An Information-Theoretic Study for Joint Sparsity Pattern Recovery With Different Sensing Matrices

Sangjun Park, Nam Yul Yu, *Member, IEEE*, and Heung-No Lee, *Senior Member, IEEE*

***Abstract*— In this paper, we study a support set reconstruction problem for multiple measurement vectors (MMV) with different sensing matrices, where the signals of interest are assumed to be jointly sparse and each signal is sampled by its own sensing matrix in the presence of noise. Using mathematical tools, we develop upper and lower bounds of the failure probability of the support set reconstruction in terms of the sparsity, the ambient dimension, the minimum signal-to-noise ratio, the number of measurement vectors, and the number of measurements. These bounds can be used to provide guidelines for determining the system parameters for various compressed sensing applications with noisy MMV with different sensing matrices. Based on the bounds, we develop necessary and sufficient conditions for reliable support set reconstruction. We interpret these conditions to provide theoretical explanations regarding the benefits of taking more measurement vectors. We then compare our sufficient condition with the existing results for noisy MMV with the same sensing matrix. As a result, we show that noisy MMV with different sensing matrices may require fewer measurements for reliable support set reconstruction, under a sublinear sparsity regime in a low noise-level scenario.**

***Index Terms*— Compressed sensing, support set reconstruction, joint sparsity structure, multiple measurement vectors model.**

I. Introduction

CONVENTIONALLY, signals sensed from sensors such as microphones and imaging devices are sampled following the Shannon and Nyquist sampling theory [1] at a rate higher than twice the maximum frequency for signal reconstruction. As the number of samples decided by this theory is often large, the samples go through a compression stage before being stored. Therefore, taking numerous samples, where most of them will be discarded in this stage, is inefficient. Because compressed sensing (CS) [2]–[7] removes the inefficiency, CS has been applied in various areas such as wireless communications [8]–[11], spectrometers [12], multiple input multiple output (MIMO) radars [13], magnetic resonance imaging [14], and imaging/signal processing [15]–[17].

The CS theory states that signals that are sparsely representable in a certain basis are compressively sampled and reconstructed from what we thought is incomplete in information. Let $\mathbf{x} \in \mathbb{R}^N$ be a *K-sparse* vector with a support set $\mathcal{I} := \{i \mid x(i) \neq 0\}$ whose indices indicate the positions of the nonzero coefficients of $\mathbf{x}$. It is compressively sampled by a model called *single measurement vector* (*SMV*) as follows:

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is a (noisy) measurement vector, $\mathbf{F} \in \mathbb{R}^{M \times N}$ is a sensing matrix, and $\mathbf{n} \in \mathbb{R}^M$ is a noise vector, whose elements are independent and identically distributed (i.i.d) Gaussian with a zero mean and a σ^2 variance. Once the support set is correctly reconstructed, then (1) can be well-posed, which allows us to obtain an accurate estimate of $\mathbf{x}$ using the least square approach. We thus aim to focus on the support set reconstruction problem.

A. Information-Theoretic Works for CS With SMV

Works [18]–[23] have studied the support set reconstruction problem from an information-theoretic perspective. For reliable support set reconstruction, sufficient and necessary conditions were established in the linear and sublinear sparsity regimes.

For support set reconstruction, Wainwright [18] used the union bound to establish a sufficient condition on the number of measurements M for a maximum likelihood (ML) decoder and used Fano's inequality [24] to obtain a necessary condition on M. This ML decoder was analyzed by Fletcher *et al.* [19] to establish a necessary condition on M. Aeron *et al.* [20] used Fano's inequality to form necessary conditions on both M and σ^2. Then, they used the union bound to obtain sufficient conditions on both M and σ^2 for their sub-optimal decoder. Akcakaya and Tarokh [21] used the union and the large deviation bounds based on empirical entropies to get sufficient conditions on M for their joint typical decoder. They used the converse of the channel coding theorem to get necessary conditions on M. Scarlett *et al.* [22] extended this decoder [21] with the assumption that the distribution of the support set is provided. For a uniform distribution case, their necessary and sufficient conditions are equivalent to those of [21]. However, they are better for a non-uniform distribution case. Scarlett and Cevher [23] linked the support set reconstruction with the problem of coding over a mixed channel, where information spectrum methods were used to obtain necessary and sufficient conditions on M.

B. Information-Theoretic Works for CS With MMV

CS has many applications in wireless sensor networks (WSNs) [8]–[11] and MIMO radars [13]. In these

Manuscript received April 1, 2016; revised November 14, 2016; accepted January 17, 2017. Date of publication May 15, 2017; date of current version August 16, 2017. This work was supported by the National Research Foundation of Korea through the South Korean Government under Grant NRF-2015R1A2A1A05001826. This paper was presented at the 2012 International Symposium on Information Theory.

The authors are with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea (e-mail: sjpark1@gist.ac.kr; nyyu@gist.ac.kr; heungno@gist.ac.kr).

Communicated by H. Rauhut, Associate Editor for Signal Processing.
Digital Object Identifier 10.1109/TIT.2017.2704111

applications, the signals of interest $\mathbf{x}^s \in \mathbb{R}^N,\ s = 1, 2, \cdots, S$ are often modeled as *jointly K-sparse vectors*, implying that $\mathcal{I} = \mathcal{I}^1 = \mathcal{I}^2 \cdots = \mathcal{I}^S$, where $\mathcal{I}^s$ is the support set of $\mathbf{x}^s$ and $|\mathcal{I}| = K$, which is referred to as a *joint sparsity structure*.

There are two models for sampling jointly *K-sparse* vectors. The first model is called *multiple measurement vectors* (MMV) with *the same sensing matrix* [25], in which they are sampled by the same sensing matrix. The second model is named as MMV with *different sensing matrices* [8], [9], in which each one is sampled by its own sensing matrix.

The authors of [26]–[28] have conducted information-theoretic research to obtain conditions under which the support set of both the models was reconstructed with a high probability. In noisy MMV with the same sensing matrix, Tang and Nehorai [26] used the hypothesis theory to obtain necessary and sufficient conditions on both the number of measurements M and the number of measurement vectors S, and proved that the success probability of the support set reconstruction increases with S, if $M = \Omega\left(K \log \frac{N}{K}\right)$. Jin and Rao [27] exploited the communication theory to establish necessary and sufficient conditions on M and demonstrated the benefits of the joint sparsity structure based on their conditions. A detailed comparison between the results of our paper and those of [27] will be presented in Section IV. Finally, Duarte *et al.* [28] studied noiseless MMV with different sensing matrices, and formed necessary and sufficient conditions on M. However, it is difficult to apply the conditions to noisy MMV with different sensing matrices.

Meanwhile, works [8], [29], [30] have presented conditions of practical algorithms for a reliable support set reconstruction. In noiseless MMV with the same sensing matrix, Blanchard and Davies [30] obtained conditions for a reliable reconstruction from rank aware orthogonal matching pursuit (OMP). In noisy MMV with the same sensing matrix, Kim *et al.* [29] created compressive MUSIC, and presented its sufficient condition. In noiseless MMV with different sensing matrices, Baron *et al.* [8] produced trivial pursuit (TP) and distributed compressed sensing-simultaneous OMP (DCS-SOMP). By analyzing TP with the assumption that each sensing matrix contains i.i.d. Gaussian elements and that the nonzero values of each sparse vector are i.i.d. Gaussian variables, they demonstrated that with $M \geq 1$, TP reconstructs the support set as S is sufficiently large. They conjectured that $M \geq K + 1$ suffice for DCS-SOMP to reconstruct the support set as S is sufficiently large, based on its empirical results.

