Support Recovery of Sparse Signals in the Presence of Multiple Measurement Vectors

Yuzhe Jin, Student Member, IEEE, and Bhaskar D. Rao, Fellow, IEEE

Abstract

This paper studies the problem of support recovery of sparse signals based on multiple measurement vectors (MMV). The MMV support recovery problem is connected to the problem of decoding messages in a Single-Input Multiple-Output (SIMO) multiple access channel (MAC), thereby enabling an information theoretic framework for analyzing performance limits in recovering the support of sparse signals. Sharp sufficient and necessary conditions for successful support recovery are derived in terms of the number of measurements per measurement vector, the number of nonzero rows, the measurement noise level, and especially the number of measurement vectors. Through the interpretations of the results, in particular the connection to the multiple output communication system, the benefit of having MMV for sparse signal recovery is illustrated providing a theoretical foundation to the performance improvement enabled by MMV as observed in many existing simulation results. In particular, it is shown that the structure (rank) of the matrix formed by the nonzero entries plays an important role on the performance limits of support recovery.

I. INTRODUCTION

Suppose the signal of interest is $X \in \mathbb{R}^{m \times l}$, and $X$ is said to be sparse when only a few of its rows contain nonzero elements whereas the rest consist of zero elements. One wishes to estimate $X$ via the linear measurements $Y = AX + Z$, where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix and $Z \in \mathbb{R}^{n \times l}$ is the measurement noise. The goal is to estimate $X$ from as few measurements as possible. Specifically, when $l = 1$, this problem is usually termed as sparse signal recovery with a single measurement vector (SMV); when $l > 1$, it is referred to as sparse signal recovery with multiple measurement vectors (MMV) \cite{1}, \cite{2}. This problem has received much attention in many disciplines motivated by a broad array of applications.

The material in this paper was presented in part at the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, USA, 2010.
such as compressed sensing [3], [4], biomagnetic inverse problems [5], [6], image processing [7], [8], robust face recognition [9], bandlimited extrapolation and spectral estimation [10], robust regression and outlier detection [11], speech processing [12], channel estimation [13], [14], echo cancellation [15], [16], body area networks [17], and wireless communication [13], [18].

A. Background on the SMV Problem

For the problem of sparse signal recovery with SMV, computationally efficient algorithms have been proposed to find or approximate the sparse solution $X \in \mathbb{R}^m$ in various settings. A partial list includes matching pursuit [19], orthogonal matching pursuit (OMP) [20], Lasso [21], basis pursuit [22], FOCUSS [5], iteratively reweighted $\ell_1$ minimization [23], iteratively reweighted $\ell_2$ minimization [24], sparse Bayesian learning (SBL) [25], [26], finite rate of innovation [27], CoSaMP [28], and subspace pursuit [29]. Analysis has been developed to shed light on the performances of these practical algorithms. For example, Donoho [3], Donoho, Elad, and Temlyakov [30], Candès and Tao [31], and Candès, Romberg, and Tao [32] presented sufficient conditions for $\ell_1$-norm minimization algorithms, including basis pursuit and its variant in the noisy setting, to successfully recover the sparse signals with respect to different performance metrics. Tropp [33], Tropp and Gilbert [34], and Donoho, Tsaig, Drori, and Starck [35] studied the performances of greedy sequential selection methods such as matching pursuit and its variants. Wainwright [36] and Zhao and Yu [37] provided sufficient and necessary conditions for Lasso to recover the support of the sparse signal, i.e., the set of indices of the nonzero entries. On the other hand, from an information theoretic perspective, a series of papers, for instance, Wainwright [38], Fletcher, Rangan, and Goyal [39], Wang, Wainwright, and Ramchandran [40], Akçakaya and Tarokh [41], Jin, Kim, and Rao [42], provided sufficient and necessary conditions to characterize the performance limits of optimal algorithms for support recovery, regardless of computational complexity.

B. Background on the MMV Problem

As a fast emerging trend, the capability of collecting multiple measurements with an array of sensors in an increasing number of applications, such as magnetoencephalography (MEG) and electroencephalography (EEG) [2], [43], blind source separation [44], multivariate regression [45], and source localization [46], gives rise to the problem of sparse signal recovery with multiple measurement vectors. Practical algorithms have been developed to address the new challenges in this scenario. One class of algorithms for solving the MMV problem can be viewed as straightforward extensions based on their counterparts in
the SMV problem. To sample a few, M-OMP [13], [47], M-FOCUSS [13], \( \ell_1/\ell_2 \) minimization method [48], multivariate group Lasso [45], and M-SBL [49] can be all viewed as examples of this kind. Another class of algorithms additionally make explicit effort to exploit the structure underlying the sparse signal \( X \), such as the temporal correlation or the autoregressive nature across the columns of \( X \) which would be otherwise unavailable when \( l = 1 \), to aim for better performance of sparse signal recovery. For instance, the improved M-FOCUSS algorithms [2] and the auto-regressive sparse Bayesian learning (AR-SBL) [50] both have the capability of explicitly taking advantage of the structural properties of \( X \) to improve the recovery performance. Along side the algorithmic advancement, a series of work have been focusing on the theoretical analysis to support the effectiveness of existing algorithms for the MMV problem. We briefly divide these results into two categories. The first category of theoretic analysis aims at specific practical algorithms for sparse signal recovery with MMV. For example, Chen and Huo [51] discovered the sufficient conditions for \( \ell_1/\ell_p \) norm minimization method and orthogonal matching pursuit to exactly recover every sparse signal within certain sparsity level in the noiseless setting. Eldar and Rauhut [52] also analyzed the performance of sparse recovery using the \( \ell_1/\ell_2 \) norm minimization method in the noiseless setting, but the sparse signal was assumed to be randomly distributed according to certain probability distribution and the performance was averaged over all possible realizations of the sparse signal. Obozinski, Wainwright, and Jordan [45] provided sufficient and necessary conditions for multivariate group Lasso to successfully recover the support of the sparse signal in the presence of measurement noise. The second category of theoretic analysis are of an information theoretic nature, and explore the performance limits that any algorithm, regardless of computational complexity could possibly achieve. In this regard, Tang and Nehorai [53] employed a hypothesis testing framework with the likelihood ratio test as the optimal decision rule to study how fast the error probability decays. Sufficient and necessary conditions are further identified in order to guarantee successful support recovery in the asymptotic sense.

C. Focus and Contributions of This Paper

We develop sharp asymptotic performance limits among the signal dimension \( m \), the number of nonzero rows \( k \), the number of measurements per measurement vector \( n \), and the number of measurement vectors

1This method is sometimes referred to as \( \ell_2/\ell_1 \) minimization, due to the naming convention in a specific paper. In this paper, we use \( \ell_1/\ell_p \) to indicate a cost of a matrix \( B \) which is defined as \( \sum_i |(\sum_j |b_{i,j}|^p)^{1/p}| \).

2We refer to the support of a matrix \( X \) as the set of indices corresponding to the nonzero rows of \( X \). It will be formally defined in Section [11].
for reliable support recovery in the noisy setting. We show that \( n = (\log m)/c(X) \) is sufficient and necessary. We give a complete characterization of \( c(X) \) that depends on the elements of the nonzero rows of \( X \). Together with interpretations, we demonstrate the potential performance improvement enabled by having MMV, and hence bolster its usage in practical applications. Our main results are inspired by the analogy to communication over a Single-Input Multiple-Output (SIMO) multiple access channel (MAC). 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 \( X \). Then, the noise corrupted linear combinations of these codewords are observed by multiple receivers, which correspond to the multiple measurement vectors. The problem can be viewed as \( k \) single-antenna users communicating over a non-frequency selective channel with a base station equipped with \( l \) receive antennas. Thus, the problem of support recovery can be interpreted as multiple receivers jointly decoding messages sent by multiple senders, i.e., a SIMO MAC channel. With appropriate modifications, the techniques for deriving multiple-user channel capacity can be leveraged to provide performance limits for support recovery.

In the literatures on sparse signal recovery with SMV, the analogy between the problems of sparse signal recovery and channel coding has been observed from various perspectives in previous work [54], [55, Section IV-D], [40, Section II-A], [41, Section III-A], [35, Section 11.2]. However, their extensions to the MMV problem are unavailable to the authors’ knowledge. Moreover, our approach differs from existing works and would be different form their possible extensions to the MMV scenario, if any. We customize tools from multiple-user information theory to address the support recovery problem and we obtain sharp performance limits in the form of tight sufficient and necessary conditions.

D. Organization of the Paper

In Section II, we formally define the problem of support recovery of sparse signals in the presence of MMV. To motivate the main results of the paper and their proof techniques, in Section III we discuss the similarities and differences between the support recovery problem and the multiple access communication problem. The main results of the paper are presented in Section IV along with the interpretations. The proofs of the main theorems are presented in Appendices A and B. Relations to existing work are discussed in Section V. Section VI concludes the paper with further discussions.
E. Notations

Throughout this paper, a set is a collection of unique objects. Let $\mathbb{R}^m$ denote the $m$-dimensional real Euclidean space. Let $\mathbf{1}$ denote a column vector whose elements are all 1’s, and its length can be determined in the context. 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 $S$, $\|x\|_2$ denotes the $\ell_2$-norm of a vector $x$, and $\|A\|_F$ denotes the Frobenius norm of a matrix $A$. For a matrix $A$, $A_i$ denotes its $i$th column, $A_i$ denotes its $i$th row, and $A_T$ denotes the submatrix formed by the rows of $A$ indexed by the set $T$.

II. Problem Formulation

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

$$X_{s,i} = \begin{cases} w_{j,i} & \text{if } s = S_j, \\ 0 & \text{if } s \notin \{S_1, ..., S_k\}. \end{cases}$$  \hspace{1cm} (1)

The support of $X$, denoted by $\text{supp}(X)$, is the set of indices corresponding to the nonzero rows of $X$, i.e., $\text{supp}(X) = \{S_1, ..., S_k\}$. According to the signal model (1), $|\text{supp}(X)| = k$. Throughout this paper, we assume $k$ is known.

We measure $X$ through the linear operation

$$Y = AX + Z$$  \hspace{1cm} (2)

where $A \in \mathbb{R}^{n \times m}$ is the measurement matrix, $Z \in \mathbb{R}^{n \times l}$ is the measurement noise, and $Y \in \mathbb{R}^{n \times l}$ is the noisy measurement. We assume that the elements of $A$ are independent and identically distributed (i.i.d.) according to the Gaussian distribution $\mathcal{N}(0, \sigma_a^2)$, and the noise $Z_{i,j}$ are i.i.d. according to $\mathcal{N}(0, \sigma_z^2)$. We assume $\sigma_a^2$ and $\sigma_z^2$ are known.

Upon observing the noisy measurement $Y$, the goal is to recover the indices of the nonzero rows of $X$. A support recovery map is defined as

$$d : \mathbb{R}^{n \times l} \rightarrow 2^{[m]}.$$  \hspace{1cm} (3)

Given the signal model (1), the measurement model (2), and the support recovery map (3), we define the average probability of error by

$$P\{d(Y) \neq \text{supp}(X(W, S))\}$$
for each (unknown) signal value matrix \(W \in \mathbb{R}^{k \times l}\). Note that the probability is averaged over the randomness of the locations of the nonzero rows \(S\), the measurement matrix \(A\), and the measurement noise \(Z\).

III. INTERPRETATION OF SUPPORT RECOVERY VIA MULTIPLE-USER COMMUNICATION

We introduce an important interpretation of the problem of support recovery of sparse signals by relating it to a single-input multiple-output (SIMO) multiple access channel (MAC) communication problem. This relationship motivates the intuition behind our main results and facilities the development of the proof techniques. It can be also viewed as an MMV extension of our earlier work [56], in which a similar connection was employed to interpret the support recovery problem with SMV.

