). \qquad (30)
\end{aligned}$$

Conditioned on $\mathcal{E}_{\text{cond}}$ and the chosen $\mathcal{Q}$, $\frac{1}{n}\left\|\sum_{j=1}^{k} w_j\mathbf{A}_j - \sum_{s'_j\in\mathcal{T}}\hat{W}_j\mathbf{A}_{s'_j} + \mathbf{Z}\right\|^2$ is a fixed quantity satisfying

$$\begin{aligned}
&\frac{1}{n}\left\|\sum_{j=1}^{k} w_j\mathbf{A}_j - \sum_{s'_j\in\mathcal{T}}\hat{W}_j\mathbf{A}_{s'_j} + \mathbf{Z}\right\|^2 \\
&\in \left(\left[\sum_{j\in[k]\setminus\mathcal{T}} w_j^2 + \sum_{s'_j\in\mathcal{T}}(w_{s'_j} - \hat{W}_j)^2\right]\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon,\right. \\
&\qquad \left.\left[\sum_{j\in[k]\setminus\mathcal{T}} w_j^2 + \sum_{s'_j\in\mathcal{T}}(w_{s'_j} - \hat{W}_j)^2\right]\sigma_a^2 + \sigma_z^2 + \delta_1\epsilon\right)
\end{aligned}$$

for some positive δ_1 that depends on $\mathbf{w}$ and ϵ only, and is nondecreasing in ϵ. Meanwhile, $\mathbf{A}_{s'_j}$ is independent of $\mathbf{A}_1, \ldots, \mathbf{A}_k$, and $\mathbf{Z}$ for $s'_j \in \mathcal{T}^c$. Hence, by Lemma 1 (with $\alpha = \left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2 + \sum_{s'_j\in\mathcal{T}}(w_{s'_j} - \hat{W}_j)^2\right)\sigma_a^2 + \sigma_z^2$ and $\gamma = \sigma_z^2 + \epsilon^2\sigma_a^2$), (30) is upperbounded by

$$\begin{aligned}
&2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2 + \sum_{s'_j\in\mathcal{T}}\left(w_{s'_j} - \hat{W}_j\right)^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}} \\
&\le 2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}}.
\end{aligned}$$

Hence, by the union of events bound

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k}|\mathcal{E}_{\text{cond}}) \\
&\le \sum_{\{s'_1,\ldots,s'_k\}=\{s_1,\ldots,s_k\}} \\
&\quad \times \mathsf{P}\left(\exists\hat{\mathbf{W}}\in\mathcal{Q}\text{s.t. }\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k}\hat{W}_j\mathbf{A}_{s'_j}\right\|^2 \le \xi \middle| \mathcal{E}_{\text{cond}}\right) \\
&\le \sum_{\{s'_1,\ldots,s'_k\}=\{s_1,\ldots,s_k\}}\ \sum_{\hat{\mathbf{W}}\in\mathcal{Q}} \\
&\quad \times \mathsf{P}\left(\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^{k}\hat{W}_j\mathbf{A}_{s'_j}\right\|^2 \le \xi \middle| \mathcal{E}_{\text{cond}}\right) \\
&\le k!\cdot|\mathcal{Q}|\cdot 2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}}.
\end{aligned}$$

Furthermore, conditioned on $\mathcal{E}_{\text{aux}}$, $\hat{\rho} < \|\mathbf{w}\| + \epsilon$ and hence $|\mathcal{Q}| \le q(\|\mathbf{w}\| + \epsilon, \epsilon)$ by Lemma 2–2). Thus

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}) \\
&\le k!\cdot q(\|\mathbf{w}\| + \epsilon, \epsilon)\cdot 2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}}. \qquad (31)
\end{aligned}$$

Note that the probability upperbound (31) depends on $s_1, \ldots, s_k$ only through $\mathcal{T}$. Grouping the $\binom{m-k}{k-|\mathcal{T}|}$ events $\{\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}\}$ with the same $\mathcal{T}$

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}) \\
&\le \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) + \sum_{\mathcal{T}\subset[k]}\binom{m-k}{k-|\mathcal{T}|} \\
&\quad \cdot k!\cdot q(\|\mathbf{w}\| + \epsilon, \epsilon)\cdot 2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}} \\
&\le \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) + k!\cdot q(\|\mathbf{w}\| + \epsilon, \epsilon) \\
&\quad \cdot \sum_{\mathcal{T}\subset[k]} 2^{(k-|\mathcal{T}|)\log m}\cdot 2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in[k]\setminus\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}} \\
&= \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) + k!\cdot q(\|\mathbf{w}\| + \epsilon, \epsilon) \\
&\quad \cdot \sum_{\mathcal{T}\subseteq[k]} 2^{|\mathcal{T}|\log m}\cdot 2^{-\frac{n}{2}\log\frac{\left(\sum_{j\in\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2}}
\end{aligned}$$

which tends to zero as $m \to \infty$, if

$$\limsup_{m\to\infty}\frac{\log m}{n_m} < \frac{1}{2|\mathcal{T}|}\log\frac{\left(\sum_{j\in\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2 - \delta_1\epsilon}{\sigma_z^2 + \epsilon^2\sigma_a^2} \qquad (32)$$

for all $\mathcal{T} \subseteq [k]$. Since $\epsilon > 0$ is arbitrarily chosen, the proof of Theorem 1 is complete.