To the best of our knowledge, no information-theoretic study has been published to get necessary and sufficient conditions for reliable support set reconstruction in noisy MMV with different sensing matrices. Besides, these conditions have not been provided from the practical recovery algorithms for CS with noisy MMV with different sensing matrices.

C. Motivations of This Paper

CS with noisy MMV with different sensing matrices has been applied in many applications and the benefits facilitated by the joint sparsity structure have been empirically reported in [10] and [14]. In WSNs, Caione *et al.* [10] used the joint sparsity structure to reduce the number of transmitted bits per sensor and reported that each sensor can reduce its transmission cost. In magnetic resonance imaging (MRI), Wu *et al.* [14] modeled multiple diffusion tensor images (DTIs) as jointly sparse vectors. They exploited the joint sparsity structure to reduce the number of samples per DTI, while retaining the reconstruction quality. Using the joint sparsity structure, they also empirically reported that the reconstruction quality of each DTI can be improved for a fixed number of samples per DTI.

To theoretically explain the above empirical benefits facilitated by the joint sparsity structure, theoretical tools are required to measure the performance of CS with noisy MMV with different sensing matrices. Such tools can be useful as guidelines for determining the system parameters in various CS applications with noisy MMV with different sensing matrices. For example, if the number of samples per DTI is fixed in the MRI [14], the theoretical tools may enable us to determine the number of DTIs required for achieving a given reconstruction quality. Thus, the first motivation of this paper is to provide the theoretical tools by establishing sufficient and necessary conditions for reliable support set reconstruction.

Next, for noiseless MMV with the same sensing matrix, let $\mathbf{Y}_{\mathrm{A}} = \mathbf{F} \times \left[\mathbf{x}^1\ \mathbf{x}^2\ \cdots\ \mathbf{x}^S\right] \in \mathbb{R}^{M \times S}$. Also, for noiseless MMV with different sensing matrices, let $\mathbf{Y}_{\mathrm{B}} = \left[\mathbf{F}^1\mathbf{x}^1\ \ \mathbf{F}^2\mathbf{x}^2\ \cdots\ \mathbf{F}^S\mathbf{x}^S\right] \in \mathbb{R}^{M \times S}$. Then, all the elements of $\mathbf{Y}_{\mathrm{B}}$ are uncorrelated because all the sensing matrices are independent. In contrast, those of $\mathbf{Y}_{\mathrm{A}}$ are correlated because they are taken from the same sensing matrix. Now, we consider a case where we set $S > K$ and $M > K$. Then, it is clear that $\operatorname{rank}(\mathbf{Y}_{\mathrm{B}}) = \min(S, M)$ with a high probability and $\operatorname{rank}(\mathbf{Y}_{\mathrm{A}}) \leq K$. Therefore, for this case, we conclude that $\operatorname{rank}(\mathbf{Y}_{\mathrm{B}}) > \operatorname{rank}(\mathbf{Y}_{\mathrm{A}})$. This implies that a more reliable support set reconstruction can be expected in noiseless MMV with different sensing matrices for this case. Thus, the second motivation is to verify this perception in the presence of noise, by comparing our results with the existing ones in noisy MMV with the same sensing matrix [27].

D. Contributions of This Paper

The contributions of this paper are as follows: First, we derive upper and lower bounds of a failure probability of the support set reconstruction from Lemmas 1 and 2, by exploiting Fano's inequality [24] and the Chernoff bound [31]. These bounds are used for measuring the performance of CS with noisy MMV with different sensing matrices.

Next, we develop necessary and sufficient conditions for reliable support set reconstruction. Theorem 1 states that

$$M > K\left(1 + \frac{1}{Sf\left(\mathrm{SNR}_{\min}\right)}\right)$$

suffices to achieve reliable support set reconstruction in the *linear sparsity* regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = \beta \in (0, 1/2)$, and it also states that

$$M > K\left(1 + \frac{1}{Sf(\mathrm{SNR}_{\min})} \log \frac{N}{K}\right)$$

suffices to achieve reliable support set reconstruction in the*sublinear sparsity* regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = 0$, where

$f(\mathrm{SNR}_{\min})$ is an increasing function with respect to the minimum signal-to-noise ratio $\mathrm{SNR}_{\min}$ defined in (4). Next, for a finite S, N, K, and $\mathrm{SNR}_{\min}$, Theorem 3 states that

$$M < \frac{2K \log \frac{N}{K} - 2\log 2}{S \log(1 + K \times \mathrm{SNR}_{\min})}$$

is necessary for reliable support set reconstruction. The necessary and sufficient conditions can be useful as guidelines to determine the system parameters of CS applications with noisy MMV with different sensing matrices. Corollaries 1 and 2 indicate that reliable support set reconstruction is possible as sufficiently many measurement vectors S for a fixed M are taken at a low $\mathrm{SNR}_{\min}$. For a fixed N and K, Theorem 2 shows that $M \geq K + 1$ measurements suffice for reconstructing the support set, as S is sufficiently large. Then, for a fixed N, K, and $M = K + 1$, Corollary 3 provides a sufficient condition on S for reliable support set reconstruction. We provide theoretical explanations of the benefits of the joint sparsity structure, which conform with the empirical results of CS applications with noisy MMV with different sensing matrices [10], [14]. Finally, we compare the sufficient condition (11) with the known one (26) for noisy MMV with the same sensing matrix [27]. Therefore, we demonstrate that if $S \geq K$, noisy MMV with different sensing matrices may require fewer measurements M for reliable support set reconstruction than noisy MMV with the same sensing matrix under a low noise-level scenario. It confirms the superiority of MMV with different sensing matrices.

II. Notations, System Model & Problem Formulation

A. Notations

The following notations will be used in the whole paper.

1. $\mathbb{P}$, $\mathbb{E}$ and $\mathbb{V}$ denote the probability, expectation and (co)variance, respectively.
2. A small (capital) bold letter $\mathbf{f}$ ($\mathbf{F}$) is a vector (matrix).
3. A sub-vector (sub-matrix) formed by the elements (columns) of $\mathbf{f}$ ($\mathbf{F}$) indexed by a set $\mathcal{I}$ is denoted by $\mathbf{f}_{\mathcal{I}}$ ($\mathbf{F}_{\mathcal{I}}$).
4. For a given matrix $\mathbf{F}$, its inversion, transpose, trace and the ith eigenvalue are denoted by $\mathbf{F}^{-1}$, $\mathbf{F}^T$, $\mathrm{tr}[\mathbf{F}]$ and $\lambda_i(\mathbf{F})$, respectively. Also, its orthogonal projection matrix is defined by
$$\mathbf{Q}(\mathbf{F}) := \mathbf{I}_M - \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T \tag{2}$$
where $\mathbf{Q}(\mathbf{F})$ maps an arbitrary vector to the space orthogonal onto the space spanned by the columns of $\mathbf{F}$.
5. For given sets $\mathcal{I}$ and $\mathcal{J}$, the relative complements of $\mathcal{J}$ in $\mathcal{I}$ is denoted as $\mathcal{J} \setminus \mathcal{I}$. The cardinality of a set $\mathcal{I}$ is denoted by $|\mathcal{I}|$.
6. For a given function $f(x)$, its nth derivation with respect to x is denoted by $f^n(x)$.
7. The *linear sparsity regime* is defined by $\lim_{N\to\infty} \frac{K}{N} = \beta \in (0, 1/2)$.
8. The *sublinear sparsity regime* is defined by $\lim_{N\to\infty} \frac{K}{N} = 0$.
9. The expression $f(x) = \mathbf{\Omega}(g(x))$ denotes $|f(x)| \geq c|g(x)|$ as $x \to \infty$ for a constant $c > 0$.