A. Brief Review on SIMO MAC

Consider the following wireless communication scenario. Suppose \(k\) senders wish to transmit information to a set of \(l\) common receivers. Each sender \(i\) has access to a codebook \(\mathcal{C}^{(i)} = \{c_{1}^{(i)}, \ldots, c_{m^{(i)}}^{(i)}\}\), where \(c_{j}^{(i)} \in \mathbb{R}^{n}\) is a codeword and \(m^{(i)}\) is the number of codewords in the codebook. The rate for 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 to \(l\) receivers leading to the SIMO MAC problem:

\[
Y_{j,i} = h_{j,1}X_{1,i} + h_{j,2}X_{2,i} + \cdots + h_{j,k}X_{k,i} + Z_{j,i}, \quad i = 1, 2, \ldots, n, \text{ and } j = 1, 2, \ldots, l
\]

where \(X_{q,i}\) denotes the input symbol from sender \(q\) to the channel at the \(i\)th use of the channel, \(h_{j,q}\) denotes the channel gain between sender \(q\) and receiver \(j\), \(Z_{j,i}\) is the additive Gaussian noise i.i.d. according to \(\mathcal{N}(0, \sigma^2_{z})\), and \(Y_{j,i}\) is the channel output at receiver \(j\) at the \(i\)th use of the channel.

After receiving \(Y_{j,1}, \ldots, Y_{j,n}\) at each receiver \(j \in [l]\), the receivers work jointly 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 error probability of decoding. By assuming each sender obeys the power constraint \(\|c_{j}^{(i)}\|^2/n \leq \sigma^2_{c}\) for all \(j \in [m^{(i)}]\) and all \(i \in [k]\), the capacity region of a SIMO MAC with known channel gains [57] is

\[
\left\{(R^{(1)}, \ldots, R^{(k)}) : \sum_{i \in T} R^{(i)} \leq \frac{1}{2} \log \left( I + \frac{\sigma^2_{z}}{\sigma^2_{c}} \sum_{i \in T} h_{i} h_{i}^{\top} \right) \right\}, \quad \forall \ T \subseteq [k]
\]
B. Similarities and Differences to the Problem of Support Recovery

Based on the measurement model (2), we can remove the columns in $A$ which correspond to the zero rows of $X$, and obtain the following effective form of the measurement procedure

$$Y_j = X_{S_1,j}A_{S_1} + \cdots + X_{S_k,j}A_{S_k} + Z_j$$

for $j \in [l]$. By contrasting (6) to the SIMO MAC (4), we can draw the following key connections that relate the two problems [58].

i) **A nonzero entry as a sender**: We can view the existence of a nonzero row index $S_i$ as sender $i$ that accesses the channel. Since there are $k$ nonzero entries, this results in $k$ users leading to the MAC analogy.

ii) **A measurement vector as a receiver**: We can view the existence of a measurement vector $Y_j$ as a measurement at receiver $j$. The multiple receivers leads to the multiple output (MO) part of the analogy.

iii) **$X_{S_i,j}$ as the channel gain**: The nonzero entry $X_{S_i,j}$, i.e., $w_{i,j}$, plays the role of the channel gain $h_{j,i}$ from the $i$th sender to the $j$th receiver.

iv) **$A_i$ as the codeword**: We treat the measurement matrix $A$ as a codebook with each column $A_i$, $i \in [m]$, as a codeword. Each element of $A_{S_i}$ is fed one by one through the channel as input symbols for the $i$th sender to the $l$ receivers, resulting in $n$ uses of the channel. Since a user transmits a single stream, this leads to the single input (SI) part of the analogy.

v) **Similarity of objectives**: In the problem of sparse signal recovery, we focus on finding the support \{$S_1, ..., S_k$\} of the signal. In the problem of MAC communication, the receiver needs to determine the indices of codewords, i.e., $S_1, ..., S_k$, that are transmitted by the senders.

Based on the abovementioned aspects, the two problems share significant similarities which enable leveraging the information theoretic methods for the SIMO MAC problem for the performance analysis of support recovery of sparse signals. However, 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 [56].

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 $h_i$ [59]. 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 [60]. 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 present the main results of this paper in the next section.

**IV. MAIN RESULTS AND THEIR INTERPRETATIONS**

**A. Main Results**

We consider the support recovery of a sequence of sparse signals generated with the same signal value matrix $W$. In particular, we assume that $k$ and $l$ are fixed. Define the auxiliary quantity

$$c(W) \triangleq \min_{T \subseteq [k]} \left[ \frac{1}{2|T|} \log \det \left( I + \frac{\sigma_0^2}{\sigma_z^2} W^T W_T \right) \right].$$  \hspace{1cm} (7)

The following two theorems summarize the main results. The proofs are presented in Appendices A and B.

**Theorem 1**: If

$$\limsup_{m \to \infty} \frac{\log m}{n_m} < c(W)$$  \hspace{1cm} (8)

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

$$\lim_{m \to \infty} P\{d(Y) \neq \text{supp}(X(W, S))\} = 0.$$  \hspace{1cm} (9)

**Theorem 2**: If

$$\limsup_{m \to \infty} \frac{\log m}{n_m} > c(W)$$  \hspace{1cm} (10)

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

$$\liminf_{m \to \infty} P\{d(Y) \neq \text{supp}(X(W, S))\} > 0.$$  \hspace{1cm} (11)
Theorems 1 and 2 together indicate that \( n = \frac{1}{c(W) \epsilon} \log m \) is the sufficient and necessary number of measurements per measurement vector to ensure asymptotically successful support recovery. The constant \( c(W) \) explicitly captures the role of the nonzero entries in the performance tradeoff.

B. Interpretations of the Main Results

We further explore the implications of having multiple measurement vectors. Due to the complicated nature of the expression for \( c(W) \), we will employ different approximations to make the interpretations more accessible.

1) The Low-Noise-Level Scenario: We consider the case where \( \sigma_z^2 \) is sufficiently small. Let \( \lambda_{T,i}, T \subseteq [k] \), denote the \( i \)th largest eigenvalue of \( W^T W_T \). For a SIMO MAC problem, the sum capacity grows as \( \min(k, l) \) leading to significant gains in the task of support recovery. This is captured in the following corollary.

Corollary 1: For a given \( W \), suppose \( \text{rank}(W^T W_T) = \min(|T|, l) \) for all \( T \subseteq [k] \). For sufficiently small \( \sigma_z^2 > 0 \), there exists a constant \( \alpha \in (0, 1) \) such that if

\[
\lim_{m \to \infty} \frac{\log m}{m} < \alpha \cdot \frac{\min(k, l)}{2k} \cdot \log \frac{\sigma_a^2}{\sigma_z^2}
\]

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

\[
\lim_{m \to \infty} P\{d(Y) \neq \text{supp}(X(W, S))\} = 0.
\]

Proof: Note that for \( T \subseteq [k] \) with \( |T| \leq l \), \( \lambda_{T,i} > 0 \) for \( i = 1, 2, ..., |T| \). Thus

\[
\frac{1}{2|T|} \log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} W^T W_T \right) = \frac{1}{2|T|} \log \prod_{i=1}^{|T|} \left( 1 + \frac{\sigma_a^2}{\sigma_z^2} \lambda_{T,i} \right)
\]

\[
= \frac{1}{2|T|} \log \prod_{i=1}^{|T|} \left( \frac{\sigma_a^2}{\sigma_z^2} + \frac{\sigma_a^2}{\sigma_z^2} \lambda_{T,i} \right)
\]

\[
= \frac{1}{2|T|} \left[ |T| \cdot \log \frac{\sigma_a^2}{\sigma_z^2} + \sum_{i=1}^{|T|} \log \left( \frac{\sigma_a^2}{\sigma_z^2} + \lambda_{T,i} \right) \right]
\]

\[
= \frac{1}{2} \log \frac{\sigma_a^2}{\sigma_z^2} \cdot \left( 1 + \frac{1}{|T|} \sum_{i=1}^{|T|} \log \left( \frac{\sigma_a^2}{\sigma_z^2} + \lambda_{T,i} \right) \right)
\]

\[
= \frac{1}{2} \log \frac{\sigma_a^2}{\sigma_z^2} \cdot \left( 1 + O \left( \frac{1}{-\log \sigma_z^2} \right) \right)
\]

\[
\geq \frac{1}{2} \log \frac{\sigma_a^2}{\sigma_z^2} \cdot \alpha_T
\]
for some $\alpha_T \in (0, 1)$. For any possible $T \subseteq [k]$ with $|T| > l$, $\lambda_{T,i} > 0$ for $i = 1, 2, \ldots, l$. Then, we have similarly
\begin{equation}
\frac{1}{2^{|T|}} \log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} W_T^T W_T \right) = \frac{l}{2^{|T|}} \log \frac{\sigma_a^2}{\sigma_z^2} \left( 1 + O \left( \frac{1}{\log \sigma_z^2} \right) \right).
\end{equation}

Thus, if $k \leq l$
\begin{equation}
\min_{T \subseteq [k]} \left[ \frac{1}{2^{|T|}} \log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} W_T^T W_T \right) \right] \geq \frac{1}{2} \log \frac{\sigma_a^2}{\sigma_z^2} \cdot \min_{T \subseteq [k]} \alpha_T
\end{equation}
and if $k > l$
\begin{equation}
\min_{T \subseteq [k]} \left[ \frac{1}{2^{|T|}} \log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} W_T^T W_T \right) \right] \geq \frac{l}{2^k} \log \frac{\sigma_a^2}{\sigma_z^2} \cdot \min_{T \subseteq [k]} \alpha_T.
\end{equation}
Combining (13) and (14) and applying Theorem 1 complete the proof.

Corollary 1 indicates the following observations. First, as the measurement noise level $\sigma_z^2$ approaches zero, the term $\frac{\min(k,l)}{2k} \log \frac{\sigma_z^2}{\sigma_a^2}$ exerts a major influence on the sufficient condition (12). The nonzero signal matrix $W$ plays its role mainly through the ranks of its row-wise submatrices, which are ensured to be full rank according to the technical assumption that $\text{rank}(W_T^T W_T) = \min(|T|, l)$ for any $T \subseteq [k]$.

Second, by rearranging the terms in (12), we obtain
\begin{equation}
m = \left( \frac{\sigma_z^2}{\sigma_a^2} \right)^{\alpha \min(k,l)} \frac{1}{\sigma_a^2}
\end{equation}
which corresponds to the maximum number of columns of $A$ that still yields a diminishing error probability in support recovery. Specifically, the term $\min(k,l)$ reveals the following insight. In the scenario with sufficiently small $\sigma_z^2$, for the challenging problem where the number of measurement vectors is less than the number of nonzero rows, i.e., $l < k$, adding one more measurement vector can lead to a much larger upper bound on the manageable number of columns of $A$. On the other hand, when $k \leq l$, the problem is much simpler and adding more measurement vectors may not significantly increase the manageable size of $A$. From an algorithmic point of view, subspace based methods can be used to recover the support in the latter case.

2) The Role of the Nonzero Signal Matrix: Next, we take a closer look at the role of the nonzero signal matrix $W$ in support recovery with MMV. We consider two different cases. In the first case, $W$ consists of identical columns. The following corollary states the corresponding sufficient condition for support recovery.
**Corollary 2:** Suppose $W \in \mathbb{R}^{k \times l}$ has identical columns, i.e., $W = [w, ..., w]$, for some $w \in \mathbb{R}^k$ with all entries being nonzero. If
\[
\lim_{m \to \infty} \frac{\log m}{n_m} < \min_{T \subseteq [k]} \frac{1}{2|T|} \log \left( 1 + l \cdot \frac{\sigma_a^2}{\sigma_z^2} \|w_T\|_2^2 \right)
\]  
then there exists a sequence of support recovery maps $\{d^{(m)}\}_{m=k}^{\infty}, d^{(m)} : \mathbb{R}^{n_m \times l} \to 2^m$, such that
\[
\lim_{m \to \infty} P\{d(Y) \neq \text{supp}(X(W, S))\} = 0.
\]

**Proof:** Note that, for any $T \subseteq [k],$
\[
\log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} W_T^T W_T \right) = \log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} [w_T, ..., w_T]^T [w_T, ..., w_T] \right)
\]
\[
= \log \det \left( I + \frac{\sigma_a^2}{\sigma_z^2} \|w_T\|_2^2 \cdot 1 \cdot 1^T \right)
\]
\[
= \log \left( 1 + l \cdot \frac{\sigma_a^2}{\sigma_z^2} \|w_T\|_2^2 \right).
\]
Applying Theorem 1 completes the proof. $\blacksquare$

Based on (15), the effect of having $l$ identical nonzero signal vectors is equivalent to decreasing the noise level by a factor of $l$, compared to the problem with SMV. This is in accordance with the intuition that when the underlying signals remain the same, taking more measurement vectors provides an opportunity to average down the measurement noise level. We hasten to add that identical columns are unlikely in practice. Even small changes in the coefficients can lead to a full rank matrix, leading to significant benefits in the high signal-to-noise ratio (SNR) case.