Now, we prove Lemma 1. For simplicity, let $\theta \equiv \sigma_v^2$. Denote $S_n = \frac{1}{n}\sum_{i=1}^{n}(u_i - V_i)^2$. The moment generating function of S_n is

$$\mathsf{E}[e^{tS_n}] = \mathsf{E}[e^{\frac{t}{n}\sum_{i=1}^{n}(u_i - V_i)^2}] = \prod_{i=1}^{n}\mathsf{E}[e^{\frac{t}{n}(u_i - V_i)^2}]. \qquad (33)$$

Note that $(u_i - V_i)^2/\theta$ is a noncentral χ^2 random variable. Its moment generating function is given by [61] as $\mathsf{E}[e^{t(u_i - V_i)^2/\theta}] = \exp(\frac{tu_i^2/\theta}{1-2t})/(1-2t)^{\frac{1}{2}}$, for $t \le 1/2$. By changing variable $\theta t/n \to t$, we have

$$\mathsf{E}[e^{t(u_i - V_i)^2/n}] = \frac{e^{\frac{\frac{t}{n}u_i^2}{1-2\theta t/n}}}{(1 - 2\theta t/n)^{\frac{1}{2}}}.$$

Back to (33), we obtain

$$\begin{aligned}\mathsf{E}[e^{tS_n}] &= \prod_{i=1}^{n} \mathsf{E}[e^{\frac{t}{n}(u_i - V_i)^2}] \\ &= \prod_{i=1}^{n} \frac{e^{\frac{\frac{t}{n}u_i^2}{1-2\theta t/n}}}{(1-2\theta t/n)^{\frac{1}{2}}} \\ &= \frac{e^{\frac{\frac{t}{n}\sum_{i=1}^n u_i^2}{1-2\theta t/n}}}{(1-2\theta t/n)^{\frac{n}{2}}}.\end{aligned}$$

The Chernoff bound implies

$$\begin{aligned}&\mathsf{P}(S_n \le \gamma) \\ &\le \min_{s>0} e^{s\gamma}\mathsf{E}[e^{-sS_n}] \\ &= \min_{s>0} e^{s\gamma} \frac{e^{\frac{-\frac{s}{n}\sum_{i=1}^n u_i^2}{1+2\theta s/n}}}{(1+2\theta s/n)^{\frac{n}{2}}} \\ &= \min_{p<0} e^{-p\gamma} \frac{e^{\frac{\frac{p}{n}\sum_{i=1}^n u_i^2}{1-2\theta p/n}}}{(1-2\theta p/n)^{\frac{n}{2}}} \\ &= \exp\left\{\min_{p<0}\left\{\log e^{-p\gamma}\frac{e^{\frac{\frac{p}{n}\sum_{i=1}^n u_i^2}{1-2\theta p/n}}}{(1-2\theta p/n)^{\frac{n}{2}}}\right\}\right\} \\ &= \exp\left\{\min_{p<0}\left\{-p\gamma + \frac{\frac{p}{n}\sum_{i=1}^n u_i^2}{1-2\theta p/n} - \frac{n}{2}\log\left(1-2\theta p/n\right)\right\}\right\}.\end{aligned}$$

Define

$$\begin{aligned}f(p) &\triangleq -p\gamma + \frac{\frac{p}{n}\sum_{i=1}^n u_i^2}{1-2\theta p/n} - \frac{n}{2}\log\left(1-2\theta p/n\right) \\ g(\lambda) &\triangleq f(n\lambda) = -n\lambda\gamma + \frac{\lambda\sum_{i=1}^n u_i^2}{1-2\theta\lambda} - \frac{n}{2}\log\left(1-2\theta\lambda\right).\end{aligned}$$

Clearly, $\min_{p<0} f(p) = \min_{\lambda<0} g(\lambda)$. Denote

$$\alpha_s \triangleq \frac{1}{n}\sum_{i=1}^{n} u_i^2.$$

Then, let us focus on the minimization problem

$$\begin{aligned}&\min_{\lambda<0} g(\lambda) \\ &= \min_{\lambda<0}\left\{-n\lambda\gamma + \frac{n\lambda\alpha_s}{1-2\theta\lambda} - \frac{n}{2}\log\left(1-2\theta\lambda\right)\right\} \\ &= n\cdot\min_{\lambda<0}\left\{-\lambda\gamma + \frac{\lambda\alpha_s}{1-2\theta\lambda} - \frac{1}{2}\log\left(1-2\theta\lambda\right)\right\} \\ &= -n\cdot\underbrace{\max_{\lambda<0}\left\{\lambda\gamma - \frac{\lambda\alpha_s}{1-2\theta\lambda} + \frac{1}{2}\log\left(1-2\theta\lambda\right)\right\}}_{\triangleq\,\Lambda(\alpha_s,\theta,\gamma)}.\end{aligned}$$

It can be shown that the minimizing λ is

$$\lambda^* = \frac{2\gamma - \theta - \sqrt{\theta^2 + 4\alpha_s\gamma}}{4\theta\gamma} < 0$$

and hence

$$\begin{aligned}&\Lambda(\alpha_s,\theta,\gamma) \\ &= \lambda^*\gamma - \frac{\lambda^*\alpha_s}{1-2\lambda^*\theta} + \frac{1}{2}\log(1-2\lambda^*\theta) \\ &= \frac{\alpha_s+\gamma}{2\theta} - \frac{1}{2} - \frac{2\alpha_s\gamma}{\theta(\theta+\sqrt{\theta^2+4\alpha_s\gamma})} \\ &\quad + \frac{1}{2}\log\frac{\theta+\sqrt{\theta^2+4\alpha_s\gamma}}{2\gamma}.\end{aligned}$$