B. System Model

Let $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^S$ be jointly *K-sparse* vectors with a support set $\mathcal{I}$ that belongs to

$$\mathcal{S} := \{\mathcal{H} | \mathcal{H} \subset \{1, 2, \cdots, N\}, |\mathcal{H}| = K\}.$$

Thus, the number of nonzero coefficients of each sparse vector is K, the indices of the nonzero coefficients of all the sparse vectors are the same and the indices belong to the support set.

In noisy MMV with different sensing matrices, each sparse vector is sampled by its own sensing matrix, i.e.,

$$\mathbf{y}^s = \mathbf{F}^s\mathbf{x}^s + \mathbf{n}^s \quad s = 1, 2, \cdots S \tag{3}$$

where all the sensing matrices have i.i.d. Gaussian elements with a zero mean and a unit variance, and all the noise vectors have i.i.d. Gaussian elements with a zero mean and a σ^2 variance. We assume that all the noise vectors and all the sensing matrices are mutually independent. Then, we let $x_{\min}$ be the smallest nonzero magnitude of all the sparse vectors and $\mathrm{SNR}_{\min}$ be the minimum signal-to-noise ratio given by

$$\mathrm{SNR}_{\min} := x_{\min}^2/\sigma^2. \tag{4}$$

C. Problem Formulation

We extend Akcakaya and Tarokh [21]'s decoder for noisy MMV with different sensing matrices. It takes all the measurement vectors as its input and yields a support set decision as its output

$$d : \left\{\forall_s \left(\mathbf{y}^s, \mathbf{F}^s\right)\right\} \mapsto \hat{\mathcal{I}} \in \mathcal{S}, \quad s = 1, 2, \cdots, S.$$

Its decision rules are given in Definition 1.

Definition 1: All the measurement vectors $\left\{\mathbf{y}^1, \mathbf{y}^2, \cdots, \mathbf{y}^S\right\}$ and a set $\mathcal{J} \in \mathcal{S}$ are δ *jointly typical* if the rank of $\mathbf{F}^s_{\mathcal{J}}$, $s = 1, \ldots, S$, is K and

$$\left|\left(\sum_{s=1}^{S} \left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{J}}\right)\mathbf{y}^s\right\|_2^2\right) - S(M-K)\sigma^2\right| < SM\delta. \tag{5}$$

As each sensing matrix contains i.i.d. Gaussian elements, the rank of each $\mathbf{F}^s_{\mathcal{J}}$, $s = 1, \ldots, S$, is K with a high probability. The decision rule is to find sets that satisfy (5) for all the given measurement vectors and $\delta > 0$. In the entire paper, the support set is denoted by $\mathcal{I}$ and any incorrect support set is denoted by $\mathcal{J}$, where their cardinalities are K, i.e., $|\mathcal{I}| = |\mathcal{J}| = K$.

We define the failure events, wherein the joint typical decoder fails to reconstruct the correct support set. First,

$$\mathcal{E}^c_{\mathcal{I}} := \left\{\left|\left(\sum_{s=1}^{S} \left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{I}}\right)\mathbf{y}^s\right\|_2^2\right) - S(M-K)\sigma^2\right| \geq SM\delta\right\} \tag{6}$$

implies that the correct support set is not δ jointly typical with all the measurement vectors. Next, for any $\mathcal{J} \in \mathcal{S} \setminus \mathcal{I}$,

$$\mathcal{E}_{\mathcal{J}} := \left\{\left|\left(\sum_{s=1}^{S} \left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{J}}\right)\mathbf{y}^s\right\|_2^2\right) - S(M-K)\sigma^2\right| < SM\delta\right\} \tag{7}$$

implies that an incorrect support set is δ jointly typical with all the measurement vectors. Based on these failure events, we define a failure probability and give its upper bound as follows:

$$\begin{aligned} p_{err} &:= \mathbb{P}\left\{\hat{\mathcal{I}} \neq \mathcal{I} \middle| \mathbf{x}^1, \cdots, \mathbf{x}^S\right\} \\ &= \mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c \bigcup_{\mathcal{J} \in \mathcal{S}\setminus\mathcal{I}} \mathcal{E}_{\mathcal{J}}\right\} \\ &\leq \mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} + \sum_{\mathcal{J} \in \mathcal{S}\setminus\mathcal{I}} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} \end{aligned} \tag{8}$$

where $\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\}$ is taken with respect to all the noise vectors and $\mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\}$ is taken with respect to all the noise vectors and all the sensing matrices. We establish Lemmas 1 and 2 given in Appendix A to give upper bounds of the probabilities of the failure events. Combining these lemmas with (8) yields

$$\begin{aligned} p_{err} &\leq \mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} + \sum_{\mathcal{J} \in \mathcal{S}\setminus\mathcal{I}} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} \\ &\leq 2p\left(d_1\right) + \binom{N}{K} p\left(d_{2,\alpha^*} - 1\right) \end{aligned}$$

where p is defined in (31), $d_1 = \frac{M\delta}{(M-K)\sigma^2}$, $d_{2,\alpha^*} = \frac{(M-K)\sigma^2 + M\delta}{(M-K)\alpha^*}$, and $\alpha^* = \sigma^2 + x_{\min}^2$.

It is of interest to examine why $\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\}$ depends only on the noise vectors. As shown in Lemma 3, the random variable to define the event $\mathcal{E}_{\mathcal{I}}^c$ in (6) is $\sum_{s=1}^S \left\|\mathbf{Q}\left(\mathbf{F}_{\mathcal{I}}^s\right)\mathbf{y}^s\right\|_2^2/\sigma^2$, where the measurement vector in (3) consists of the two parts: the noise part $\mathbf{n}^s$ and the signal part $\mathbf{F}_{\mathcal{I}}^s\mathbf{x}_{\mathcal{I}}^s$. The signal part belongs to the space spanned by the columns of $\mathbf{F}_{\mathcal{I}}^s$. Then, as specified in (2), the orthogonal projection matrix $\mathbf{Q}\left(\mathbf{F}_{\mathcal{I}}^s\right)$ maps the measurement vector to the space orthogonal onto the space spanned by the columns of $\mathbf{F}_{\mathcal{I}}^s$. Thus, the random variable is a function of the noise vectors only.

III. Main Results

As the main contribution of this paper, this section presents sufficient and necessary conditions on M for reliable support set reconstruction, i.e., p_{err} converges to zero, in noisy MMV with different sensing matrices. We then interpret the conditions to demonstrate the benefits facilitated by the joint sparsity structure.

A. Sufficient Conditions on M

In [18] and [21], the authors have shown that fewer measurements M for a reliable support set reconstruction are required for noisy SMV in the linear sparsity regime, compared to the sublinear sparsity regime. Based on the results of [18] and [21], we are motivated to examine if the same result can be observed in noisy MMV with different sensing matrices.

Theorem 1: For any $\rho > 1$, we let $\delta = \rho^{-1}\left(1 - K/M\right)x_{\min}^2$. If the number of measurements satisfies

$$M > K + \upsilon_1 \frac{K}{S} \tag{9}$$

then the failure probability p_{err} defined in (8) converges to zero in the linear sparsity regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = \beta \in (0, 1/2)$, where

$$\upsilon_1 = -\frac{2\left(1 - \log\beta\right)}{\log\left(1 - \frac{1-\rho^{-1}}{1+\mathrm{SNR}_{\min}^{-1}}\right) + \frac{1-\rho^{-1}}{1+\mathrm{SNR}_{\min}^{-1}}} > 0. \tag{10}$$

Also, under the same conditions on ρ and δ, if the number of measurements satisfies

$$M > K + \upsilon_2 \frac{K}{S} \log\frac{N}{K} \tag{11}$$

then the failure probability p_{err} defined in (8) converges to zero in the sublinear sparsity regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = 0$, where

$$\upsilon_2 = -\frac{2}{\log\left(1 - \frac{1-\rho^{-1}}{1+\mathrm{SNR}_{\min}^{-1}}\right) + \frac{1-\rho^{-1}}{1+\mathrm{SNR}_{\min}^{-1}}} > 0. \tag{12}$$

Proof: The proof is given in Appendix C.