In the second case, we construct a special example to achieve a large performance improvement via a second measurement. This is demonstrated in the following corollary.

**Corollary 3:** Suppose $W = [w_1, w_2] \in \mathbb{R}^{k \times 2}$, where $k$ is even, $w_1 = 1 \in \mathbb{R}^k$, and $w_2$ is defined as
\[
w_{i,2} = \begin{cases} 1 & \text{if } 1 < i \leq \frac{k}{2}, \\ -1 & \text{if } \frac{k}{2} < i \leq k. \end{cases}
\]  
(16)

If
\[
\lim_{m \to \infty} \frac{\log m}{n} < \frac{1}{k} \log \left( 1 + k \cdot \frac{\sigma_a^2}{\sigma_z^2} \right)
\]  
(17)
then there exists a sequence of support recovery maps $\{d^{(m)}\}_{m=k}^{\infty}, d^{(m)} : \mathbb{R}^{n_m \times l} \to 2^m$, such that
\[
\lim_{m \to \infty} P\{d(Y) \neq \text{supp}(X(W, S))\} = 0.
\]

**Proof:** Please see Appendix C. $\blacksquare$
For the ease of illustration, we compare the performances among the problems with (i) SMV where \( W = 1 \in \mathbb{R}^{k \times 1} \), (ii) MMV where \( W = [1, 1] \in \mathbb{R}^{k \times 2} \), and (iii) MMV where \( W \) is defined in Corollary 3, for an even \( k \). The following table summarizes the results.

|                           | lower bound on \( n \)                              | upper bound on \( m \)                              |
|---------------------------|----------------------------------------------------|----------------------------------------------------|
| (i) SMV (\( W = 1 \))     | \( n > \frac{\log m}{k \log (1 + \frac{k \sigma^2}{\sigma^2})} \) | \( m < \left(1 + \frac{k \sigma^2}{\sigma^2}\right)^\frac{1}{k} \) |
| (ii) MMV (\( W = [1, 1] \), Corollary 2) | \( n > \frac{\log m}{k \log (1 + 2 \cdot \frac{k \sigma^2}{\sigma^2})} \) | \( m < \left(1 + 2 \cdot \frac{k \sigma^2}{\sigma^2}\right)^\frac{1}{k} \) |
| (iii) MMV (\( W \) as defined in Corollary 3) | \( n > \frac{\log m}{k \log (1 + \frac{k \sigma^2}{\sigma^2})} \) | \( m < \left(1 + \frac{k \sigma^2}{\sigma^2}\right)^\frac{1}{k} \) |

Based on this table, we have the following observations for this specific setup. First, compared with the SMV problem, having MMV can improve the performance of support recovery by enabling a relaxed condition on the number of measurements \( n \). Equivalently, for the same number of measurements per measurement vector, the MMV setup permits a measurement matrix \( A \) with more columns. Second, the performance improvement enabled by having MMV is closely related to \( c(W) \), and it can be quite different for different nonzero signal value matrices. In case (ii), we achieve a moderate performance gain which is equivalent to reducing the noise level by half. On the contrary, in case (iii), a larger performance gain can be achieved due to the structure of the nonzero signal value matrix. Note that the change occurs in the factor in the exponent in the upper bound for \( m \). In summary, these examples are specially constructed as representative cases to illustrate the effect of the nonzero signal value matrix \( W \) in support recovery. Generally, the difficulty of a support recovery problem is inherently determined by the model parameters and Theorems 1 and 2 together characterize their exact roles.

3) A Generalization of \( W \): Thus far, we have assumed \( w_{i,j} \neq 0 \) for all \( i, j \) in the discussion above. Now, we generalize \( W \) in the following manner: for each \( i \in [k] \), there exist a \( j \in [l] \) such that \( w_{i,j} \neq 0 \); meanwhile, for each \( j \in [l] \), there exist a \( i \in [k] \) such that \( w_{i,j} \neq 0 \). This relaxed assumption indicates that neither a zero row nor a zero column exists but zero elements are allowed in \( W \), as opposed to the original assumption that all elements of \( W \) are nonzeros. Accordingly,

\[
\text{supp}(X) = \bigcup_{j=1}^{l} \text{supp}(X_j)
\]

Note that \( \|w_1\|_2 = \|w_2\|_2 \), which can be viewed as a way of normalization to make comparison meaningful.
which means the support of $X$ is equivalent to the union of the supports of all columns of $X$. Following
the proofs for Theorem 1 and 2, one can readily see that the two theorems still hold in this case.

It is worthwhile to note that having more measurement vectors does not necessarily result in performance improvement. To illustrate this point, we construct a simple example. Let $W^{(1)} = [0.1, 5]^T$, $W^{(2)} = \begin{bmatrix} 0.1 & 0 \\ 5 & 6 \end{bmatrix}$, and $\frac{\sigma_a^2}{\sigma_z^2} = 10$. As a result, $c(W^{(1)}) = c(W^{(2)}) = \frac{1}{2} \log 1.1$. This means that the performance limits for these two setups are the same. Intuitively, by inspecting the definition of $c(W)$, it can be seen that if a submatrix composed of certain rows of $W$ is ill-conditioned, the minimization inside $c(W)$ may likely be determined by that submatrix. Hence, for an extra measurement vector to benefit support recovery, this measurement vector should correspond to a column of $W$ whose presence improves the small eigenvalues of the previous worst-case submatrix that causes the performance bottleneck. The observations are reminiscent of some of the intuition developed in space time wireless communication systems [61]. The $l$ receivers can be viewed as an $l$ antenna receiver and it is known that the rank of the channel matrix plays an important role in the high SNR case. The correlation between the channel gains is not as harmful in this context. The gains of having multiple receive antennas is lower at low SNR [61].

V. RELATION TO EXISTING RESULTS

We discuss the relation between the main results in this paper and existing results in the literature.

A. Relation to the Performance of Practical Algorithms

Our analysis provides the performance limit that governs all possible support recovery algorithms. This is achieved by a theoretic support recovery method which has exponential complexity and therefore is impractical. However, it is interesting to make comparisons with performance limits of practical algorithms, since it provides insight into the potential gap between the performance of a practical algorithm and the fundamental performance limit, and suggests possibilities for performance improvement.

We note that the model employed in Obozinski, Wainwright, and Jordan [45] is similar to the measurement model (2). Sufficient and necessary conditions are derived therein for multivariate group Lasso to successfully recover the support of the sparse signal in the presence of noise, as $m, n$, and $k$ jointly grow.
to infinity in certain manner. This is different from our assumption that $k$ is fixed. Although a direct comparison may seem difficult, we wish to draw the following intuitive discussion. Note that Example 1 in [45, Section 2.3] considered the case for identical regression, which means the nonzero signal matrix $W$ has identical columns. The conclusion therein is that multivariate group Lasso offers no performance improvement under the MMV formulation compared with using Lasso on an SMV formulation with one measurement vector. However, our Corollary 2 indicates that the effect of having $l$ identical columns in $W$ is equivalent to lowering the noise level by a factor of $l$. The different performances indicated by multivariate group Lasso and the information theoretic analysis lead to the following observation. In general, if the sparse signal to be recovered possesses strong structural property, an algorithm needs to take advantage of this factor in order to achieve better performance. For multivariate group Lasso, the $\ell_1/\ell_p$ cost term completely ignores the row-wise structure presented in the nonzero entries. In contrast, AR-SBL [50] is developed based on the assumption that the elements of $W$ are drawn from an auto-regressive process, and it explicitly attempts to learn this correlation structure. Based on the experimental study presented in [50], notable performance improvement in support recovery was observed when such correlation is present, including the case when the columns of $W$ were highly correlated.

B. Relation to Information Theoretic Performance Analysis

Under the assumption that $\sigma^2_a = 1$ and the elements of $W$ are i.i.d. according to $\mathcal{N}(0,1)$, Tang and Nehorai [53] identifies sufficient and necessary conditions, involving the model parameters (i.e., $m, n, k, l, \text{and } \sigma^2_z$), to ensure diminishing error probability in support recovery as the problem size grows to infinity. We restate the sufficient condition to facilitate the discussion.

Theorem 3 ([53, Theorem 5]): Suppose that $n = \Omega(k \log \frac{m}{k})$ and $\frac{1}{2} \log \frac{n}{\sigma^2_z} \gg \log(k(m-k))$, then with probability one the error probability vanishes. In particular, if $n = \Omega(k \log \frac{m}{k})$ and $l \gg \frac{\log m}{\log \log m}$, the error probability vanishes as $m \to \infty$.

As noted in [53], heuristically, when $l = 1$, $n \gg m$ is needed to guarantee asymptotically successful support recovery. Although our main results aim for the case with fixed $W$, intuitive observations can

\footnote{Note that it is stated at the end of Section 3.3 of [45] that the requirement on $k$ growing to infinity can be removed. The remark therein provided an alternative probability upper bound for the intermediate term $T_1$ such that this bound can drop to zero even for a fixed $k$. However, it seems that the other intermediate term $T_2$ still relies on a probability upper bound that involves a term scaling as $\exp(-\frac{k}{4})$, which requires an increasing $k$ to drive it to zero.}

\footnote{We only consider the real case in this discussion.}
still be drawn to provide more insight into the behavior of the support recovery with random $W$. To see this, recall that for a sequence of support recovery problems with a fixed $W$, the quantity $c(W)$ inherently determines the performance limit and the sufficient condition is $n > \frac{1}{c(W)} \log m$. Now, let us assume that the elements of $W$ are i.i.d. according to certain distribution with bounded support. Thus, in general, for any constant $\delta > 0$, the probability $P(c(W) < \delta)$ may be strictly positive. This implies that for the scaling $n = \Theta(\log m)$, the error probability will not converge to zero because there is a nontrivial probability of poor realizations of $W$ such that the sufficient condition above cannot be satisfied. As one plausible solution, we need $n$ to grow with $m$ at a much faster rate to ensure that the sufficient condition above can be met with probability converging to one.

**VI. CONCLUSION AND DISCUSSION**

We have developed performance limits for support recovery of sparse signals when multiple measurement vectors are available. Sufficient and necessary conditions are obtained for support recovery to be asymptotically successful. Especially, the role of nonzero entries in the performance limits is explicitly characterized, and the quantity $c(W)$ captures the effect of all nonzero entries. The key technique that enabled our analysis is motivated by the connection between sparse signal recovery with MMV and multiple access communication over SIMO channels. This leads to the opportunity of leveraging the methodology for deriving SIMO MAC capacity to help understand the performance limits of sparse signal recovery with MMV. Interpretations of the main results were provided in order to demonstrate the performance improvement by having MMV, and relations to existing results were also discussed.

The proposed methodology also has the potential to address other theoretical and practical issues associated with sparse signal recovery. First, this analytical approach can be extended to deal with the case where the signal value matrix $W$ is random. Outage analysis for fading channels can be leveraged to reveal the performance limits for sparse signal recovery in this case. Second, one can consider the problem where recovering a partial support is also desirable, if recovering the full support is not possible [62]. This can be achieved by treating a subset of users as noise and examining the capacity region of the remaining users. The connection between sparse signal recovery and multiple access communication offers the opportunity to explore the connection between sparse recovery algorithms and multiuser detection techniques with potential for cross-fertilization. A sender with larger channel gain may be easier to detect compared to a sender with weaker channel gain. The successive interference cancellation (SIC) scheme is aimed to detected users in a sequential manner, where the remaining undetected users are treated as noise bearing a strong resemblance to the matching pursuit algorithms for sparse signal recovery. It is
conceivable that by appropriately utilizing the techniques for channel coding, performance limits could be obtained for partial support recovery of sparse signals.

Further, according to the interpretations of the main results, we can see that the structure of $W$ plays an important role in the performance limits. Roughly speaking, high correlation among the columns of $W$ may decrease the performance limit for support recovery, in the sense that, given other parameters fixed, the dimension of the signal $m$ should be reduced to guarantee successful support recovery. However, as observed in practice, when only a finite number of measurements per measurement vector are available, a strong correlation among columns of $W$ actually facilitates the estimation of the nonzero signal values, and hence can be beneficial to the performance. Hence, there is an interplay that is not revealed by the asymptotic analysis. It will be interesting to study an analytical approach which links the estimation quality of nonzero values in the finite case and performance limits of support recovery in the asymptotic case.