Next, for fixed α_s and γ

$$\begin{aligned}&\frac{\partial\Lambda(\alpha_s,\theta,\gamma)}{\partial\theta} \\ &= -\frac{\alpha_s+\gamma}{2\theta^2} \\ &\quad + \frac{2\alpha_s\gamma\left[\theta+\sqrt{\theta^2+4\alpha_s\gamma}+\theta\left(1+\frac{2\theta}{2\sqrt{\theta^2+4\alpha_s\gamma}}\right)\right]}{\theta^2(\theta+\sqrt{\theta^2+4\alpha_s\gamma})^2} \\ &\quad + \frac{1}{2(\theta+\sqrt{\theta^2+4\alpha_s\gamma})}\left(1+\frac{2\theta}{2\sqrt{\theta^2+4\alpha_s\gamma}}\right) \\ &= -\frac{\alpha_s+\gamma}{2\theta^2} + \frac{\sqrt{4\alpha_s\gamma+\theta^2}}{2\theta^2}.\end{aligned}$$

For $\theta > 0$, there is only one stationary point $\theta' = \alpha_s - \gamma$, which is a solution to $\frac{\partial\Lambda(\alpha_s,\theta,\gamma)}{\partial\theta} = 0$. Check the second derivative

$$\left.\frac{\partial^2\Lambda(\alpha_s,\theta,\gamma)}{\partial\theta^2}\right|_{\theta=\alpha_s-\gamma} = \frac{1}{2(\alpha_s+\gamma)(\alpha_s-\gamma)} > 0.$$

This confirms that $\theta' = \alpha_s - \gamma$ is the minimum point of $\Lambda(\alpha_s,\theta,\gamma)$, for $\theta > 0$. Hence, for fixed α_s and γ with $\gamma < \alpha_s$

$$\Lambda(\alpha_s,\theta,\gamma) \ge \Lambda(\alpha_s,\theta',\gamma) = \frac{1}{2}\log\frac{\alpha_s}{\gamma}.$$

As a result

$$\begin{aligned}&\mathsf{P}(S_n \le \gamma) \\ &\le \exp\left\{\min_{p<0}\left\{-p\gamma + \frac{\frac{p}{n}\sum_{i=1}^{n} u_i^2}{1-2\theta p/n} - \frac{n}{2}(1-2\theta p/n)\right\}\right\} \\ &= \exp\left\{\min_{\lambda<0} g(\lambda)\right\} \\ &= \exp\left\{-n\Lambda(\alpha_s,\theta,\gamma)\right\} \\ &\le \exp\left\{-n\Lambda(\alpha_s,\theta',\gamma)\right\} \\ &= \exp\left\{-\frac{n}{2}\log\left(\frac{\alpha_s}{\gamma}\right)\right\} \\ &\le \exp\left\{-\frac{n}{2}\log\left(\frac{\alpha-\beta}{\gamma}\right)\right\}.\end{aligned}$$

Hence, by changing the base of logarithm

$$\mathsf{P}\left(\frac{1}{n}\sum_{i=1}^{n}(u_i - V_i)^2 \le \gamma\right) \le 2^{-\frac{n}{2}\log\left(\frac{\alpha-\beta}{\gamma}\right)}.$$

Finally, we verify Lemma 2–1). For any $\zeta_1 > 0$, according to LLN

$$\lim_{m\to\infty}\mathsf{P}(|\hat{\rho} - \|\mathbf{w}\|| \le \zeta_1) = 1.$$

Note that $\mathbf{W}_0 \triangleq \frac{\hat{\rho}}{\|\mathbf{w}\|}\mathbf{w} \in \mathcal{B}_k(\hat{\rho})$. According to the definition of $\mathcal{Q}(\hat{\rho},\zeta)$, there must exist $\hat{\mathbf{W}} \in \mathcal{Q}(\hat{\rho},\zeta)$ such that $\|\hat{\mathbf{W}} - \mathbf{W}_0\| \le \frac{\zeta}{2}$. Fundamental geometry implies

$$\begin{aligned}\|\hat{\mathbf{W}} - \mathbf{w}\| &\le \|\hat{\mathbf{W}} - \mathbf{W}_0\| + \|\mathbf{W}_0 - \mathbf{w}\| \\ &\le \frac{\zeta}{2} + \big|\hat{\rho} - \|\mathbf{w}\|\big|.\end{aligned}$$

Hence

$$\lim_{m\to\infty}\mathsf{P}\left(\|\hat{\mathbf{W}} - \mathbf{w}\| \le \frac{\zeta}{2} + \zeta_1\right) = 1.$$

Choosing $\zeta_1 \in (0, \zeta/2)$ completes the proof.

Appendix II
Proof of Theorem 2

The main techniques for the proof of Theorem 2 include Fano's inequality and the properties of entropy. It mimics the proof of the converse for the channel coding theorem [49] with proper modifications.

For any $\mathcal{T} \subseteq [k]$, denote the tuple of random variables $(S_l : l \in \mathcal{T})$ by $S(\mathcal{T})$. From Fano's inequality [49], we have

$$\begin{aligned}H(S(\mathcal{T})|\mathbf{Y}, A) &\le H(S_1,\ldots,S_k|\mathbf{Y}, A) \\ &\le \log k! + H(\{S_1,\ldots,S_k\}|\mathbf{Y}, A) \\ &\le \log k! + \overline{P}_e^{(m)}\log\binom{m}{k} + 1 \qquad (34)\end{aligned}$$

where $\overline{P}_e^{(m)} \triangleq \mathsf{P}\{d^{(m)}(A\mathbf{X} + \mathbf{Z}) \ne \mathrm{supp}(\mathbf{X}(\mathbf{w},\mathbf{S}))\}$ for notation simplicity. On the other hand, by a basic permutation argument