In terms of N, K, and S, the asymptotic order of the sufficient condition on M for the linear sparsity regime is $\Omega\left(K + \frac{K}{S}\right)$, whereas the order for the sublinear sparsity regime is $\Omega\left(\frac{K}{S}\log\frac{N}{K}\right)$. It confirms that fewer measurements are required in the linear sparsity regime, compared to the sublinear sparsity regime. Next, from the sufficient conditions, we observe an inverse relationship between M and S, owing to the joint sparsity structure. This relationship implies that taking more measurement vectors S reduces the number of required measurements M for reliable support set reconstruction. Then, the relationship can be used for explaining the empirical results of Caione *et al.* [10] and Wu *et al.* [14]. In [10], the authors have reported that the number of transmitted bits per sensor could be inversely reduced by the number of sensors, which implies that the transmission cost of each sensor could be saved. The result can be confirmed by our inverse relationship by considering S and M as the number of sensors and the number of transmitted bits per sensor, respectively. In [14], S and M are considered as the number of DTIs and the number of samples of each DTI, respectively. Again, it has been observed from [14] that the joint sparsity structure enabled the number of samples of each DTI to be inversely reduced by the number of DTIs, reducing the acquisition time for each DTI. These results can be confirmed by our inverse relationship.

Theorem 2: For any $\rho > 1$, we let $\delta = \rho^{-1}\left(1 - K/M\right)x_{\min}^2$, N and K be fixed. If the number of measurements satisfies $M \geq K + 1$, the failure probability p_{err} defined in (8) converges to zero as the number of measurement vectors is increased to the infinity.

Proof: The proof is given in Appendix C.

Theorem 2 suggests that with $M \geq K + 1$, reliable support set reconstruction for noisy MMV with different sensing matrices is possible when a large number of measurement vectors is available. The sufficient conditions in Theorem 1, i.e., (9) and (11) have $\mathrm{SNR}_{\min}$ values as shown in (10) and (12). They disappear in the sufficient condition of Theorem 2, i.e., $M \geq K + 1$. The support set reconstruction problem becomes robust against noise when the number of measurement vectors is large.

B. Discussions on the Sufficient Conditions

We now examine the effect of $\mathrm{SNR}_{\min}$ on the sufficient conditions of Theorem 1. The aim is to determine the relationship among S, M and $\mathrm{SNR}_{\min}$ for reliable support set reconstruction.

Corollary 1: For any $\rho > 1$, we let $\delta = \rho^{-1}(1 - K/M)\, x_{\min}^2$. The sufficient conditions of Theorem 1 are rewritten as

$$M > K + \left(\frac{\sqrt[-1]{S} + \left(\sqrt{S} \times \mathrm{SNR}_{\min}\right)^{-1}}{1 - \rho^{-1}} \right)^2 4K \log \frac{N}{K} \quad (13)$$

in the sublinear sparsity regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = 0$, and

$$M > K + \left(\frac{\sqrt[-1]{S} + \left(\sqrt{S} \times \mathrm{SNR}_{\min}\right)^{-1}}{1 - \rho^{-1}} \right)^2 4K\,(1 - \log \beta) \quad (14)$$

in the linear sparsity regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = \beta \in (0, 1/2)$.

Proof: The proof is given in Appendix D.

Corollary 1 suggests that for a fixed M, reliable support set reconstruction is possible as the number of measurement vectors S is increased to infinity, although $\mathrm{SNR}_{\min}$ is low. Namely, we observe a noise reduction effect, which shows that using the joint sparsity structure leads to an increase in $\mathrm{SNR}_{\min}$ or a decrease in σ^2 by the square root of S. This effect can explain the improvement in the reconstruction quality of the DTIs, as empirically reported in [14].

We then improve our noise reduction effect by considering that $\mathrm{SNR}_{\min}$ is larger than a certain value.

Corollary 2: For any $\rho > 3$, we let $\delta = \rho^{-1}(1 - K/M)\, x_{\min}^2$ and $\alpha = 2/3$. If

$$\mathrm{SNR}_{\min} \geq \frac{\alpha}{1 - \rho^{-1} - \alpha} = \frac{2\rho}{\rho - 3}, \quad (15)$$

the sufficient conditions of Theorem 1 are rewritten as

$$M > K + \frac{S^{-1} + (S \times \mathrm{SNR}_{\min})^{-1}}{1 - \rho^{-1}} 4K \log \frac{N}{K} \quad (16)$$

in the sublinear sparsity regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = 0$, and

$$M > K + \frac{S^{-1} + (S \times \mathrm{SNR}_{\min})^{-1}}{1 - \rho^{-1}} 4K\,(1 - \log \beta) \quad (17)$$

in the linear sparsity regime, i.e., $\lim_{N\to\infty} \frac{K}{N} = \beta \in (0, 1/2)$.

Proof: The proof is given in Appendix D.

First of all, Corollary 2 requires $\rho > 3$ to ensure that the lower bound in (15) is positive. A simple computation shows that Corollary 2 requires fewer measurements in both the regimes compared to Corollary 1 because

$$\begin{aligned} \left(\frac{\sqrt[-1]{S} + \left(\sqrt{S} \times \mathrm{SNR}_{\min}\right)^{-1}}{1 - \rho^{-1}} \right)^2 &= S^{-1} \left(\frac{1 + \mathrm{SNR}_{\min}^{-1}}{1 - \rho^{-1}} \right)^2 \\ &\geq S^{-1} \left(\frac{1 + \mathrm{SNR}_{\min}^{-1}}{1 - \rho^{-1}} \right) \\ &= \frac{S^{-1} + (S \times \mathrm{SNR}_{\min})^{-1}}{1 - \rho^{-1}} \end{aligned}$$

where the second inequality is owing to $\frac{1+\mathrm{SNR}_{\min}^{-1}}{1-\rho^{-1}} = \frac{1}{t} > 1$ for any $\rho > 3$ and t defined in (61). Besides, Corollary 2 improves the noise reduction effect observed in Corollary 1 by showing that $\mathrm{SNR}_{\min}$ is increased by S for the region of $\mathrm{SNR}_{\min}$ in (15).

Theorem 2 suggests, it is to be noted, that $M = K+1$ is sufficient for reliable support set reconstruction if S is sufficiently large with a fixed N and K. Then, it would be interesting to determine how large S should be required for achieving the minimum number of measurements at each sensor, i.e., $M = K + 1$. In wireless sensor networks [34], energy sources used in sensors are very limited due to limitation of sensor sizes. Thus, minimizing the energy used for transmission of data at each sensor which often leads to extending the lifetime of the sensor battery is a value of importance. This point is noted in Caione *et al*. [10] as an advantage of using distributed compressed sensing on joint sparse model-2 signal ensembles (see Section V there). Corollary 3 which aims to provide a sufficient condition on S for achieving $M = K + 1$ thus is motivated.

Corollary 3: Let N and K be fixed and finite. For any $\rho > 1$, we let $\delta = \rho^{-1}(K+1)^{-1} x_{\min}^2$ and $M = K+1$. If the number of measurement vectors satisfies

$$S > \underbrace{\left(\log\left(\binom{N}{K} + 2\right) - \log \varepsilon\right) \times \max\left[\left|\frac{1}{\log \mu_{\mathcal{I}}}\right|, \left|\frac{1}{\log \mu_{\mathcal{J}}}\right|\right]}_{:=S^*} \quad (18)$$

reliable support set reconstruction is possible, i.e., $p_{err} < \varepsilon$ for sufficiently small $\varepsilon \in (0, 1)$, where $\log \mu_{\mathcal{I}}$ and $\log \mu_{\mathcal{J}}$ are defined in (63) and (65), respectively. The sufficient condition on S is decreasing with respect to $\mathrm{SNR}_{\min}$.