**ACKNOWLEDGMENT**

This research was supported by NSF Grants CCF-0830612 and CCF-1144258. The authors wish to thank Professor Young-Han Kim for insightful discussions on the problem of sparse signal recovery and its connection to multiple access communication. The proof techniques extend upon our previous methodology for the SMV case, to which Professor Kim made significant contributions.
**APPENDIX A**

**PROOF OF THEOREM 1**

For the ease of exposition, we consider two distinct cases on the number of nonzero rows of $X$.

**Case 1:** $k = 1$. In this case, the signal of interest is $X = X(W, S_1)$, where $W = [w_{1,1}, ..., w_{1,l}]$. Fix $\epsilon > 0$. We first form an estimate $\hat{\rho}_i$ of $|w_{1,i}|$ for $i \in [l]$ as

$$\hat{\rho}_i \triangleq \sqrt{\frac{1}{nm} \|Y_i\|_2^2 - \sigma_z^2}.$$  

(18)

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

$$\frac{1}{nl} \|Y - A_{\hat{s}_1} [(-1)^{q_1} \hat{\rho}_1, ..., (-1)^{q_l} \hat{\rho}_l]_F \leq \sigma_z^2 + \epsilon^2 \sigma_a^2.$$  

(19)

for $q_i = 1$ or $q_i = 2$, $i \in [l]$. If there is none or more than one such index, pick an arbitrary index.

We analyze the average probability of error

$$P(\mathcal{E}) = P\{d^{(m)}(Y) \neq \text{supp}(X(W, S_1))\}.$$  

(20)

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

$$Y = A_1 W + Z$$  

(21)

for some $W = [w_{1,1}, ..., w_{1,l}] \in \mathbb{R}^{1 \times l}$. In the following analysis, we drop superscripts and subscripts on $m$ for notational simplicity when no ambiguity arises. Define the events

$$\mathcal{E}_s \triangleq \left\{ \forall i \in [l], \exists q_i \in \{1, 2\}, \text{ such that } \frac{1}{nl} \|Y - A_s [(-1)^{q_1} \hat{\rho}_1, ..., (-1)^{q_l} \hat{\rho}_l]_F \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \right\}, \quad s \in [m].$$

Then,

$$P(\mathcal{E}) \leq P(\mathcal{E}^c \cup (\cup_{s=2}^{m} \mathcal{E}_s))$$  

(22)

where $\mathcal{E}^c$ denotes the compliment event of $\mathcal{E}$. Let

$$\mathcal{E}_{\text{aux}} \triangleq \left\{ \text{det} \left( \frac{1}{n} (A_1 W + Z)^T (A_1 W + Z) \right) - \text{det} \left( \sigma_a^2 W^T W + \sigma_z^2 I \right) \in (-\epsilon, \epsilon) \right\} \cap \left\{ \bigcap_{i=1}^l \{\hat{\rho}_i - |w_{1,i}| \in (-\epsilon, \epsilon) \} \right\}.$$  

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

$$P(\mathcal{E}) \leq P(\mathcal{E}^c_{\text{aux}}) + P(\mathcal{E}^c_1) + \sum_{s=2}^{m} P(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}).$$  

(23)
We bound each term in (23). First, by the weak law of large numbers (LLN), \( \lim_{m \to \infty} P(\mathcal{E}_{\text{aux}}^c) = 0 \). Next, we consider \( P(\mathcal{E}_f^c) \). It can be readily seen that, with \( q_i = (3 + \text{sign}(w_{1,i})) / 2 \),

\[
\lim_{m \to \infty} P \left( \frac{1}{m} \left\| Y - A_1[(−1)^{q_1} \hat{ρ}_1, \ldots, (−1)^{q_l} \hat{ρ}_l] \right\|_F^2 \leq \sigma_1^2 + \epsilon^2 \sigma_0^2 \right) = 1.
\]

(24)

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

Next, we consider the third term in (23). We need the following lemma, whose proof is presented at the end of this appendix.

**Lemma 1:** Let \( B \in \mathbb{R}^{n \times l} \) be a fixed matrix satisfying \( \prod_{i=1}^{l} \left[ \frac{1}{n} B^T B \right]_{i,i} \) \( \equiv \alpha > 0 \). Let \( S \subseteq [l] \) be a fixed set. Let \( D \in \mathbb{R}^{n \times l} \) be a matrix such that, for \( j \in S \), \( D_j \sim \mathcal{N}(0, \theta_j I) \) with some \( \theta_j > 0 \); for \( j \in [l] \setminus S \), \( D_j \equiv 0 \). All columns of \( D \) are independent. Then, for any \( \gamma \in (0, \alpha) \),

\[
P \left( \frac{1}{n} \left\| B - D \right\|_F^2 \leq \gamma \right) \leq 2^{-\frac{n}{2} \log \frac{\alpha}{\gamma}}.
\]

(25)

We continue the proof of Theorem 1. Consider \( P(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}) \) for \( s \neq 1 \). Note that

\[
P(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}) \leq P(\mathcal{E}_s | \mathcal{E}_{\text{aux}}) = \int_{Y_1 \in \mathcal{E}_{\text{aux}}} P(\mathcal{E}_s \mid \{Y = Y_1\} \cap \mathcal{E}_{\text{aux}}) f(Y_1 | \mathcal{E}_{\text{aux}}) dY_1.
\]

Let \( [(−1)^{q_1} \hat{ρ}_1, \ldots, (−1)^{q_l} \hat{ρ}_l] = U \Theta V^T \) denote the singular value decomposition. Since \( A_s \) is independent of \( Y \) and \( \hat{ρ}_i \) for \( s \neq 1 \), it follows from Lemma 1 that (by treating \( B = YV \) and \( D = A_s U \Theta \)), for \( q_i \in \{1, 2\} \), \( i \in [l] \) and sufficiently small \( \epsilon \),

\[
P \left( \frac{1}{n} \left\| Y - A_s[(−1)^{q_1} \hat{W}_1, \ldots, (−1)^{q_l} \hat{W}_l] \right\|_F^2 \leq \sigma_1^2 + \epsilon^2 \sigma_0^2 \right) \{Y = Y_1\} \cap \mathcal{E}_{\text{aux}}
\]

\[
\leq 2^{-\frac{n}{2} \log \frac{\alpha}{\gamma}}.
\]

(26)

where (26) follows from the Hadamard’s inequality [63]. Thus,

\[
P(\mathcal{E}_s \mid \{Y = Y_1\} \cap \mathcal{E}_{\text{aux}}) \leq 2^l \cdot 2^{-\frac{n}{2} \log \left( \frac{\alpha \sigma_0^2}{\sigma_0^2 + \epsilon^2 \sigma_0^2} \right)}
\]

and hence

\[
\sum_{s = 2}^{m} P(\mathcal{E}_s \cap \mathcal{E}_{\text{aux}}) \leq 2^l \cdot m \cdot 2^{-\frac{n}{2} \log \left( \frac{\alpha \sigma_0^2}{\sigma_0^2 + \epsilon^2 \sigma_0^2} \right)}
\]
which tends to zero as \( m \to \infty \), if
\[
\limsup_{m \to \infty} \frac{\log m}{n_m} < \frac{1}{2} \log \left( \frac{\det (\sigma^2 W^\tau W + \sigma^2 I) - \epsilon}{(\sigma^2 + \epsilon \sigma^2)} \right).
\] (27)

Since \( \epsilon > 0 \) is chosen arbitrarily, we have the desired proof of Theorem 1.

**Case 2:** \( k \geq 2 \). In this case, the signal of interest is \( X = X(W, S) \). Fix \( \epsilon > 0 \). First, for \( i \in [l] \), we form an estimate of \( \|w_i\|_2 \) as
\[
\hat{\rho}_i \triangleq \sqrt{\frac{\|Y_i\|_2^2 - \sigma_a^2}{\sigma_a^2}}.
\] (28)

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

i) \( Q \subseteq B_k(r) \), where \( B_k(r) \) is the \( k \)-dimensional hypersphere of radius \( r \).

ii) For any \( b \in B_k(r) \), there exists \( \hat{w} \in Q \) such that \( \|\hat{w} - b\|_2 \leq \frac{\zeta}{2} \).

The following properties can be easily proved:

**Lemma 2:**
1) For \( i \in [l] \), \( \lim_{m \to \infty} P \left( \exists \hat{W} \in Q(\hat{W}_i, \zeta) \text{ such that } \|\hat{W} - w_i\|_2 < \zeta \right) = 1.

2) \( q(r, \zeta) \triangleq |Q(r, \zeta)| \) is monotonically non-decreasing in \( r \) for fixed \( \zeta \).

For \( i \in [l] \), given \( \hat{\rho}_i \) and \( \epsilon \), fix \( Q_i = Q_{\hat{\rho}_i}(\hat{\rho}_i, \epsilon) \). Declare \( d(Y) = \{\hat{s}_1, ..., \hat{s}_k\} \subseteq [m] \) is the recovered set of indices of nonzero rows of \( W \), if it is the unique set of indices such that
\[
\frac{1}{nl} \left\| Y - [A_{\hat{s}_1}, ..., A_{\hat{s}_k}] \left[ \hat{W}_1, ..., \hat{W}_l \right] \right\|_F^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2
\] (29)

for some \( \hat{W}_i \in Q_i, i \in [l] \). 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
\[
P(\mathcal{E}) = P\{d(Y) \neq X(W, S)\}.
\] (30)

Without loss of generality, we assume that \( S_j = j \) for \( j = 1, 2, ..., k \), which gives
\[
Y = [A_1, ..., A_k]W + Z
\] (31)

for some \( W \). Define the event
\[
\mathcal{E}_{s_1, s_2, ..., s_k} \triangleq \left\{ \exists \hat{W}_i \in Q_i \text{ and } \{s_1', ..., s_k'\} = \{s_1, ..., s_k\} \text{ s.t. } \frac{1}{nl} \left\| Y - [A_{s_1'}, ..., A_{s_k'}] \left[ \hat{W}_1', ..., \hat{W}_l' \right] \right\|_F^2 \leq \sigma_z^2 + \epsilon^2 \sigma_a^2 \right\}.
\]

Define \( \sigma^2_{\max} \) and \( \sigma^2_{\min} \) to be the largest and smallest eigenvalues of the matrix
\[
\frac{1}{n \sigma^2_z} [A_1, ..., A_k, \frac{\sigma_a}{\sigma_z} Z]^T [A_1, ..., A_k, \frac{\sigma_a}{\sigma_z} Z]
\]
respectively. Then
\[
P(\mathcal{E}) = P \left( \mathcal{E}_{1,2,\ldots,k} \cup \bigcup_{s_1 < \ldots < s_k; \{s_1, \ldots, s_k\} \neq [k]} \mathcal{E}_{s_1,s_2,\ldots,s_k} \right) \\
\leq P \left( \mathcal{E}_{\text{aux}}^c \cup \mathcal{E}_{1,2,\ldots,k}^c \cup \bigcup_{s_1 < \ldots < s_k; \{s_1, \ldots, s_k\} \neq [k]} (\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}) \right) \\
\leq P(\mathcal{E}_{\text{aux}}^c) + P(\mathcal{E}_{1,2,\ldots,k}^c) + \sum_{s_1 < \ldots < s_k; \{s_1, \ldots, s_k\} \neq [k]} P(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}) \tag{32}
\]
where
\[
\mathcal{E}_{\text{aux}} \triangleq \{ \sigma_{\text{max}}^2 \in (1 - \epsilon, 1 + \epsilon) \} \cap \{ \sigma_{\text{min}}^2 \in (1 - \epsilon, 1 + \epsilon) \} \cap \left( \bigcap_{i=1}^l \{ \hat{W}_i - \|W_i\|_2 \in (-\epsilon, \epsilon) \} \right).
\]

First, note that \( \lim_{m \to \infty} P(\mathcal{E}_{\text{aux}}) = 1 \) due to LLN and the properties of the extreme eigenvalues of random matrices \([64]\). Next, consider
\[
\frac{1}{nl} \left\| Y - [A_1, \ldots, A_k] \left[ \hat{W}_1, \ldots, \hat{W}_l \right] \right\|_F^2 \\
= \frac{1}{nl} \left\| [A_1, \ldots, A_k] W + Z - [A_1, \ldots, A_k] \left[ \hat{W}_1, \ldots, \hat{W}_l \right] \right\|_F^2 \\
= \frac{1}{nl} \left\| [A_1, \ldots, A_k, \frac{\sigma_a}{\sigma_z} Z \left[ W - \left[ \hat{W}_1, \ldots, \hat{W}_l \right] \right] \right\|_F^2 \\
\leq \frac{1}{l} \sigma_{\text{max}}^2 \left( \frac{\sigma_a^2}{\sigma_z^2} \left\| W - \left[ \hat{W}_1, \ldots, \hat{W}_l \right] \right\|^2_F + \sigma_z^2 \right)
\tag{33}
\]
By using the fact that \( \sigma_{\text{max}}^2 \to 1 \) almost surely as \( n \to \infty \) \([64]\) and Lemma 2-1), we have \( \lim_{m \to \infty} P(\mathcal{E}_{1,2,\ldots,k}^c) = 0 \).