$$\begin{aligned}H(S(\mathcal{T})|S(\mathcal{T}^c), A) &= \log\left(\prod_{q=0}^{|\mathcal{T}|-1}(m - (k - |\mathcal{T}|) - q)\right) \\ &= |\mathcal{T}|\log m - n\epsilon_{1,n} \qquad (35)\end{aligned}$$

where $\mathcal{T}^c \triangleq [k]\backslash\mathcal{T}$ and

$$\epsilon_{1,n} \triangleq \frac{1}{n}\log\left(m^{|\mathcal{T}|}/\prod_{q=0}^{|\mathcal{T}|-1}(m - (k - |\mathcal{T}|) - q)\right)$$

which tends to zero as $n \to \infty$. Hence, combining (34) and (35), we have

$$\begin{aligned}&|\mathcal{T}|\log m \\ &\quad = H(S(\mathcal{T})|S(\mathcal{T}^c), A) + n\epsilon_{1,n} \\ &\quad = I(S(\mathcal{T});\mathbf{Y}|S(\mathcal{T}^c), A) + H(S(\mathcal{T})|\mathbf{Y}, S(\mathcal{T}^c), A) + n\epsilon_{1,n} \\ &\quad \le I(S(\mathcal{T});\mathbf{Y}|S(\mathcal{T}^c), A) + H(S(\mathcal{T})|\mathbf{Y}, A) + n\epsilon_{1,n} \qquad (36) \\ &\quad \le I(S(\mathcal{T});\mathbf{Y}|S(\mathcal{T}^c), A) + \log k! \\ &\qquad + \overline{P}_e^{(m)}\log\binom{m}{k} + 1 + n\epsilon_{1,n} \\ &\quad = \sum_{i=1}^{n} I(Y_i; S(\mathcal{T})|Y_1^{i-1}, S(\mathcal{T}^c), A) + \log k! \\ &\qquad + \overline{P}_e^{(m)}\log\binom{m}{k} + 1 + n\epsilon_{1,n} \qquad (37) \\ &\quad \le \sum_{i=1}^{n}\left(h(Y_i|S(\mathcal{T}^c)) - h(Y_i|S_1,\ldots,S_k, A)\right) + \log k! \\ &\qquad + \overline{P}_e^{(m)}\log\binom{m}{k} + 1 + n\epsilon_{1,n} \\ &\quad \le \sum_{i=1}^{n}(h(Y_i|S(\mathcal{T}^c)) - h(Z_i)) + \log k! \\ &\qquad + \overline{P}_e^{(m)}\log\binom{m}{k} + 1 + n\epsilon_{1,n} \qquad (38)\end{aligned}$$

where (36) follows the fact that conditioning reduces entropy, (37) follows the chain rule of mutual information [49], and (38) follows since we condition on the measurement matrix A and Z_i is independent of $(S_1,\ldots,S_k)$ and A.

Consider

$$\begin{aligned}&h(Y_i|S(\mathcal{T}^c)) \\ &\quad = h\left(\sum_{j=1}^{k} w_j a_{i,S_j} + Z_i\middle|S(\mathcal{T}^c)\right) \\ &\quad = h\left(\sum_{j\in\mathcal{T}} w_j a_{i,S_j} + Z_i\middle|S(\mathcal{T}^c)\right) \\ &\quad \le h\left(\sum_{j\in\mathcal{T}} w_j a_{i,S_j} + Z_i\right) \\ &\quad \le \frac{1}{2}\log\left(2\pi e\cdot\mathrm{Var}\left(\sum_{j\in\mathcal{T}} w_j a_{i,S_j} + Z_i\right)\right) \qquad (39)\end{aligned}$$

where the last inequality follows since the Gaussian random variable maximizes the differential entropy given a variance constraint. To further upperbound (39), note that

$$\mathsf{E}\left(\sum_{j\in\mathcal{T}} w_j a_{i,S_j} + Z_i\middle|S_1,\ldots,S_k\right) = 0 \qquad (40)$$

and

$$\mathrm{Var}\left(\sum_{j\in\mathcal{T}} w_j a_{i,S_j} + Z_i\middle|S_1,\ldots,S_k\right) = \sigma_a^2\sum_{j\in\mathcal{T}} w_j^2 + \sigma_z^2.$$

According to the law of total variance

$$\mathrm{Var}\left(\sum_{j\in\mathcal{T}} w_j a_{i,S_j} + Z_i\right) = \sigma_a^2\sum_{j\in\mathcal{T}} w_j^2 + \sigma_z^2.$$

Returning to (38), we have

$$\begin{aligned}|\mathcal{T}|\log m &\le \sum_{i=1}^{n}\frac{1}{2}\log\left[2\pi e\left(\sigma_a^2\sum_{j\in\mathcal{T}}w_j^2+\sigma_z^2\right)\right]-\frac{n}{2}\log(2\pi e\sigma_z^2)\\ &\quad+\log k!+\overline{P}_e^{(m)}\log\binom{m}{k}+1+n\epsilon_{1,n}\\ &=\frac{n}{2}\log\left(\frac{\sigma_a^2}{\sigma_z^2}\sum_{j\in\mathcal{T}}w_j^2+1\right)+\log k!+\overline{P}_e^{(m)}\log\binom{m}{k}+1\\ &\quad+n\epsilon_{1,n}.\end{aligned}\tag{41}$$