Proof: The proof is given in Appendix D.

To the best of our knowledge, the sufficient conditions on S for a reliable support set reconstruction have not yet been developed. A similar result has been reported by Tang and Nehorai [26], in which they reported that $M = \Omega\left(K \log \frac{N}{K}\right)$ and $S = \frac{\log N}{\log \log N}$ suffice for a reliable support set reconstruction in noisy MMV with the same sensing matrix, as N is sufficiently large.

It is of interest to examine whether the sufficient condition S^* in (18) is good. For this, we implement the joint typical decoder in (5) and conduct experiments for different values of $\mathrm{SNR}_{\min}$ and K, for a fixed $N = 50$. We count the number of failure occurrences, wherein the joint typical decoder fails to reconstruct the support set. We obtain the smallest S^{emp} such that the ratio of the failure occurrences is smaller than $\varepsilon = 0.01$. By comparing S^{emp} with S^* in (18), we see that S^* approaches S^{emp}, as $\mathrm{SNR}_{\min}$ is sufficiently large. For example, we see that $S^{emp} = 8$ and $S^* = 12$ at $\mathrm{SNR}_{\min} = 20$ [dB], $K = 2$, and $S^{emp} = 5$ and $S^* = 6$ at $\mathrm{SNR}_{\min} = 30$ [dB], $K = 2$. A similar trend is observed with a bigger K, i.e., at $K = 5$. For example, we see that $S^{emp} = 12$ and $S^* = 19$ at $\mathrm{SNR}_{\min} = 20$ [dB], and $S^{emp} = 7$ and $S^* = 10$ at $\mathrm{SNR}_{\min} = 30$ [dB].

Fletcher *et al.* [19] have reported that the ML decoder requires $M = K + 1$ measurements for a reliable support set reconstruction in noisy SMV, when the signal-to-noise ratio is sufficiently large. This result can be observed from Corollary 3. Specifically, we assume that $\mathrm{SNR}_{\min}$ is sufficiently large for a fixed N and K. Then, from (63) and (65), it is easy to see that

$$\lim_{\mathrm{SNR}_{\min}\to\infty} \log \mu_{\mathcal{I}} = -\infty,$$
$$\lim_{\mathrm{SNR}_{\min}\to\infty} \log \mu_{\mathcal{J}} = 2^{-1}\left(1 - \rho^{-1} - \log \rho\right).$$

Hence, (18) is simplified to

$$S > \left(\log\left(\binom{N}{K} + 2\right) - \log \varepsilon\right) \times \left|2\left(1 - \rho^{-1} - \log \rho\right)^{-1}\right|. \tag{19}$$

Note that N, K, and ε are fixed. Thus, for a large ρ, we have

$$\left|1 - \rho^{-1} - \log \rho\right| \gg 2\left(\log\left(\binom{N}{K} + 2\right) - \log \varepsilon\right), \tag{20}$$

which leads to $S \geq 1$. This result suggests that the joint typical decoder requires $M = K+1$ measurements for reliable support set reconstruction in noisy SMV, whenever $\mathrm{SNR}_{\min}$ is sufficiently large and ρ satisfies (20).

C. Necessary Condition on M

We specify a necessary condition that must be satified by a decoder for reliable support set reconstruction in noisy MMV with different sensing matrices. Unlike the sufficient conditions of Theorem 1, the necessary condition is presented for a finite N and K.

We begin by transforming (3) into

$$\underbrace{\begin{bmatrix} \mathbf{y}^1 \\ \vdots \\ \mathbf{y}^S \end{bmatrix}}_{=:\mathbf{y}\in\mathbb{R}^{SM}} = \underbrace{\begin{bmatrix} \mathbf{F}^1 & & \\ & \ddots & \\ & & \mathbf{F}^S \end{bmatrix}}_{=:\tilde{\mathbf{F}}\in\mathbb{R}^{SM\times SN}} \underbrace{\begin{bmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^S \end{bmatrix}}_{=:\mathbf{x}\in\mathbb{R}^{SN}} + \underbrace{\begin{bmatrix} \mathbf{n}^1 \\ \vdots \\ \mathbf{n}^S \end{bmatrix}}_{=:\mathbf{n}\in\mathbb{R}^{SM}} \tag{21}$$

where $\mathbf{x}$ is an *SK-sparse* vector belonging to an infinite set

$$\mathcal{X}_{x_{\min}} := \left\{\mathbf{x} \in \mathbb{R}^{SN} \,\middle|\, |x(i)| \geq x_{\min}, \forall i \in \mathcal{I}, |\mathcal{I}| = SK\right\}$$

where $x(i)$ is the ith element of $\mathbf{x}$ and $\mathcal{I}$ is the support set of $\mathbf{x}$. Owing to the joint sparsity structure, the number of possible support sets is $\binom{N}{K}$. Then, we define a failure probability as:

$$p_{err} := \mathbb{E}_{\tilde{\mathbf{F}}} \sup_{\mathbf{x}\in\mathcal{X}_{x_{\min}}} \mathbb{P}\left\{\hat{\mathcal{I}} \neq \mathcal{I} \,\middle|\, \mathbf{x}, \tilde{\mathbf{F}}\right\} \tag{22}$$

where $\hat{\mathcal{I}}$ is an estimate of the support set based on $\mathbf{y}$ and $\tilde{\mathbf{F}}$ in (21). Then, Lemma *III*-3 of [20] yields

$$\sup_{\mathbf{x}\in\mathcal{X}_{x_{\min}}} \mathbb{P}\left\{\hat{\mathcal{I}} \neq \mathcal{I} \,\middle|\, \mathbf{x}, \tilde{\mathbf{F}}\right\} \geq \min_{\hat{\mathbf{x}}\in\mathcal{X}_{\{x_{\min}\}}} \max_{\mathbf{x}\in\mathcal{X}_{\{x_{\min}\}}} \mathbb{P}\left\{\hat{\mathbf{x}} \neq \mathbf{x} \,\middle|\, \mathbf{x}, \tilde{\mathbf{F}}\right\} \tag{23}$$

where $\hat{\mathbf{x}}$ is an estimate for $\mathbf{x}$ based on $\mathbf{y}$ and $\tilde{\mathbf{F}}$ in (21) and

$$\mathcal{X}_{\{x_{\min}\}} := \left\{\mathbf{x} \in \mathbb{R}^{SN} \,\middle|\, x(i) = x_{\min}, \forall i \in \mathcal{I}, |\mathcal{I}| = SK\right\}$$

which is a finite set. Assume that $\mathbf{x}$ is uniformly distributed over this finite set. Applying Fano's inequality [24] to (23) yields

$$\max_{\mathbf{x}\in\mathcal{X}_{\{x_{\min}\}}} \mathbb{P}\left\{\hat{\mathbf{x}} \neq \mathbf{x} \,\middle|\, \mathbf{x}, \tilde{\mathbf{F}}\right\} \geq \mathbb{P}\left\{\hat{\mathbf{x}} \neq \mathbf{x} \,\middle|\, \tilde{\mathbf{F}}\right\}$$
$$\geq 1 - \frac{\mathbb{I}\left(\mathbf{x}; \mathbf{y} | \tilde{\mathbf{F}}\right) + \log 2}{\log\left(\left|\mathcal{X}_{\{x_{\min}\}}\right| - 1\right)} \tag{24}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ belong to the finite set $\mathcal{X}_{\{x_{\min}\}}$ and $\mathbb{I}(\mathbf{x};\ \mathbf{y})$ is the mutual information between $\mathbf{x}$ and$\mathbf{y}$. We get a necessary condition on M to ensure that the lower bound in (24) is bounded away from zero, as follows:

Theorem 3: Let N and K are fixed and finite. In (21), if the number of measurements satisfies

$$M < \frac{2K \log \frac{N}{K} - 2\log 2}{S \log\left(1 + K \times \mathrm{SNR}_{\min}\right)} \tag{25}$$

then the failure probability p_{err} defined in (22) is bounded away from zero.