Next, we consider \( P(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}) \) for \( \{s_1, s_2, \ldots, s_k\} \neq [k] \). Note that
\[
P(\mathcal{E}_{s_1,s_2,\ldots,s_k} \cap \mathcal{E}_{\text{aux}}) \\
\leq P(\mathcal{E}_{s_1,s_2,\ldots,s_k} | \mathcal{E}_{\text{aux}}) \\
= \int \ldots \int_{\{a_1, \ldots, a_k, Z_0\} \in \mathcal{E}_{\text{aux}}} P(\mathcal{E}_{s_1,s_2,\ldots,s_k} | \{A_1 = a_1\} \cap \ldots \cap \{A_k = a_k\} \cap \{Z = Z_0\} \cap \mathcal{E}_{\text{aux}}) \\
\times f(a_1, \ldots, a_k, Z_0 | \mathcal{E}_{\text{aux}}) da_1 \cdots da_k dZ_0. \tag{34}
\]
For notational simplicity, define \( \xi \triangleq \sigma_z^2 + \xi^2 \sigma_a^2 \), \( \mathcal{T} \triangleq \{s_1, s_2, \ldots, s_k\} \cap [k] \), \( \mathcal{T}^c \triangleq \{s_1, s_2, \ldots, s_k\} \setminus \mathcal{T} \), and \( \mathcal{E}_{\text{cond}} \triangleq \{A_1 = a_1\} \cap \ldots \cap \{A_k = a_k\} \cap \{Z = Z_0\} \cap \mathcal{E}_{\text{aux}} \). For any permutation \( (s'_1, s'_2, \ldots, s'_k) \) of
\{s_1, s_2, ..., s_k\} and any \( \hat{W}_i \in Q_i, i \in [l] \),
\[
P \left( \frac{1}{nl} \| Y - [A_{s_1'}, ..., A_{s_k'}] \left[ \hat{W}_1, ..., \hat{W}_l \right] \|_F^2 \leq \xi \| \mathcal{E}_{\text{cond}} \right)
= P \left( \frac{1}{nl} \| [A_1, ..., A_k]W + Z - [A_{s_1'}, ..., A_{s_k'}] \left[ \hat{W}_1, ..., \hat{W}_l \right] \|_F^2 \leq \xi \| \mathcal{E}_{\text{cond}} \right) \tag{35}
\]
Define the matrix \( W' \in \mathbb{R}^{k \times l} \) as
\[
W'_j = \begin{cases} 
W_j & \text{if } j \in [k] \setminus T \\
W_j - \tilde{W}_i & \text{if } j = s_i' \in T
\end{cases}
\tag{36}
\]
where \( \tilde{W}_i \) denotes the \( i \)th row of the matrix \([\hat{W}_1, ..., \hat{W}_l]\). Define \( \tilde{W}' \in \mathbb{R}^{k \times l} \) as
\[
\tilde{W}'_j = \begin{cases} 
\tilde{W}_j & \text{if } s_j' \notin T \\
\mathbf{0} & \text{if } s_j' \in T
\end{cases}
\tag{37}
\]
where \( \mathbf{0} \) is a zero row vector of a proper size. Then, continue from (35), we have
\[
P \left( \frac{1}{nl} \| Y - [A_{s_1'}, ..., A_{s_k'}] \left[ \hat{W}_1, ..., \hat{W}_l \right] \|_F^2 \leq \xi \| \mathcal{E}_{\text{cond}} \right)
= P \left( \frac{1}{nl} \| [A_1, ..., A_k, \frac{\sigma_a}{\sigma_z} Z] \left[ \begin{array}{c} W' \\ \frac{\sigma_a}{\sigma_z} I \end{array} \right] - [A_{s_1'}, ..., A_{s_k'}] \tilde{W}' \right\|_F^2 \leq \xi \| \mathcal{E}_{\text{cond}} \right) \tag{38}
\]
\[
\equiv P \left( \frac{1}{nl} \| [A_1, ..., A_k, \frac{\sigma_a}{\sigma_z} Z] \left[ \begin{array}{c} W' \\ \frac{\sigma_a}{\sigma_z} I \end{array} \right] - \tilde{A} \tilde{W}'_1 \right\|_F^2 \leq \xi \| \mathcal{E}_{\text{cond}} \right) \tag{39}
\]
where in (38) \( \tilde{W}'_1 \) denotes matrix formed by removing the zero rows in \( \tilde{W}' \), and \( \tilde{A} \) denotes the matrix by removing columns of \([A_{s_1'}, ..., A_{s_k'}]\) indexed by the indices of the zero rows of \( \tilde{W}' \). To reach (39), let \( \tilde{W}'_1 = U \Theta V^\top \) denote the singular value decomposition. The follow lemma, the proof of which is presented at the end of this appendix, is useful.

**Lemma 3:** Let \( B \in \mathbb{R}^{p \times q}, D \in \mathbb{R}^{q \times r} \). Let \( \sigma_B^2 \) denote the smallest eigenvalue of \( B^\top B \). Then
\[
det((BD)^\top BD) \geq (\sigma_B^2)^r \det(D^\top D).
\]
Let \( M \triangleq [A_1, ..., A_k, \frac{\sigma_a}{\sigma_z} Z] \left[ \begin{array}{c} W' \\ \frac{\sigma_a}{\sigma_z} I \end{array} \right] V \). Conditioned on \( \mathcal{E}_{\text{cond}} \) and the chosen \( Q_i \) for \( i \in [l] \), \( M \) is
fixed. According to Lemma 3 (treating $B = \frac{1}{\sqrt{n}}[A_1, ..., A_k, \frac{\sigma}{\sigma_a} Z]$ and $D = \left[ \begin{array}{c} W' \\ \frac{\sigma}{\sigma_a} I \end{array} \right] V$),
\[
\det\left( \frac{1}{n} M^T M \right) \geq \left( (1 - \epsilon) \sigma_a^2 \right)^l \det\left( \left[ \begin{array}{c} W' \\ \frac{\sigma}{\sigma_a} I \end{array} \right] \right) .
\]
(40)

Continue with (39). Using Lemma 2 (treating $B = M$ and $D = \tilde{A} U \Theta$), we have
\[
P\left( \frac{1}{n l} \left\| Y - [A_{s_1}', ..., A_{s_k}'] \left[ \tilde{W}_1, ..., \tilde{W}_l \right] \right\|_F^2 \leq \xi \left| \mathcal{E}_{\text{cond}} \right| \right)
\leq 2^{-\frac{\xi}{2} \log \frac{\det\left( \frac{A}{B} M^T M \right)}{(\sigma^2 + \epsilon \sigma_a^2)^l}}
\leq 2^{-\frac{\xi}{2} \log \left( \frac{\det\left( \frac{A}{B} M^T M \right)}{(\sigma^2 + \epsilon \sigma_a^2)^l} \right)}
\leq 2^{-\frac{\xi}{2} \log \left( \frac{(1 - \epsilon) \sigma_a^2 \sigma^2}{(\sigma^2 + \epsilon \sigma_a^2)^l} \right)}
\leq 2^{-\frac{\xi}{2} \log \left( \frac{(1 - \epsilon) \sigma_a^2 \sigma^2}{(\sigma^2 + \epsilon \sigma_a^2)^l} \right)}
\]
(41)

where (41) uses the fact that
\[
\left[ \begin{array}{c} W' \\ \frac{\sigma}{\sigma_a} I \end{array} \right] \left[ \begin{array}{c} W' \\ \frac{\sigma}{\sigma_a} I \end{array} \right]^T = \left[ \begin{array}{c} W_{[k]} \end{array} \right]^T \left[ \begin{array}{c} W_{[k]} \end{array} \right] + \left[ \begin{array}{c} W_{O} \end{array} \right]^T \left[ \begin{array}{c} W_{O} \end{array} \right]
\]
where $O$ denotes the matrix with elements all being zeros, and the fact that [65, Corollary 8.4.15], for positive semidefinite $B, D \in \mathbb{R}^{l \times l}$, $\det(B + D) \geq \det(B)$. By the union of events bound,
\[
P(\mathcal{E}_{s_1, s_2, ..., s_k} | \mathcal{E}_{\text{cond}})
\leq \sum_{\{s_1, ..., s_k\} = \{s_1, ..., s_k\}} P\left( \forall i \in Q_i \text{ such that } \frac{1}{n l} \left\| Y - [A_{s_1}', ..., A_{s_k}'] \left[ \tilde{W}_1, ..., \tilde{W}_l \right] \right\|_F^2 \leq \xi \left| \mathcal{E}_{\text{cond}} \right| \right)
\leq \sum_{\{s_1, ..., s_k\} = \{s_1, ..., s_k\}} \sum_{\tilde{W}_i \in Q_i} \cdots \sum_{\tilde{W}_i \in Q_i} P\left( \frac{1}{n l} \left\| Y - [A_{s_1}', ..., A_{s_k}'] \left[ \tilde{W}_1, ..., \tilde{W}_l \right] \right\|_F^2 \leq \xi \left| \mathcal{E}_{\text{cond}} \right| \right)
\leq k! \left( \prod_{i=1}^t |Q_i| \right) \cdot 2^{-\frac{\xi}{2} \log \left( \frac{(1 - \epsilon) \sigma_a^2 \sigma^2}{(\sigma^2 + \epsilon \sigma_a^2)^l} \right)} .
\]
Furthermore, conditioned on $\mathcal{E}_{\text{aux}}$, $\tilde{\rho}_i < \|w_i\|_2 + \epsilon$ for $i \in [l]$ and hence $|Q_i| \leq q_i(\|w_i\|_2 + \epsilon, \epsilon)$ by Lemma \ref{lemma:perturbation}. Thus,

$$P(\mathcal{E}_{s_1, s_2, \ldots, s_k} \cap \mathcal{E}_{\text{aux}}) \leq k! \cdot \left( \prod_{i=1}^l q_i(\|w_i\|_2 + \epsilon, \epsilon) \right) \cdot 2^{-\frac{m}{2} \log \frac{(1-\epsilon)\sigma^2_n}{\sigma^2_n + \epsilon^2}}.$$

Note that the probability upper-bound (43) 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 $\mathcal{T}$,

$$P(\mathcal{E}) \leq P(\mathcal{E}_{\text{aux}}^c) + P(\mathcal{E}_{1, 2, \ldots, k}^c) + \sum_{\mathcal{T} \subseteq [k]} \binom{m-k}{k-|\mathcal{T}|} \cdot k! \cdot \left( \prod_{i=1}^l q_i(\|w_i\|_2 + \epsilon, \epsilon) \right) \cdot 2^{-\frac{m}{2} \log \frac{(1-\epsilon)\sigma^2_n}{\sigma^2_n + \epsilon^2}}.$$  

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

$$\limsup_{m \to \infty} \frac{\log m}{n_m} < \frac{1}{2|\mathcal{T}|} \log \frac{(1-\epsilon)\sigma^2_n}{\sigma^2_n + \epsilon^2}.$$

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

Next, we prove Lemma 1. For $j \in \mathcal{S}$, $(b_{i,j} - D_{i,j})^2/\theta_j$ is a noncentral $\chi^2$ random variable. Its moment generating function is \ref{lemma:mgf} (for $t < 1/2$)

$$E[e^{t(b_{i,j} - D_{i,j})^2/\theta_j}] = \frac{e^{\frac{t^2}{1-2t}}}{(1-2t)^{\frac{1}{2}}}. \quad (45)$$