Therefore

$$\begin{aligned}&\limsup_{m\to\infty}\frac{\log m}{n_m}-\frac{\log k!+\overline{P}_e^{(m)}\log\binom{m}{k}+1+n_m\epsilon_{1,n_m}}{|\mathcal{T}|n_m}\\ &\quad\le\frac{1}{2|\mathcal{T}|}\log\left(1+\frac{\sigma_a^2}{\sigma_z^2}\sum_{j\in\mathcal{T}}w_j^2\right)\end{aligned}\tag{42}$$

for all $\mathcal{T}\subseteq[k]$. Due to the fact that $\log\binom{m}{k}\le k\log m$, we have

$$\begin{aligned}&\limsup_{m\to\infty}\frac{(1-k\overline{P}_e^{(m)}/|\mathcal{T}|)\log m}{n_m}-\frac{\log k!+n_m\epsilon_{1,n_m}+1}{|\mathcal{T}|n_m}\\ &\quad\le\frac{1}{2|\mathcal{T}|}\log\left(1+\frac{\sigma_a^2}{\sigma_z^2}\sum_{j\in\mathcal{T}}w_j^2\right)\end{aligned}\tag{43}$$

for all $\mathcal{T}\subseteq[k]$. Since $\lim_{m\to\infty}\overline{P}_e^{(m)}=0$, we reach the conclusion

$$\limsup_{m\to\infty}\frac{\log m}{n_m}\le\frac{1}{2|\mathcal{T}|}\log\left(1+\frac{\sigma_a^2}{\sigma_z^2}\sum_{j\in\mathcal{T}}w_j^2\right)$$

for all $\mathcal{T}\subseteq[k]$, which completes the proof of Theorem 2.

APPENDIX III
PROOF OF THEOREM 3

We show that

$$\lim_{m\to\infty}\mathsf{P}\{d^{(m)}(A\mathbf{X}(\mathbf{w}^{(m)},\mathbf{S})+\mathbf{Z})\neq\mathrm{supp}(\mathbf{X}(\mathbf{w}^{(m)},\mathbf{S}))\}=0$$

provided that the condition

$$\limsup_{m\to\infty}\frac{1}{n_m}\max_{j\in[k_m]}\left[\frac{6k_m\log k_m+2j\log\frac{me}{j}}{\log\left(\frac{jw_{\min}^2\sigma_a^2}{\sigma_z^2}+1\right)}\right]<1\tag{44}$$

is satisfied. Note that (44) implies that $n=\max[\Omega(k\log k),\Omega(\frac{k}{\log k}\log\frac{m}{k})]$, which in turn implies that $k=o(n)$.

We follow the proof of Theorem 1 in Appendix I. Recall that in Case 2 of the proof of Theorem 1, we first proposed the support recovery rule (26). Then, we formed estimates of the nonzero values, and used them to test all possible sets of k indices. The key step was to analyze two types of errors. On the one hand, the true support should satisfy the reconstruction rule (26) with high probability. On the other hand, the probability that at least one incorrect support possibility satisfies this rule was controlled to diminish as the problem size increases.

By mainly replicating the steps in Appendix I with necessary accommodations to the new setting with growing number of nonzero entries, we present the proof of Theorem 3 as follows.

1) We first modify the support recovery rule by replacing (26) with

$$\frac{1}{n}\left\|\mathbf{Y}-\sum_{j=1}^{k}\hat{W}_j\mathbf{A}_{\hat{s}_j}\right\|^2\le(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2.\tag{45}$$

2) The cardinality $q(r,\zeta)$ of a minimal $\mathcal{Q}(r,\zeta)$ can be upper-bounded by

$$q(r,\zeta)\le\left(\frac{\eta_1 kr}{\zeta}\right)^k$$

for some $\eta_1>0$. This can be easily shown by first partitioning the k-dimensional hypercube of side $2r$ into identical elementary hypercubes with side not exceeding $\frac{\zeta}{4k}$ and then, for each elementary hypercube that intersects the hypersphere, picking an arbitrary point on the hypersphere within that elementary hypercube. The resulting set of points provides the upper bound above for $q(r,\zeta)$.

3) Define $\sigma_{\max}^2$ and $\sigma_{\min}^2$ to be the largest and smallest eigenvalues of the matrix

$$\frac{1}{n\sigma_a^2}[\mathbf{A}_1,\dots,\mathbf{A}_k,\frac{\sigma_a}{\sigma_z}\mathbf{Z}]^\mathsf{T}[\mathbf{A}_1,\dots,\mathbf{A}_k,\frac{\sigma_a}{\sigma_z}\mathbf{Z}]$$

respectively. We replace the definition of $\mathcal{E}_{\mathrm{aux}}$ by

$$\begin{aligned}\mathcal{E}_{\mathrm{aux}}\triangleq&\{\hat{\rho}-\|\mathbf{w}\|\in(-\epsilon,\epsilon)\}\\ &\cap\left\{\sigma_{\max}^2\in(1-\epsilon,1+\epsilon)\right\}\\ &\cap\left\{\sigma_{\min}^2\in(1-\epsilon,1+\epsilon)\right\}.\end{aligned}$$

Consider the asymptotic behaviors of the events. First, note that

$$\begin{aligned}&\sqrt{\frac{1}{n\sigma_a^2}\|\mathbf{Y}\|^2}\\ &\quad=\sqrt{\frac{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}{n\sigma_a^2}}\sqrt{\left\|\frac{\mathbf{Y}}{\sqrt{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}}\right\|^2}\end{aligned}\tag{46}$$

where $\sqrt{\left\|\frac{\mathbf{Y}}{\sqrt{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}}\right\|^2}$ is χ-distributed with mean $\sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}$ and variance $\left(n-\frac{2\Gamma^2((n+1)/2)}{\Gamma^2(n/2)}\right)$. Then, $\sqrt{\frac{1}{n\sigma_a^2}\|\mathbf{Y}\|^2}$ has mean $\sqrt{\frac{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}{n\sigma_a^2}}\sqrt{2}\frac{\Gamma((n+1)/2)}{\Gamma(n/2)}$ and variance $\frac{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}{n\sigma_a^2}\left(n-\frac{2\Gamma^2((n+1)/2)}{\Gamma^2(n/2)}\right)$.