Proof: The proof is given in Appendix C.

IV. Relations to the Existing Information-Theoretic Results

A. Relations to Noisy MMV With the Same Sensing Matrix [27]

Jin and Rao [27] have exploited the Chernoff bound to obtain a tight sufficient condition on M for a reliable support set reconstruction for noisy MMV with the same sensing matrix in the sublinear sparsity regime. Owing to the complicated form of their sufficient condition, they could not clearly show the benefits facilitated by the joint sparsity structure. Thus, they simplified their condition under scenarios such as: *i*) a low noise-level scenario and *ii*) a scenario with S identical sparse vectors. In Table I, we summarize our sufficient conditions on M, and compare them to that of [27] under the low noise-level scenario in the sublinear sparsity regime.

First, in a low noise-level scenario, as shown in Table I, the sufficient condition [27] for noisy MMV with the same sensing matrix is

$$M = \Omega\left(\frac{K \log N}{\min(K, S)}\right). \tag{26}$$

If $S < K$, the sufficient conditions (11) and (26) have the same order, implying that there is no significant performance gap in the support set reconstruction between the models. However, if $S > K$, (26) is $M = \Omega(\log N)$, whereas (11) is

TABLE I
SUFFICIENT CONDITIONS ON M FOR SUPPORT SET RECONSTRUCTION

	This paper	Yuzhe and Rao [27]
Linear sparsity regime $\lim_{N\to\infty}\frac{K}{N}=\beta\in(0,1/2)$	$M=\Omega\left(K+\frac{K}{S}\right)$	Not presented
Sublinear sparsity regime $\lim_{N\to\infty}\frac{K}{N}=0$	$M=\Omega\left(\frac{K}{S}\log\frac{N}{K}\right)$	$M=\Omega\left(\frac{K\log N}{\min(K,S)}\right)$
N and K are finite ($\mathrm{SNR}_{\min}\to\infty$ or $S\to\infty$)	$M\geq K+1$	Not presented

$M=\Omega\left(\frac{K}{S}\log N\right)$. It implies that noisy MMV with different sensing matrices is superior to noisy MMV with the same sensing matrix or $S>K$, with respect to M for reliable support set reconstruction. The result of this comparison supports the perception presented in Section I-C, wherein a more reliable support set reconstruction could be expected in a noiseless MMV with different sensing matrices owing to the linear independency of the measurement vectors. Moreover, it validates the perception, even in the presence of noise.

Second, we consider a scenario with S identical sparse vectors. Then, the sufficient condition of [27] is

$$M=\Omega\left(\frac{K\log N}{\log\left(1+S\left\|\mathbf{x}\right\|_2^2/\sigma^2\right)}\right). \tag{27}$$

From (27), we observe that σ^2 is reduced by a factor of S. However, the noise reduction effect for noisy MMV with the same sensing matrix requires a restriction, where all the sparse vectors should be identical, which can be hardly achieved in practice. In contrast, the noise reduction effect for noisy MMV with different sensing matrices does not require this restriction, as shown in Corollaries 1 and 2.

B. Relations to Noisy SMV [21]

Akcakaya and Tarokh [21] have used the joint typical decoder to establish the sufficient conditions on M for a reliable support set reconstruction in noisy SMV. They exploited the exponential inequalities [32] to obtain the upper bounds on the sum of the weighted chi-square random variables. In this subsection, we demonstrate that the approaches developed in this paper are superior to the use of the exponential inequalities. Thus, we use the exponential inequalities to generalize their bounds for noisy MMV with different sensing matrices. We give Propositions 1 and 2 to prove that the generalized bounds are worse than the bounds of Lemmas 1 and 2.

Proposition 1: For any positive δ, we have

$$\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\}\leq 2p\left(d_1\right)\leq 2p_{1,\exp}$$

where both $p\left(d_1\right)$ and d_1 are given in Lemma 1, and

$$p_{1,\exp}:=\exp\left(-\frac{S\delta^2}{4\sigma^4}\frac{M^2}{M-K+2\delta M/\sigma^2}\right). \tag{28}$$

Proof: The proof is given in Appendix E.

Proposition 2: For any $\mathcal{J}\in\mathcal{S}\setminus\mathcal{I}$ and any $\delta>0$ such that

$$0<\delta<(1-K/M)\,x_{\min,\mathcal{J}}^2, \tag{29}$$

we have

$$\mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\}\leq p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}-1\right)\leq p_{2,\mathcal{J},\exp}$$

where both $p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}-1\right)$ and $d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}$ are given in Lemma 2 and

$$p_{2,\mathcal{J},\exp}:=\exp\left(-\frac{S^2\left(M-K\right)}{4\sum_{s=1}^{S}\alpha_{\mathcal{J},s}^2}\left(x_{\min,\mathcal{J}}^2-\frac{M\delta}{M-K}\right)^2\right) \tag{30}$$

and $\alpha_{\mathcal{J},s}$ is defined in (39) and $x_{\min,\mathcal{J}}^2$ is defined in (43).

Proof: The proof is given in Appendix E.

If $S=1$, we can see that $p_{1,\exp}$ and $p_{2,\mathcal{J},\exp}$ are equivalent to the bounds of Akcakaya and Tarokh [21]. Propositions 1 and 2 state that the bounds on the failure probability of Lemmas 1 and 2 are tighter than the bounds of [21] for noisy SMV.

V. CONCLUSIONS

We have studied a support set reconstruction problem for CS with noisy MMV with different sensing matrices. The union and Chernoff bounds have been used to obtain the upper bound of the failure probability of the support set reconstruction, and Fano's inequality has been used to obtain the lower bound of this failure probability. As we have obtained the upper bound by analyzing an exhaustive search decoder, the bound is used to measure the performance of CS with noisy MMV with different sensing matrices. We have then developed the necessary and sufficient conditions in terms of the sparsity K, the ambient dimension N, the number of measurements M, the number of measurement vectors S, and the minimum signal-to-noise ratio $\mathrm{SNR}_{\min}$. They can be useful as guidelines to determing the system parameters in various CS applications with noisy MMV with different sensing matrices.

The conditions are interpreted to provide theoretical explanations for the benefits facilitated by the joint sparsity structure in noisy MMV with different sensing matrices:

i. From the sufficient conditions of Theorem 1, we have observed an inverse relationship between M and S. Due to the inverse relation, we take fewer measurements M per each measurement vector for reliable support set reconstruction by taking more measurement vectors S.

ii. From the sufficient conditions of Corollaries 1 and 2, we have observed a noise reduction effect, which shows that the usage of the joint sparsity structure results in an increase in $\mathrm{SNR}_{\min}$ or a decrease in σ^2 by a factor of S. Therefore, the support set reconstruction can be robust against noise as the number of measurement vectors is increased to infinity.

iii. From Theorem 2, we have shown that $M=K+1$ is achieved for a fixed N and K, as S is sufficiently large. From Corollary 3, we have provided the sufficient condition on S to reconstruct the support set for a fixed N, K, and $M=K+1$.