By changing variable $\frac{\theta_j t}{m} \to t$, we have

$$E[e^{\frac{t(b_{i,j} - D_{i,j})^2}{nt}}] = \frac{e^{\frac{t^2}{1-2t}}}{(1-2t)^{\frac{1}{2}}}. \quad (46)$$

For $j \in [l] \setminus \mathcal{S}$ with $D_j \equiv 0$, we additionally define $\theta_j = 0$. In this case,

$$E[e^{\frac{t(b_{i,j} - D_{i,j})^2}{nt}}] = E[e^{\frac{t^2}{nt}}] = \frac{e^{\frac{t^2}{nt}}}{(1-2t)^{\frac{1}{2}}}. \quad (47)$$
Define

\[ S_n \triangleq \frac{1}{m} \| B - D \|_F^2 = \frac{1}{l} \sum_{j=1}^{l} \frac{1}{n} \| b_j - D_j \|_2^2. \]  

(48)

Then, we have

\[
E[e^{tS_n}] = E[e^{\frac{t}{m} \| B - D \|_F^2}]
\]

(49)

\[
= E[e^{\frac{t}{2} \sum_{j=1}^{l} \frac{1}{n} \| b_j - D_j \|_2}]
\]

(50)

\[
= \prod_{j=1}^{l} E[e^{\frac{t}{2n} \| b_j - D_j \|_2}]
\]

(51)

\[
= \prod_{j=1}^{l} \frac{e^{\frac{t}{2n} \| b_j \|_2}}{e^{1 - 2\theta \frac{1}{m} t}}
\]

(52)
The Chernoff bound indicates that
\[
P(S_n \leq \gamma) \leq \min_{s > 0} e^{s \gamma} E[e^{-s S_n}]
\]  
\[
= \min_{s > 0} \prod_{j=1}^{l} \frac{e^{-\frac{\|b_j\|_2^2}{2}}}{(1 + \frac{20s}{nl})^\frac{1}{2}}
\]  
\[
= \min_{p < 0} e^{-p \gamma} \prod_{j=1}^{l} \frac{e^{-\frac{\|b_j\|_2^2}{2}}}{(1 - \frac{20p}{nl})^\frac{1}{2}}
\]  
\[
= \exp \left\{ \min_{p < 0} \left\{ -p \gamma + \sum_{j=1}^{l} \left[ \frac{\|b_j\|_2^2}{2} - \frac{n}{2} \log \left( 1 - \frac{2\theta_j p}{n} \right) \right] \right\} \right\}
\]  
\[
= \exp \left\{ \min_{p < 0} \left\{ -lp \gamma + \sum_{j=1}^{l} \left[ \frac{\|b_j\|_2^2}{2} - \frac{n}{2} \log \left( 1 - \frac{2\theta_j p}{n} \right) \right] \right\} \right\}
\]  
\[
= \exp \left\{ \min_{p < 0} \left\{ -lp \gamma - \sum_{j=1}^{l} \left[ \frac{\|b_j\|_2^2}{2} - \frac{n}{2} \log \prod_{j=1}^{l} \left( 1 - \frac{2\theta_j p}{n} \right) \right] \right\} \right\}
\]  
\[
\leq \exp \left\{ \min_{p < 0} \left\{ -lp \gamma - l \left( \prod_{j=1}^{l} \frac{\|b_j\|_2^2}{2} \right)^\frac{1}{l} - \frac{n}{2} \log \prod_{j=1}^{l} \left( 1 - \frac{2\theta_j p}{n} \right) \right\} \right\}
\]  
\[
= \exp \left\{ \min_{p < 0} \left\{ -lp \gamma - l \left( \prod_{j=1}^{l} \frac{\|b_j\|_2^2}{2} \right)^\frac{1}{l} - \frac{n l}{2} \log \left( \prod_{j=1}^{l} \left( 1 - \frac{2\theta_j p}{n} \right) \right)^\frac{1}{l} \right\} \right\}
\]  
\[
= \exp \left\{ \min_{p < 0} \left\{ -lp \gamma + lp - l \left( \prod_{j=1}^{l} \frac{\|b_j\|_2^2}{2} \right)^\frac{1}{l} - \frac{n l}{2} \log \left( \prod_{j=1}^{l} \left( 1 - \frac{2\theta_j p}{n} \right) \right)^\frac{1}{l} \right\} \right\}
\]  
\[
= \exp \left\{ \min_{p < 0} f(p) \right\}
\]  
\where (59) follows from the fact that the arithmetic mean is no smaller than the geometric mean. On the other hand, define the function
\[
g(p, \theta) = -lp \gamma + lp - l \left( \prod_{j=1}^{l} \frac{\|b_j\|_2^2}{2} \right)^\frac{1}{l} - \frac{n l}{2} \log \left( \prod_{j=1}^{l} \left( 1 - \frac{2\theta_j p}{n} \right) \right)^\frac{1}{l}.
\]  
(63)
Recall that \( \left( \prod_{j=1}^{f} \frac{1}{n} \| b_j \|_2^2 \right)^{1/2} = \alpha \). It can be readily seen that, for a fixed \( p < 0 \),

\[
    f(p) \leq \max_{\theta > 0} g(p, \theta)
\]

(64)

which is because there exists \( \theta > 0 \) such that \( 1 - \frac{2\theta_p}{n} = \left( \prod_{j=1}^{f} \left( 1 - \frac{2\theta_p}{n} \right) \right)^{1/2} \). Thus,

\[
    \min_{p < 0} f(p) \leq \min_{p < 0} \left( \max_{\theta > 0} g(p, \theta) \right).
\]

(65)

Our goal is to show

\[
    \min_{p < 0} \left( \max_{\theta > 0} g(p, \theta) \right) = \frac{-n l}{2} \log \frac{\alpha}{\gamma}
\]

(66)

which will lead to \( P(S_n \leq \gamma) \leq \frac{-n l}{2} \log \frac{\alpha}{\gamma} \) as desired. To this end, we first consider, for a fixed \( p \),

\[
    \frac{\partial g(p, \theta)}{\partial \theta} = \frac{pl(2\alpha - 2\theta + 1)}{(1 - 2\theta_p/n)^2}.
\]

(67)

By setting \( \frac{\partial g(p, \theta)}{\partial \theta} = 0 \), we have the stationary point \( \theta^* = \alpha + \frac{n}{2p} \). Examine the second derivative

\[
    \frac{\partial^2 g(p, \theta)}{\partial \theta^2} \bigg|_{\theta = \theta^*} = \frac{(1 - \frac{2\theta_p}{n})(-2\theta + \alpha)(-2\theta - 4\alpha - 1)}{(1 - \frac{2\theta_p}{n})^4} \bigg|_{\theta = \theta^*} = \frac{(-2\gamma^2)}{(2\theta_p/n)^2} < 0.
\]

(68)

Due to the constraint \( \theta > 0 \), we have

\[
    \max_{\theta > 0} g(p, \theta) = \begin{cases} 
    -pl\gamma - \frac{n l}{2} - \frac{n l}{2} \log(-\frac{2\alpha}{n}) & \text{if } p \leq -\frac{n}{2\alpha}, \\
    pl(\alpha - \gamma) & \text{if } -\frac{n}{2\alpha} < p < 0.
    \end{cases}
\]

(69)

Next, we calculate \( \min_{p < 0} (\max_{\theta > 0} g(p, \theta)) \). First,

\[
    \min_{-\frac{n}{2\alpha} < p < 0} \left( \max_{\theta > 0} g(p, \theta) \right) = \min_{-\frac{n}{2\alpha} < p < 0} pl(\alpha - \gamma) = \frac{-n l}{2} \left( 1 - \frac{\gamma}{\alpha} \right).
\]

(70)

Then, to figure out \( \min_{p \leq -\frac{n}{2\alpha}} (\max_{\theta > 0} g(p, \theta)) \), we compute

\[
    \frac{\partial}{\partial p} \max_{\theta > 0} g(p, \theta) = \frac{\partial}{\partial p} \left( -pl\gamma - \frac{n l}{2} - \frac{n l}{2} \log(-\frac{2\alpha}{n}) \right) = -l\gamma - \frac{n l}{2p} = \text{set 0}
\]

(71)

which gives the stationary point \( p^* = -\frac{n}{2\gamma} \). Check for the second derivative,

\[
    \frac{\partial^2}{\partial p^2} \max_{\theta > 0} g(p, \theta) \bigg|_{p=p^*} = \frac{n l}{2(p^*)^2} > 0.
\]

(72)

Therefore, \( p^* = -\frac{n}{2\gamma} \leq -\frac{n}{2\alpha} \) is the minimizer. As a result,

\[
    \min_{p \leq -\frac{n}{2\alpha}} \left( \max_{\theta > 0} g(p, \theta) \right) = -pl\gamma - \frac{n l}{2} - \frac{n l}{2} \log(-\frac{2\alpha}{n}) \bigg|_{p=p^*} = -\frac{n l}{2} \log \frac{\alpha}{\gamma}.
\]

(73)
Overall,

$$\min_{p<0} \left( \max_{\theta>0} g(p, \theta) \right) = \min \left( -\frac{n}{2} \left( 1 - \frac{\gamma}{\alpha} \right), -\frac{n}{2} \log \frac{\alpha}{\gamma} \right)$$ \quad (74)

Using the fact that $0 \leq 1 - \frac{1}{x} \leq \log x$ for $x > 1$, we finally have

$$\min_{p<0} \left( \max_{\theta>0} g(p, \theta) \right) = -\frac{n}{2} \log \frac{\alpha}{\gamma}. \quad (75)$$

Therefore,

$$P(S_n \leq \gamma) \leq \exp \left\{ \min_{p<0} f(p) \right\} \quad (76)$$

$$\leq \min_{p<0} \left( \max_{\theta>0} g(p, \theta) \right) \quad (77)$$

$$= 2^{-\frac{n}{2} \log \frac{\prod_{i=1}^{q} (b_i \tilde{b}_i^\top)^{\frac{1}{2}}}{\gamma}} \quad (78)$$

$$= 2^{-\frac{n}{2} \log \frac{\prod_{i=1}^{q} \left( \frac{1}{\gamma} \alpha \gamma \right)^{\frac{1}{2}}}{\gamma}}. \quad (79)$$

The remaining task is to prove Lemma 3. Let $\sigma_{b,1}^2 \geq \cdots \geq \sigma_{b,q}^2$ be the $q$ eigenvalues of $B^\top B$, where $\sigma_{b,q}^2 = \sigma_{b}^2$. The eigen-decomposition states that there exists a unitary matrix $J \in \mathbb{R}^{q \times q}$, such that $B^\top B = JGGJ^\top$, where $G \in \mathbb{R}^{q \times q}$ is a diagonal matrix with the $i$th diagonal element being $\sigma_{b,i}$. Thus,
\[ D^\top B^\top BD = D^\top JGG^\top J^\top D = FT, \] where \( F = D^\top JG \) and \( T = F^\top \). Note that

\[ \det((BD)^\top BD) = \det(FT) \]
\[ = \sum_{1 \leq j_1 < \cdots < j_r \leq q} \det \begin{bmatrix} f_{1,j_1} & \cdots & f_{1,j_r} \\ \vdots & & \vdots \\ f_{r,j_1} & \cdots & f_{r,j_r} \end{bmatrix} \det \begin{bmatrix} t_{j_1,1} & \cdots & t_{j_1,r} \\ \vdots & & \vdots \\ t_{k_r,1} & \cdots & t_{k_r,r} \end{bmatrix} \]
\[ = \sum_{1 \leq j_1 < \cdots < j_r \leq q} \left( \det \begin{bmatrix} f_{1,j_1} & \cdots & f_{1,j_r} \\ \vdots & & \vdots \\ f_{r,j_1} & \cdots & f_{r,j_r} \end{bmatrix} \right)^2 \]
\[ = \sum_{1 \leq j_1 < \cdots < j_r \leq q} \left( \det \begin{bmatrix} [D^\top J]_{1,j_1} & \cdots & [D^\top J]_{1,j_r} \\ \vdots & & \vdots \\ [D^\top J]_{r,j_1} & \cdots & [D^\top J]_{r,j_r} \end{bmatrix} \right)^2 \]
\[ \geq (\sigma_0^2)^r \sum_{1 \leq j_1 < \cdots < j_r \leq q} \left( \det \begin{bmatrix} [D^\top J]_{1,j_1} & \cdots & [D^\top J]_{1,j_r} \\ \vdots & & \vdots \\ [D^\top J]_{r,j_1} & \cdots & [D^\top J]_{r,j_r} \end{bmatrix} \right)^2 \]
\[ = (\sigma_0^2)^r \det(D^\top J^\top JD) \]
\[ = (\sigma_0^2)^r \det(D^\top D) \]

where \(80\) is due to the Binet-Cauchy formula \(67\).