It has been shown [62] that

$$\lim_{x\to\infty}\frac{x\Gamma(x)}{\sqrt{x+1/4}\Gamma(x+1/2)}=1.$$

Then, as $n\to\infty$, $\sqrt{\frac{1}{n\sigma_a^2}\|\mathbf{Y}\|^2}$ has asymptotic mean $\sqrt{\frac{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}{\sigma_a^2}}$ and variance $\frac{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}{2n\sigma_a^2}$. Since

$k = o(n)$, we have $\frac{\|\mathbf{w}\|^2\sigma_a^2+\sigma_z^2}{2n\sigma_a^2} \to 0$. Hence, $\lim_{m\to\infty} \mathsf{P}\{\hat{\rho} - \|\mathbf{w}\| \in (-\epsilon,\epsilon)\} = 1$.
Second, $\sigma_{\max}^2$ and $\sigma_{\min}^2$ are shown [63] to almost surely converge to $(1+q)^2$ and $(1-q)^2$, respectively, where $q \triangleq \lim_{m\to\infty}\sqrt{(k+1)/n} = 0$. Thus, $\lim_{m\to\infty}\mathsf{P}(\mathcal{E}_{\text{aux}}^c) = 0$.

4) Next, we analyze the probability that the true support satisfies the recovery rule. Note that

$$\begin{aligned}\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^k \hat{W}_j\mathbf{A}_j\right\|^2 &= \frac{1}{n}\left\|\sum_{j=1}^k w_j\mathbf{A}_j + \mathbf{Z} - \sum_{j=1}^k \hat{W}_j\mathbf{A}_j\right\|^2\\ &= \frac{1}{n}\left\|\left[\mathbf{A}_1,\ldots,\mathbf{A}_k,\frac{\sigma_a}{\sigma_z}\mathbf{Z}\right]\begin{bmatrix}\mathbf{w}-\hat{\mathbf{W}}\\ \frac{\sigma_z}{\sigma_a}\end{bmatrix}\right\|^2\\ &\le \sigma_{\max}^2\sigma_a^2\left\|\begin{bmatrix}\mathbf{w}-\hat{\mathbf{W}}\\ \frac{\sigma_z}{\sigma_a}\end{bmatrix}\right\|^2\\ &= \sigma_{\max}^2\sigma_a^2\|\mathbf{w}-\hat{\mathbf{W}}\|^2 + \sigma_{\max}^2\sigma_z^2. \end{aligned} \tag{47}$$

By using the fact that $\sigma_{\max}^2 \to 1$ almost surely as $n\to\infty$ and Lemma 2–1), we have $\lim_{m\to\infty}\mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) = 0$.

5) Now, suppose we have proceeded to a step similar to (30) [that is, to be exact, equipped with the modified rule (45) and a proper $\mathcal{E}_{\text{cond}}$]. Define the auxiliary vector $\mathbf{w}' \in \mathbb{R}^{k+1}$ as

$$w_j' = \begin{cases} w_j, & \text{if } j \in [k]\backslash\mathcal{T}\\ w_j - \hat{W}_i, & \text{if } j = s_i' \in \mathcal{T}\\ \frac{\sigma_z}{\sigma_a}, & \text{if } j = k+1. \end{cases} \tag{48}$$

Then

$$\begin{aligned}\frac{1}{n}\left\|\sum_{j=1}^k w_j\mathbf{A}_j - \sum_{s_j'\in\mathcal{T}}\hat{W}_j\mathbf{A}_{s_j'} + \mathbf{Z}\right\|^2 &= \frac{1}{n}\left\|\left[\mathbf{A}_1,\ldots,\mathbf{A}_k,\frac{\sigma_a}{\sigma_z}\mathbf{Z}\right]\mathbf{w}'\right\|^2\\ &\ge (1-\epsilon)\|\mathbf{w}'\|^2\sigma_a^2\\ &\ge (1-\epsilon)\left(\left(\sum_{j\in[k]\backslash\mathcal{T}} w_j^2\right)\sigma_a^2 + \sigma_z^2\right).\end{aligned}$$

From Lemma 1, it follows that (for sufficiently small ϵ)

$$\begin{aligned}&\mathsf{P}\left(\frac{1}{n}\left\|\mathbf{Y} - \sum_{j=1}^k \hat{W}_j\mathbf{A}_{s_j'}\right\|^2 \le (1+\epsilon)\sigma_z^2 + 2\epsilon^2\sigma_a^2 \middle| \mathcal{E}_{\text{cond}}\right)\\ &\le 2^{-\frac{n}{2}\log\frac{(1-\epsilon)\left(\left(\sum_{j\in[k]\backslash\mathcal{T}} w_j^2\right)\sigma_a^2+\sigma_z^2\right)}{(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2}}.\end{aligned}$$