The theoretical explanations confirm the benefits of the joint sparsity structure, as empirically shown in CS applications with noisy MMV with different sensing matrices [10], [14].

We have compared our sufficient conditions for noisy MMV with different sensing matrices with the other existing results [27] for noisy MMV with the same sensing matrix. For a low-level noise scenario with $S \geq K$, we have shown that the number of measurements for reliable support set reconstruction for noisy MMV with different sensing matrices is lesser than that for noisy MMV with the same sensing matrix. Also, [27] has shown the noise reduction effect. This was achieved under a rather restricted condition though, i.e., all sparse vectors are the same. While such a restricted condition is not required in the noisy MMV problem with *different* sensing matrices studied in this paper, the noise reduction effect has also been observed, which is a significant improvement.

Appendix A
Lemmas 1 and 2

This section presents Lemmas 1 and 2, which give upper bounds of the probabilities of the failure events defined in (6) and (7), respectively. Also, for simplicity, we define

$$p(x) = \exp\left(-\frac{S(M-K)}{2}x\right)(1+x)^{\frac{S(M-K)}{2}}. \quad (31)$$

Lemma 1: For any positive δ, we have

$$\begin{aligned}\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} &\leq 2\exp\left(-\frac{S(M-K)}{2}d_1\right)(1+d_1)^{\frac{S(M-K)}{2}}\\ &= 2p(d_1)\end{aligned} \quad (32)$$

where the function p is defined in (31), and

$$d_1 := \frac{M\delta}{(M-K)\sigma^2} > 0. \quad (33)$$

Proof: From (6), we have

$$\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} = \mathbb{P}\left\{\mathrm{Z}_{\mathcal{I}} \leq W_1\right\} + \mathbb{P}\left\{\mathrm{Z}_{\mathcal{I}} \geq W_2\right\} \quad (34)$$

where $\mathrm{Z}_{\mathcal{I}}$ is defined in Lemma 3, and

$$W_i = S(M-K) + (-1)^i SM\delta/\sigma^2, \quad i = 1,\ 2.$$

Applying the Chernoff bound [31] to (34) yields

$$\begin{aligned}\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} &\leq \sum_{i=1}^2 \exp(-t_i W_i)\,\mathbb{E}\left[\exp(t_i \mathrm{Z}_{\mathcal{I}})\right]\\ &= \sum_{i=1}^2 \underbrace{\exp(-t_i W_i)(1-2t_i)^{-S(M-K)/2}}_{=:f(t_i;W_i)}\end{aligned} \quad (35)$$

where the equality is from Lemma 3, $t_1 < 0$ and $t_2 \in \left(0, \frac{1}{2}\right)$. As each $f(t_i; W_i)$ is convex, $t_i = t_i^*$at $f^{(1)}(t_i; W_i) = 0$ yields the minimizer of $f(t_i; W_i)$, where

$$t_i^* = 2^{-1}\left(1 - W_i^{-1} S(M-K)\right), \quad i = 1,\ 2.$$

Thus, $f(t_i; W_i) \geq f\left(t_i^*; W_i\right)$ for each i. If $W_1 \leq 0$, it is clear that $\mathbb{P}\{\mathrm{Z}_{\mathcal{I}} \leq W_1\} = 0$ because $\mathrm{Z}_{\mathcal{I}}$ is quadratic. Thus,

$$\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} = \mathbb{P}\left\{\mathrm{Z}_{\mathcal{I}} \geq W_2\right\} \leq f\left(t_2^*; W_2\right) = p(d_1) \quad (36)$$

where $p(d_1)$ and d_1 are defined in (32) and (33), respectively. If $W_1 > 0$ then $f\left(t_1^*; W_1\right) \leq f\left(t_2^*; W_2\right)$ because

$$\begin{aligned}&\log f\left(t_1^*; W_1\right) - \log f\left(t_2^*; W_2\right)\\ &= S(M-K)\left[d_1 + 2\log(1-d_1) - 2\log(1+d_1)\right] < 0.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} &= f\left(t_1^*; W_1\right) + f\left(t_2^*; W_2\right) \leq 2f\left(t_2^*; W_2\right)\\ &= 2\exp\left(-\frac{S(M-K)}{2}d_1\right)(1+d_1)^{\frac{S(M-K)}{2}}.\end{aligned} \quad (37)$$

Finally, combining (36) and (37) leads to (32). ■

Lemma 2: Let $\mathcal{J} \in \mathcal{S} \setminus \mathcal{I}$ and a matrix $\mathbf{R}_{\mathcal{J}}$ be

$$\mathbf{R}_{\mathcal{J}} = \begin{bmatrix}\alpha_{\mathcal{J},1}\mathbf{I}_{M-K} & & \\ & \ddots & \\ & & \alpha_{\mathcal{J},S}\mathbf{I}_{M-K}\end{bmatrix} \quad (38)$$

where

$$\alpha_{\mathcal{J},s} := \sigma^2 + \left\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{J}}^s\right\|_2^2 > 0. \quad (39)$$

Consider any positive δ such that

$$0 < \delta < (1 - K/M)\left(\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right) - \sigma^2\right)$$

where $\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)$ is the smallest eigenvalue of $\mathbf{R}_{\mathcal{J}}$. Then,

$$\begin{aligned}\mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} &\leq \exp\left(-\frac{S(M-K)}{2}\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right)\right) d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}^{\frac{S(M-K)}{2}}\\ &= p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right)\\ &\leq p\left(d_{2,\alpha^*} - 1\right)\end{aligned} \quad (40)$$

where the function p is defined in (31),

$$d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} := \frac{(M-K)\sigma^2 + M\delta}{(M-K)\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)} \in (0, 1), \quad (41)$$

$$\alpha^* := \sigma^2 + x_{\min}^2, \quad (42)$$

and

$$x_{\min}^2 = \min_{\mathcal{J}\in\mathcal{S}\setminus\mathcal{I}} \underbrace{\min_{s\in\{1,2,\cdots,S\}} \left\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{J}}^s\right\|_2^2}_{=:x_{\min,\mathcal{J}}^2}. \quad (43)$$

Proof: From (7), we have

$$\mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} = \mathbb{P}\left\{\mathrm{Z}_{\mathcal{J}} < W_1\right\} - \mathbb{P}\left\{\mathrm{Z}_{\mathcal{J}} < W_2\right\} \leq \mathbb{P}\left\{\mathrm{Z}_{\mathcal{J}} < W_1\right\} \quad (44)$$

where $\mathrm{Z}_{\mathcal{J}}$ is defined in Lemma 4, and

$$W_i = S(M-K)\sigma^2 - (-1)^i SM\delta, \quad i = 1, 2. \quad (45)$$

Applying the Chernoff bound [31] to (44) yields for $t < 0$,

$$\begin{aligned}\mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} &\leq \exp(-tW_1)\,\mathbb{E}\left[\exp\left(t\mathrm{Z}_{\mathcal{J}}\right)\right]\\ &= \exp(-tW_1)\prod_{i=1}^{S(M-K)}\left(1 - 2t\lambda_i\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-1/2}\\ &\leq \exp(-tW_1)\left(1 - 2t\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-S(M-K)/2}\\ &=: f(t; W_1)\end{aligned} \quad (46)$$

where the equality is from Lemma 4 and the second inequality is due to that all the eigenvalues are positive. We then define a function $h(t) := \log f(t; W_1)$. Then,

$$h^{(2)}(t) = 2S(M-K)\lambda_{\min}^2\left(\mathbf{R}_{\mathcal{J}}\right)\left(1 - 2t\lambda_{\min}\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-2} > 0$$

which implies that h is convex with respect to t. It leads to that f in (46) is logarithmically convex. Thus $t = t^*$ at $f^{(1)}(t; W_1) = 0$ yields the minimizer of $f(t; W_1)$ where

$$t^* = 2^{-1}\left(\lambda_{\min}^{-1}(\mathbf{R}_{\mathcal{J}}) - W_1^{-1} S(M-K)\right) < 0.$$

Substituting t^* in (46) yields

$$\begin{aligned}\mathbb{P}\{\mathcal{E}_{\mathcal{J}}\} &\le f(t^*; W_1) \\ &= \exp\left(-\frac{S(M-K)}{2}\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right)\right) d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}^{\frac{S(M-K)}{2}} \\ &= p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right) \end{aligned} \quad (47)$$

where $d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}$ is defined in (41) and p is defined in (31).