**APPENDIX B**

**PROOF OF THEOREM 2**

To establish this theorem, we prove the following equivalent statement:

If there exist a sequence of matrices \(\{A^{(m)}\}_{m=k}^\infty, A^{(m)} \in \mathbb{R}^{n_m \times m}\), and a sequence of support recovery maps \(\{d^{(m)}\}_{m=k}^\infty, d^{(m)} : \mathbb{R}^{n_m} \mapsto 2^{\{1,2,\ldots,m\}}\), such that

\[ \frac{1}{n_mm} \|A^{(m)}\|_F^2 \leq \sigma_a^2 \]

and

\[ \lim_{m \to \infty} P\{d^{(m)}(A^{(m)}X + Z) \neq \text{supp}(X(W,S))\} = 0 \]
then
\[ \limsup_{m \to \infty} \frac{\log m}{n_m} \leq c(W). \]

For any \( T \subseteq [k] \), denote the tuple of random variables \((S_l : l \in T)\) by \( S(T) \). For notation simplicity, let \( \overline{P}^{(m)}_e \triangleq \mathbb{P}\{d^{(m)}(A^{(m)}X + Z) \neq \text{supp}(X(W,S))\} \). From Fano’s inequality \[63\], we have
\[
H(S(T)|Y) \leq H(S_1, \ldots, S_k|Y) \\
\leq \log k! + H(\{S_1, \ldots, S_k\}|Y) \\
\leq \log k! + \overline{P}^{(m)}_e \log \left( \frac{m}{k} \right) + 1. \tag{81}
\]

On the other hand,
\[
H(S(T)|S(T^c)) = \log \left( \prod_{q=0}^{T-1} (m - (k - |T|) - q) \right) \\
= |T| \log m - n\epsilon_{1,n} \tag{82}
\]
where \( T^c \triangleq [k]\setminus T \) and
\[
\epsilon_{1,n} \triangleq \frac{1}{n} \log \left( \frac{m^{(|T|)}}{\prod_{q=0}^{T-1} (m - (k - |T|) - q)} \right)
\]
which tends to zero as \( n \to \infty \). Hence, combining \eqref{81} and \eqref{82}, we have
\[
|T| \log m = H(S(T)|S(T^c)) + n\epsilon_{1,n} \\
= I(S(T); Y|S(T^c)) + H(S(T)|Y, S(T^c)) + n\epsilon_{1,n} \\
\leq I(S(T); Y|S(T^c)) + H(S(T)|Y) + n\epsilon_{1,n} \tag{83}
\]
\[
\leq I(S(T); Y|S(T^c)) + \log k! + \overline{P}^{(m)}_e \log \left( \frac{m}{k} \right) + 1 + n\epsilon_{1,n} \\
= \sum_{i=1}^{n} I(Y_i; S(T)|Y_{[i-1]}, S(T^c)) + \log k! + \overline{P}^{(m)}_e \log \left( \frac{m}{k} \right) + 1 + n\epsilon_{1,n} \\
= \sum_{i=1}^{n} \left( h(Y_i|Y_{[i-1]}, S(T^c)) - h(Y_i|Y_{[i-1]}, S([k])) \right) + \log k! + \overline{P}^{(m)}_e \log \left( \frac{m}{k} \right) + 1 + n\epsilon_{1,n} \\
\leq \sum_{i=1}^{n} \left( h(Y_i|S(T^c)) - h(Y_i|S_1, \ldots, S_k) \right) + \log k! + \overline{P}^{(m)}_e \log \left( \frac{m}{k} \right) + 1 + n\epsilon_{1,n} \tag{84}
\]
\[
= \sum_{i=1}^{n} \left( h(Y_i|S(T^c)) - h(Z_i) \right) + \log k! + \overline{P}^{(m)}_e \log \left( \frac{m}{k} \right) + 1 + n\epsilon_{1,n} \tag{85}
\]
where $\mathbf{Y}_{[i-1]}$ denotes the set $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_{i-1}\}$. To explain some intermediate steps, (83) follows from the fact that conditioning reduces entropy. (84) holds because $\mathbf{Y}_i$ is independent of $\mathbf{Y}_{[i-1]}$ when conditioned on $S([k])$, and (85) follows since the measurement matrix is fixed and $\mathbf{Z}_i$ is independent of $(S_1, \ldots, S_k)$.

Consider

$$h(\mathbf{Y} | S(T^c))$$

$$= h \left( A_i, S([k]) | W + \mathbf{Z}_i \right)$$

$$= h \left( A_i, S(T) | W_T + \mathbf{Z}_i \right)$$

$$\leq h \left( A_i, S(T) | W_T + \mathbf{Z}_i \right)$$

$$\leq \frac{1}{2} \log \left( (2\pi)^{\frac{1}{2}} \cdot \det \left( [E(\mathbf{A}_i, S(T) | W_T + \mathbf{Z}_i)] - E[A_i, S(T) | W_T + \mathbf{Z}_i] \right) \right)$$

$$\leq \frac{1}{2} \log \left( (2\pi)^{\frac{1}{2}} \cdot \det \left( W_T \left( [E[A_i, S(T)] A_i, S(T)] - E[A_i, S(T)] [E[A_i, S(T)]] \right) W_T + \sigma^2 I \right) \right)$$

where (86) follows from the fact that with the same covariance the Gaussian random vector maximizes the entropy (63), and the randomness in $A_i, S(T)$ is due to the randomness of the index set $S(T)$. Note that

$$E[A_i, S(T)] = \frac{1}{m} \sum_{p=1}^{m} a_{i,p} 1^T.$$  \hspace{1cm} (88)

Meanwhile

$$E[A_{i,S(T)}^T A_{i,S(T)}] = \frac{1}{m} \sum_{p=1}^{m} a_{i,p}^2 I + \frac{1}{m(m-1)} \sum_{p=1}^{m} \sum_{q=1}^{m} a_{i,p} a_{i,q} (1 \cdot 1^T - I).$$  \hspace{1cm} (89)

Thus

$$E[A_{i,S(T)}^T A_{i,S(T)}] - E[A_{i,S(T)}] [E[A_{i,S(T)}]]$$

$$= \frac{1}{m} \sum_{p=1}^{m} a_{i,p}^2 I + \frac{1}{m(m-1)} \left( \sum_{p=1}^{m} a_{i,p} \right)^2 (1 \cdot 1^T - I) - \frac{1}{m^2} \left( \sum_{p=1}^{m} a_{i,p} \right)^2 1 \cdot 1^T$$

$$= \frac{1}{m} \sum_{p=1}^{m} a_{i,p}^2 I + \frac{1}{m(m-1)} \left( \sum_{p=1}^{m} a_{i,p} \right)^2 - \frac{1}{m^2} \left( \sum_{p=1}^{m} a_{i,p} \right)^2 1 \cdot 1^T$$

$$= \frac{1}{m} \sum_{p=1}^{m} a_{i,p}^2 \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) + \left( \sum_{p=1}^{m} a_{i,p} \right)^2 \left( \frac{1}{m(m-1)} (1 \cdot 1^T - I) - \frac{1}{m^2} 1 \cdot 1^T \right).$$  \hspace{1cm} (90)
Note that \(\frac{1}{m(m-1)}(1 \cdot 1^T - I) - \frac{1}{m} 1 \cdot 1^T = \frac{1}{m(m-1)} 1 \cdot 1^T - \frac{1}{m(m-1)} I\) is negative semidefinite for sufficiently large \(m\), and so is \(W_T^\top \left(\frac{1}{m(m-1)} 1 \cdot 1^T - \frac{1}{m-1} I\right) W_T\). Hence

\[
\det \left( W_T^\top \left( E[A_i^\top, S(T)] A_i^\top, S(T) - E[A_i^\top, S(T)] E[A_i, S(T)] \right) W_T + \sigma_z^2 I \right)
\]

\[
\leq \det \left( \frac{1}{m} \sum_{p=1}^{m} a_{i,p}^2 W_T^\top \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) W_T + \sigma_z^2 I \right)
\]

as a result of [65, Corollary 8.4.15]. Therefore

\[
|\mathcal{T}| \log m
\]

\[
\leq \sum_{i=1}^{n} \left[ \frac{1}{2} \log \left( (2\pi e)^l \cdot \det \left( \frac{1}{m} \sum_{p=1}^{m} a_{i,p}^2 W_T^\top \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) W_T + \sigma_z^2 I \right) \right) \right] - \frac{1}{2} \log \left( (2\pi e \sigma_z^2)^l \right)
\]

\[
+ \log k! + \mathcal{P}_e^{(m)} \log \left( \frac{m}{k} \right) + 1 + n \epsilon_{1,n}
\]

\[
= \sum_{i=1}^{n} \frac{1}{2} \log \det \left( \frac{1}{m \sigma_z^2} \sum_{i=1}^{m} \sum_{p=1}^{m} a_{i,p}^2 W_T^\top \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) W_T + I \right)
\]

\[
+ \log k! + \mathcal{P}_e^{(m)} \log \left( \frac{m}{k} \right) + 1 + n \epsilon_{1,n}
\]

\[
\leq \sum_{i=1}^{n} \frac{1}{2} \log \det \left( \frac{\sigma_z^2}{2} W_T^\top \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) W_T + I \right)
\]

\[
+ \log k! + \mathcal{P}_e^{(m)} \log \left( \frac{m}{k} \right) + 1 + n \epsilon_{1,n}
\]

\[
\leq \sum_{i=1}^{n} \frac{1}{2} \log \det \left( \frac{\sigma_z^2}{\sigma_z^2} W_T^\top \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) W_T + I \right)
\]

\[
+ \log k! + \mathcal{P}_e^{(m)} \log \left( \frac{m}{k} \right) + 1 + n \epsilon_{1,n}
\]

(91)

Then, we have

\[
\limsup_{m \to \infty} \frac{(1 - k \mathcal{P}_e^{(m)}/|\mathcal{T}|) \log m}{n_m} - \frac{\log k! + n_m \epsilon_{1,n} + 1}{|\mathcal{T}| n_m}
\]

\[
\leq \limsup_{m \to \infty} \frac{1}{2|\mathcal{T}|} \log \det \left( \frac{\sigma_z^2}{\sigma_z^2} W_T^\top \left( I - \frac{1}{m-1} (1 \cdot 1^T - I) \right) W_T + I \right)
\]

\[
= \frac{1}{2|\mathcal{T}|} \log \det \left( \frac{\sigma_z^2}{\sigma_z^2} W_T^\top W_T + I \right)
\]

(92)

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

\[
\limsup_{m \to \infty} \frac{\log m}{n_m} \leq \frac{1}{2|\mathcal{T}|} \log \det \left( \frac{\sigma_z^2}{\sigma_z^2} W_T^\top W_T + I \right)
\]

(93)

for all \(\mathcal{T} \subseteq [k]\). This completes the proof of Theorem 2.
APPENDIX C
PROOF OF COROLLARY \([3]\)

To justify this corollary, we need to show

\[
\min_{T \subseteq [k]} \left[ \frac{1}{2|T|} \log \det \left( I + \frac{\sigma^2}{\sigma_z^2} W_T^T W_T \right) \right] = 2 \cdot \frac{1}{2k} \log \left( 1 + k \cdot \frac{\sigma^2}{\sigma_z^2} \right).
\]

To begin with, recall that \(k\) is even, and \(w_2\) is defined in (16). For a given \(T \subseteq [k]\), let \(T_1 = T \cap \left[ \frac{k}{2} \right], T_2 = T \setminus T_1, t = |T|, t_1 = |T_1|, \) and \(t_2 = |T_2|\). One can obtain

\[
W_T^T W_T = \begin{bmatrix} t & t_1 - t_2 \\ t_1 - t_2 & t \end{bmatrix}.
\]