6) Note that, from [33]

$$\binom{m}{k} \le \left(\frac{me}{k}\right)^k.$$

Together with the modifications above, we follow the proof steps of Theorem 1 to reach

$$\begin{aligned}\mathsf{P}(\mathcal{E}) &\le \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) + k!\cdot q(\|\mathbf{w}\|+\epsilon,\epsilon)\\ &\quad\cdot\sum_{\mathcal{T}\subseteq[k]}\left(\frac{me}{|\mathcal{T}|}\right)^{|\mathcal{T}|}\cdot 2^{-\frac{n}{2}\log\frac{(1-\epsilon)\left(\left(\sum_{j\in\mathcal{T}} w_j^2\right)\sigma_a^2+\sigma_z^2\right)}{(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2}}\\ &\le \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) + k!\cdot q(\|\mathbf{w}\|+\epsilon,\epsilon)\\ &\quad\cdot\sum_{\mathcal{T}\subseteq[k]}\left(\frac{me}{|\mathcal{T}|}\right)^{|\mathcal{T}|}\cdot 2^{-\frac{n}{2}\log\frac{(1-\epsilon)\left(|\mathcal{T}|w_{\min}^2\sigma_a^2+\sigma_z^2\right)}{(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2}}\\ &\le \mathsf{P}(\mathcal{E}_{\text{aux}}^c) + \mathsf{P}(\mathcal{E}_{1,2,\ldots,k}^c) + k!\cdot q(\|\mathbf{w}\|+\epsilon,\epsilon)\\ &\quad\cdot 2^k\cdot\max_{j\in[k]}\left[\left(\frac{me}{j}\right)^j\cdot 2^{-\frac{n}{2}\log\frac{(1-\epsilon)\left(jw_{\min}^2\sigma_a^2+\sigma_z^2\right)}{(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2}}\right].\end{aligned} \tag{49}$$

Note that

$$\begin{aligned}&\log\Bigg(k!\cdot q(\|\mathbf{w}\|+\epsilon,\epsilon)\cdot 2^k\\ &\qquad\cdot\max_{j\in[k]}\left[\left(\frac{me}{j}\right)^j\cdot 2^{-\frac{n}{2}\log\frac{(1-\epsilon)\left(jw_{\min}^2\sigma_a^2+\sigma_z^2\right)}{(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2}}\right]\Bigg)\\ &\le k\log k + k\log(\eta_1 k^2 w_{\max}/\epsilon) + k\\ &\quad+\max_{j\in[k]}\left[j\log\frac{me}{j} - \frac{n}{2}\log\frac{(1-\epsilon)\left(jw_{\min}^2\sigma_a^2+\sigma_z^2\right)}{(1+\epsilon)\sigma_z^2+2\epsilon^2\sigma_a^2}\right].\end{aligned} \tag{50}$$

It can be readily seen that from condition (44), the upper bound in (50) becomes negative and thus $\mathsf{P}(\mathcal{E}) \to 0$ as $m\to\infty$.

Appendix IV
Proof of Theorem 4

The proof of Theorem 2 can be adapted to establish Theorem 4; see [64] for detail. Since we need a bound corresponding to only the sum rate, however, we use the following simple argument.

Suppose that each user $i \in [k_m]$ uses a codebook of size m/k_m given by $\{\mathbf{A}_j : (i-1)m/k_m < j \le im/k_m\}$. (Assume without loss of generality that m/k_m is an integer.) This is equivalent to assuming that each nonzero entry appears in its predefined subset of $[m]$, i.e.,

$$S_i \in \{(i-1)m/k_m+1,\ldots,im/k_m\}, i\in[k_m]. \tag{51}$$

Under this specific setup, if exact support recovery is asymptotically successful, it follows that every user can operate at the rate $R^{(i)} = \log(m/k_m)/n$. Immediately, (5) implies the necessary condition

$$\limsup_{m\to\infty}\frac{\log(m/k_m)}{n_m c(\mathbf{w}^{(m)})} \le 1$$

which leads to

$$\limsup_{m\to\infty} \frac{2k_m\left(\log m - \log k_m\right)}{n_m \log\left(\frac{k_m w_{\max}^2 \sigma_a^2}{\sigma_z^2}+1\right)} \leq 1.$$

We conclude the proof by noting that the special setup in (51) is equivalent to the original setup in Section II in terms of the average probability of error in support recovery due to the symmetry in the random matrix A.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful comments. Y. Jin would like to thank L. Yu for insightful discussions on perspectives of MAC communication.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[2] E. J. Candes, "Compressive sampling," in *Proc. Int. Congr. Mathematicians*, 2006, pp. 1433–1452.
[3] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
[4] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *J. Electroencephalog. Clin. Neurophysiol.*, vol. 95, pp. 231–251, 1995.
[5] B. D. Jeffs, "Sparse inverse solution methods for signal and image processing applications," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 1998, pp. 1885–1888.
[6] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[7] S. D. Cabrera and T. W. Parks, "Extrapolation and spectral estimation with iterative weighted norm modification," *IEEE Trans. Signal Process.*, vol. 39, no. 4, pp. 842–851, Apr. 1991.
[8] Y. Jin and B. D. Rao, "Algorithms for robust linear regression by exploiting the connection to sparse signal recovery," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 3830–3833.
[9] W. C. Chu, *Speech Coding Algorithms*. New York: Wiley-Interscience, 2003.
[10] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 374–377, Mar. 2002.
[11] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing," in *Proc. Conf. Inf. Sci. Syst.*, 2008, pp. 5–10.
[12] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.
[13] B. D. Rao and B. Song, "Adaptive filtering algorithms for promoting sparsity," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2003, pp. 361–364.
[14] D. Guo, "Neighbor discovery in ad hoc networks as a compressed sensing problem," in *Proc. Inf. Theory Appl. Workshop*, 2009.
[15] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
[16] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals Syst. Comput.*, 1993, pp. 40–44.
[17] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
[18] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
[19] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
[20] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.
[21] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May. 2009.
[22] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
[23] D. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[24] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[25] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[26] J. A. Tropp, "Greedy is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[27] 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.
[28] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," 2006, to be published.
[29] 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.
[30] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Process. Mag.*, vol. 18, no. 6, pp. 14–30, Nov. 2001.
[31] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, no. 3, pp. 947–966, 2008.
[32] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2007, pp. 1357–1360.
[33] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *Proc. Int. Symp. Inf. Theory*, Jun. 2007, pp. 961–965.
[34] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, Dec. 2009.
[35] K. R. Rad, "Sharp upper bound on error probability of exact sparsity recovery," in *Proc. Conf. Inf. Sci. Syst.*, 2009, pp. 14–17.
[36] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Resolution limits of sparse coding in high dimensions," in *Proc. Neural Inf. Process. Syst.*, 2008, pp. 449–456.
[37] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.
[38] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement," in *Proc. Int. Symp. Inf. Theory*, 2008, pp. 2197–2201.
[39] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, Jun. 2008.
[40] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.
[41] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Proc. 44th Allerton Conf. Commun. Control Comput.*, 2006.
[42] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On the rate-distortion performance of compressed sensing," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, pp. 885–888.
[43] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4299–4308, Sep. 2009.
[44] H. V. Pham, W. Dai, and O. Milenkovic, "Sublinear compressive sensing reconstruction via belief propagation decoding," in *Proc. Int. Symp. Inf. Theory*, 2009, pp. 674–678.
[45] F. Zhang and H. D. Pfister, "Compressed sensing and linear codes over real numbers," in *Proc. Inf. Theory Appl. Workshop*, 2008, pp. 558–561.
[46] Y. Jin and B. D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 3921–3924.