Let \(\alpha \triangleq \frac{\sigma^2}{\sigma_z^2}\) for notational simplicity. Thus

\[
\frac{1}{2|T|} \log \det (I + \alpha W_T^T W_T) = \frac{1}{2t} \log \begin{bmatrix} 1 + \alpha t & \alpha(t_1 - t_2) \\ \alpha(t_1 - t_2) & 1 + \alpha t \end{bmatrix} = \frac{1}{2t} \log \left( 1 + 2\alpha t + 4\alpha^2 t_1 t_2 \right)
\]

where we use the fact that \(t = t_1 + t_2\). Note that, for a given \(t \in [k]\),

\[
\min_{T: T \subseteq [k], |T| = \frac{k}{2}} \frac{1}{2t} \log \left( 1 + 2\alpha t + 4\alpha^2 t_1 t_2 \right) = \frac{1}{2t} \log(1 + 2\alpha t)
\]

and

\[
\min_{T: T \subseteq [k], |T| = t \geq \frac{k}{2}} \frac{1}{2t} \log \left( 1 + 2\alpha t + 4\alpha^2 t_1 t_2 \right) = \frac{1}{2t} \log \left( 1 + 2\alpha t + 4\alpha^2 \frac{k}{2} \left( t - \frac{k}{2} \right) \right)
\]

where we use the implicit constraints that \(t_1, t_2 \leq \frac{k}{2}\). Then, the problem becomes evaluating

\[
\min_{t: t \in [k]} f(t), \text{ where } f(t) = \begin{cases} \frac{1}{2t} \log(1 + 2\alpha t) & \text{if } 0 < t \leq \frac{k}{2}, \\
\frac{1}{2t} \log \left( 1 + 2\alpha t + 4\alpha^2 \frac{k}{2} \left( t - \frac{k}{2} \right) \right) & \text{if } \frac{k}{2} + 1 \leq t \leq k.
\end{cases}
\]

First, it can be readily seen that \(\min_{t: t \in [k]} f(t) = \frac{1}{k} \log(1 + \alpha k)\). Next, we consider the function

\[
g(t) \triangleq \frac{\log(\beta_1 + \beta_2 t)}{2t}
\]

where \(\beta_1 \triangleq 1 - \alpha^2 k^2\) and \(\beta_2 \triangleq 2\alpha(1 + \alpha k)\) for \(t \in [\frac{k}{2}, k]\). Note that

\[
\frac{\partial g(t)}{\partial t} = \frac{1}{\beta_1 + \beta_2 t} - \frac{\log(\beta_1 + \beta_2 t)}{t^2}.
\]

\(^6\)For the purpose of analysis, the base of logarithm is not important, as long as all of them are consistent. Here, we choose natural logarithm to simplify the calculation.
To obtain stationary points, we solve

\[ 1 - \frac{\beta_1}{\beta_1 + \beta_2 t} + \log \frac{1}{\beta_1 + \beta_2 t} = 0, \quad t \neq 0 \]  

which is equivalent to

\[ 1 + v(t) = \beta_1 e^{v(t)}, \quad t \neq 0 \]  

where \( v(t) \triangleq \log \frac{1}{\beta_1 + \beta_2 t} \). Note that \( \beta_1 < 1 \). We will consider three different cases. The first case is \( 0 < \beta_1 < 1 \). By comparing the curves of \( 1 + v \) and \( \beta_1 e^v \) as functions of \( v \), we see that there are two solutions with opposite signs, namely \( v_1 < 0 \) and \( v_2 > 0 \), to (100). Note that

\[ g \left( \frac{k}{2} \right) = g(k) = \frac{1}{k} \log(1 + \alpha k). \]

Meanwhile, \( v(t) \) is monotonically decreasing on \( \left[ \frac{k}{2}, k \right] \), and

\[ v(k) = \log \frac{1}{(1 + \alpha k)^2} < v \left( \frac{k}{2} \right) = \log \frac{1}{1 + \alpha k} < 0. \]

Therefore, it is evident that \( v(k) < v_1 < v \left( \frac{k}{2} \right) < v_2 \). Further, it can be readily seen that

\[ \frac{\partial g(t)}{\partial t} \bigg|_{t=\frac{k}{2}} = \frac{1 + v(t) - \beta_1 e^{v(t)}}{t^2} \bigg|_{t=\frac{k}{2}} > 0 \]

and

\[ \frac{\partial g(t)}{\partial t} \bigg|_{t=k} = \frac{1 + v(t) - \beta_1 e^{v(t)}}{t^2} \bigg|_{t=k} < 0. \]

In summary, \( g(t) \) is increasing at \( t = \frac{k}{2} \) and decreasing at \( t = k \), it takes the same value at these two points, and there exists only one stationary point in between. These observations lead to the conclusion that

\[ \min_{t: t \in \left[ \frac{k}{2} \right]} f(t) = f(k) = \frac{1}{k} \log(1 + \alpha k). \]

To analyze the cases for \( \beta_1 = 0 \) and \( \beta_1 < 0 \), we only need to note that there is only one solution \( v_1 \) to (100). Thus, similar argument applies to these two cases.
REFERENCES

[1] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

[2] R. Zdunek and A. Cichocki, “Improved m-focuss algorithm with overlapping blocks for locally smooth sparse signals,” *IEEE Transactions on Signal Processing*, 2008.

[3] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. J. Candes, “Compressive sampling,” *Proceedings of the Int. Congress of Mathematicians*, pp. 1433–1452, 2006.

[5] I. Gorodnitsky and B. Rao, “Sparse signal reconstruction from limited data using focuss: A re-weighted norm minimization algorithm,” *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[6] I. F. Gorodnitsky, J. S. George, and B. D. Rao, “Neuromagnetic source imaging with focuss: A recursive weighted minimum norm algorithm,” *J. Electroencephalog. Clinical Neurophysiol.*, vol. 95, pp. 231–251, 1995.

[7] B. D. Jeffs, “Sparse inverse solution methods for signal and image processing applications,” *Proc. ICASSP*, pp. 1885–1888, 1998.

[8] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. G. Baraniuk, “Single-pixel imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, pp. 83–91, 2008.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 1–18, 2009.

[10] S. D. Cabrera and T. W. Parks, “Extrapolation and spectral estimation with iterative weighted norm modification,” *IEEE Transactions on Acoust., Speech, Signal Process.*, vol. 4, pp. 842–851, 1991.

[11] Y. Jin and B. D. Rao, “Algorithms for robust linear regression by exploiting the connection to sparse signal recovery,” *ICASSP*, 2010.

[12] W. C. Chu, *Speech coding algorithms*. Wiley-Interscience, 2003.

[13] S. F. Cotter and B. D. Rao, “Sparse channel estimation via matching pursuit with application to equalization,” *IEEE Transactions on on Communications*, vol. 50, pp. 374–377, 2002.

[14] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, “Compressed channel sensing,” *Proc. of CISS*, 2008.

[15] D. L. Duttweiler, “Proportionate normalized least-mean-squares adaptation in echo cancelers,” *IEEE Transactions on Acoust., Speech, Signal Process.*, vol. 8, pp. 508–518, 2000.

[16] B. D. Rao and B. Song, “Adaptive filtering algorithms for promoting sparsity,” *ICASSP*, pp. 361–364, 2003.

[17] H. Garudadri, P. K. Baheti, and S. Majumdar, “Low complexity sensors for body area networks,” *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments*, 2010.

[18] D. Guo, “Neighbor discovery in ad hoc networks as a compressed sensing problem,” *Presented at Information Theory and Application Workshop, UCSD*, 2009.

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

[20] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” *27th Annual Asilomar Conf. Sig. Sys. Comp.*, 1993.

[21] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[22] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIREV*, vol. 43, no. 1, pp. 129–159, 2001.
[23] E. J. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted $\ell_1$ minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[24] R. Chartrand and W. Yin, “Iteratively reweighted algorithms for compressive sensing,” *ICASSP*, 2008.

[25] M. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *JMLR*, 2001.

[26] D. Wipf and B. D. Rao, “Sparse bayesian learning for basis selection,” *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, August 2004.

[27] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 50, pp. 1417–1428, 2002.

[28] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *preprint*, 2008.

[29] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing: Closing the gap between performance and complexity,” *preprint*, 2008.

[30] D. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.

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

[32] E. J. Candes, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, 2006.

[33] J. A. Tropp, “Greedy is good: Algorithmic results for sparse approximation,” *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[34] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory preprint*, 2007.

[35] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, “Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit,” *preprint*, 2006.

[36] M. Wainwright, “Sharp thresholds for high-dimensional and noisy sparsity recovery using $\ell_1$-constrained quadratic programming (lasso),” *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

[37] P. Zhao and B. Yu, “On model selection consistency of lasso,” *Journal of Machine Learning Research*, vol. 7, pp. 2541–2563, 2006.

[38] M. Wainwright, “Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting,” *IEEE Transactions on Information Theory*, Dec 2009.

[39] A. K. Fletcher, S. Rangan, and V. K. Goyal, “Necessary and sufficient conditions for sparsity pattern recovery,” *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5758–5772, Dec 2009.

[40] W. Wang, M. J. Wainwright, and K. Ramchandran, “Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement,” *Proc. of ISIT*, pp. 2197–2201, 2008.

[41] M. Akçakaya and V. Tarokh, “Shannon theoretic limits on noisy compressive sampling,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 492–504, 2010.

[42] Y. Jin, Y.-H. Kim, and B. D. Rao, “Support recovery of sparse signals,” *arXiv:1003.0888v1[cs.IT]*, 2010.

[43] D. Wipf and S. Nagarajan, “A unified bayesian framework for meg/eeg source imaging,” *NeuroImage*, pp. 947–966, 2008.

[44] A. Cichocki, “Blind source separation: New tools for extraction of source signals and denoising,” *SPIE*, pp. 11–25, 2005.

[45] G. Obozinski, M. J. Wainwright, and M. I. Jordan, “Support union recovery in high-dimensional multivariate regression,” *The Annals of Statistics*, vol. 39, no. 1, pp. 1–47.
[46] D. Malioutov, M. Cetin, and A. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[47] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Simultaneous sparse approximation via greedy pursuit,” *ICASSP*, 2005.

[48] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[49] D. P. Wipf and B. D. Rao, “An empirical bayesian strategy for solving the simultaneous sparse approximation problem,” *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, July 2007.

[50] Z. Zhang and B. D. Rao, “Sparse signal recovery in the presence of correlated multiple measurement vectors,” *ICASSP*, 2010.

[51] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 4634–4643, 2006.

[52] Y. C. Eldar and H. Rauhut, “Average case analysis of multichannel sparse recovery using convex relaxation,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519.

[53] G. Tang and A. Nehorai, “Performance analysis for sparse support recovery,” *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[54] S. Sarvotham, D. Baron, and R. G. Baraniuk, “Measurements vs. bits: Compressed sensing meets information theory,” *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing*, 2006.

[55] J. Tropp, “Just relax: Convex programming methods for identifying sparse signal in noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 3, 2006.

[56] Y. Jin, Y.-H. Kim, and B. D. Rao, “Support recovery of sparse signals,” arXiv:1003.0888v1, 2010.

[57] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of mimo channels,” *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, 2003.

[58] Y. Jin and B. D. Rao, “Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory,” *ICASSP*, 2008.

[59] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[60] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Transactions on Information Theory*, vol. 49, no. 4, 2003.

[61] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[62] G. Reeves and M. Gastpar, “Sampling bounds for sparse support recovery in the presence of noise,” *Proc. of International Symposium on Information Theory*, pp. 2187–2191, 2008.

[63] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.

[64] J. W. Silverstein, “The smallest eigenvalue of a large dimensional wishart matrix,” *Annals of Probability*, vol. 13, pp. 1364–1368, 1985.

[65] D. S. Bernstein, *Matrix Mathematics: Theory, facts, and formulas (second edition)*. Princeton University Press, 2009.

[66] H. O. Lancaster, *The \( \chi^2 \) distribution*. Wiley, 1969.

[67] H. Dym, *Linear Algebra in Action (Graduate Studies in Mathematics, Volume 78)*. American Mathematical Society, 2007.