[47] Y. Jin and B. D. Rao, “Performance limits of matching pursuit algorithms,” in *Proc. Int. Symp. Inf. Theory*, 2008, pp. 2444–2448.
[48] J. Tropp, “Just relax: Convex programming methods for identifying sparse signal in noise,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.
[50] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[51] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?,” *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2000.
[52] A. Lapidoth, “Nearest neighbor decoding for additive non-Gaussian noise channels,” *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.
[53] S. I. Resnick, *A Probability Path*. Boston, MA: Birkhauser, 1999.
[54] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
[55] D. P. Wipf and B. D. Rao, “An empirical Bayesian strategy for solving the simultaneous sparse approximation problem,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.
[56] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.
[57] R. Zdunek and A. Cichocki, “Improved M-FOCUSS algorithm with overlapping blocks for locally smooth sparse signals,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4752–4761, Oct. 2008.
[58] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
[59] M. Akçakaya and V. Tarokh, “A frame construction and a universal distortion bound for sparse representations,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2443–2450, Jun. 2008.
[60] D. Baron, S. Sarvotham, and R. G. Baraniuk, “Bayesian compressive sensing via belief propagation,” *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.
[61] H. O. Lancaster, *The χ^2 Distribution*. New York: Wiley, 1969.
[62] C.-P. Chen and F. Qi, “Completely monotonic function associated with the Gamma functions and proof of Wallis’ inequality,” *Tamkang J. Math.*, vol. 36, pp. 303–307, 2005.
[63] J. W. Silverstein, “The smallest eigenvalue of a large dimensional Wishart matrix,” *Ann. Probab.*, vol. 13, pp. 1364–1368, 1985.
[64] Y. Jin, “Algorithm development for sparse signal recovery and performance limits using multiple-user information theory,” Ph.D. dissertation, Dept. Electr. Comput. Eng., Univ. California San Diego, La Jolla, CA, 2011.

Yuzhe Jin (S’07) received the B.E. degree in computer science and technology from Tsinghua University, Beijing, China, in 2005 and the Ph.D. degree in electrical and computer engineering from University of California, San Diego, La Jolla, in 2011.

His research interests are in statistical signal processing, sparse signal recovery, compressed sensing, information theory, natural language processing, and machine learning.

Young-Han Kim (S’99–M’06) received the B.S. degree with honors in electrical engineering from Seoul National University, Seoul, Korea, in 1996 and the M.S. degrees in electrical engineering and statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2001, 2006, and 2006, respectively.

In July 2006, he joined the University of California San Diego, La Jolla, where he is currently an Assistant Professor of Electrical and Computer Engineering. His research interests are in statistical signal processing and information theory, with applications in communication, control, computation, networking, data compression, and learning.

Dr. Kim is a recipient of the 2008 NSF Faculty Early Career Development (CAREER) Award and the 2009 U.S.–Israel Binational Science Foundation Bergmann Memorial Award.

Bhaskar D. Rao (F’00) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 1979 and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1981 and 1983, respectively.

Since 1983, he has been with the University of California San Diego, La Jolla, where he is currently a Professor with the Electrical and Computer Engineering Department and holder of the Ericsson endowed chair in wireless access networks. He is the Director of the Center for Wireless Communications. His interests are in the areas of digital signal processing, estimation theory, and optimization theory, with applications to digital communications, speech signal processing, and human–computer interactions.

Dr. Rao was elected to an IEEE Fellow for his contributions in high-resolution spectral estimation. His research group has received several paper awards. His paper received the best paper award at the 2000 Speech Coding Workshop and his students have received student paper awards at the 2005 and 2006 International Conference on Acoustics, Speech and Signal Processing as well as the best student paper award at the 2006 Neural Information Processing Systems (NIPS). A paper he coauthored with B. Song and R. Cruz received the 2008 Stephen O. Rice Prize Paper Award in the Field of Communications Systems. He has been a member of the Statistical Signal and Array Processing technical committee, the Signal Processing Theory and Methods Technical Committee, the Communications Technical Committee of the IEEE Signal Processing Society. He has also served on the editorial board of the *EURASIP Signal Processing Journal*.