Next, let $\beta = 2^{-1}S(M-K)$ and $x = d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})}$ in the upper bound (47). Then, we have $p(x-1) = x^{\beta}\exp(-\beta(x-1))$, where

$$\frac{\partial p(x-1)}{\partial x} = \beta x^{\beta}\exp(-\beta(x-1))\left(x^{-1}-1\right) > 0 \quad (48)$$

and

$$\frac{\partial x}{\partial \lambda_{\min}(\mathbf{R}_{\mathcal{J}})} = -x < 0. \quad (49)$$

Due to (48) and (49),

$$\begin{aligned}\frac{\partial p(x-1)}{\partial \lambda_{\min}(\mathbf{R}_{\mathcal{J}})} &= \frac{\partial p(x-1)}{\partial x}\frac{\partial x}{\partial \lambda_{\min}(\mathbf{R}_{\mathcal{J}})} \\ &= -\beta x^{\beta-1}\exp(-\beta(x-1))\left(x^{-1}-1\right) < 0\end{aligned}$$

which shows that the upper bound in (47) is decreasing with respect to $\lambda_{\min}(\mathbf{R}_{\mathcal{J}})$. Then, remind that the matrix in (38) is the covariance matrix of a multivariate Gaussian vector $\mathbf{b}$ in (58). Then for any incorrect support set, its smallest eigenvalue can be easily computed and lower bounded by

$$\lambda_{\min}(\mathbf{R}_{\mathcal{J}}) = \min_{s\in\{1,2,\cdots,S\}} \alpha_{\mathcal{J},s} = \sigma^2 + x^2_{\min,\mathcal{J}} \ge \alpha^* \quad (50)$$

where $x^2_{\min,\mathcal{J}}$ is defined in (43) and α^* is defined in (42). Thus, for any incorrect support set $\mathcal{J} \in \mathcal{S}\setminus\mathcal{I}$, we conclude that

$$\mathbb{P}\{\mathcal{E}_{\mathcal{J}}\} \le p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right) \le p\left(d_{2,\alpha^*} - 1\right)$$

which completes the proof. ■

Appendix B
Lemmas 3 and 4

First of all, we give the Scharf's theorem [33] to compute the moment generating function of a quadratic random variable. We then make Lemmas 3 and 4 to give the moment generating functions of the random variables of $\mathcal{E}^c_{\mathcal{I}}$ and $\mathcal{E}_{\mathcal{J}}$ that were used in the proofs of Lemmas 1 and 2, respectively.

Scharf's Theorem [33, p. 64]: Let $\mathbf{b} \in \mathbb{R}^N$ be a multivariate Gaussian vector with a mean $\mathbf{m}$ and a covariance $\mathbf{R}$. Then a random variable $Q \triangleq (\mathbf{b}-\mathbf{m})^T(\mathbf{b}-\mathbf{m})$ is quadratic with $\mathbb{E}[Q] = \mathrm{tr}[\mathbf{R}], \mathbb{V}[Q] = 2\mathrm{tr}\left[\mathbf{R}^T\mathbf{R}\right]$ and for any t

$$\mathbb{E}\left[\exp(tQ)\right] = \prod_{i=1}^{N}(1-2t\lambda_i(\mathbf{R}))^{-1/2}.$$

Lemma 3: In (6), define a quadratic random variable

$$\mathrm{Z}_{\mathcal{I}} := \sum_{s=1}^{S}\left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{I}}\right)\mathbf{y}^s\right\|_2^2/\sigma^2. \quad (51)$$

Then, $\mathbb{E}[\mathrm{Z}_{\mathcal{I}}] = S(M-K)$, $\mathbb{V}[\mathrm{Z}_{\mathcal{I}}] = 2S(M-K)$ and for any $0 < t < 0.5$,

$$\mathbb{E}\left[\exp(t\mathrm{Z}_{\mathcal{I}})\right] = (1-2t)^{-S(M-K)/2}. \quad (52)$$

Proof: The orthogonal projection matrix is decomposed as

$$\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{I}}\right) = \mathbf{U}^s_{\mathcal{I}}\mathbf{D}^s\left(\mathbf{U}^s_{\mathcal{I}}\right)^T$$

where $\mathbf{D}^s$ is a diagonal matrix, whose first $M-K$ diagonals are ones and the remains are zeros, and $\mathbf{U}^s_{\mathcal{I}}$ is a unitary matrix. Then,

$$\begin{aligned}Z_{\mathcal{I}} &= \sum_{s=1}^{S}\left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{I}}\right)\mathbf{y}^s\right\|_2^2/\sigma^2 = \sum_{s=1}^{S}\left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{I}}\right)\mathbf{n}^s\right\|_2^2/\sigma^2 \\ &= \sum_{s=1}^{S}\left\|\mathbf{D}^s\underbrace{\left(\mathbf{U}^s_{\mathcal{I}}\right)^T\mathbf{n}^s/\sigma^2}_{=:\mathbf{w}^s}\right\|_2^2 = \sum_{s=1}^{S}\left\|\mathbf{D}^s\mathbf{w}^s\right\|_2^2 \end{aligned} \quad (53)$$

where $\mathbf{w}^s$ is a multivariate Gaussian vector with mean $\mathbf{0}_M$ and covariance $\mathbf{I}_M$. Since the first $M-K$ diagonal elements of each diagonal matrix are ones, we have

$$\begin{aligned}Z_{\mathcal{I}} &= \sum_{s=1}^{S}\left\|\mathbf{D}^s\mathbf{w}^s\right\|_2^2 = \sum_{s=1}^{S}\sum_{i=1}^{M-K}\left|w^s(i)\right|^2 \\ &= \sum_{s=1}^{S}\left(\mathbf{w}^s_{\mathcal{P}}\right)^T\mathbf{w}^s_{\mathcal{P}} = \mathbf{w}^T\mathbf{w}\end{aligned} \quad (54)$$

which is quadratic, where

$$\mathbf{w}^s_{\mathcal{P}} = \left[\, w^s(1) \quad w^s(2) \quad \cdots \quad w^s(M-K)\,\right]^T$$

and

$$\mathbf{w} = \left[\,\left(\mathbf{w}^1_{\mathcal{P}}\right)^T \quad \left(\mathbf{w}^2_{\mathcal{P}}\right)^T \quad \cdots \quad \left(\mathbf{w}^S_{\mathcal{P}}\right)^T\,\right]^T. \quad (55)$$

In (53), $\mathbf{w}^s$ is determined by $\mathbf{U}^s_{\mathcal{I}}$ and $\mathbf{n}^s$. Since the elements of $\mathbf{U}^s_{\mathcal{I}}$ and $\mathbf{n}^s$ are independent, $\mathbf{w}^i$ and $\mathbf{w}^j$ are mutually independent for any $1 \le i \ne j \le S$. The covariance matrix of $\mathbf{w}$ is an identity matrix. Thus, applying the Scharf's theorem to $Z_{\mathcal{I}}$ completes the proof. ■

Lemma 4: In (7), for any $\mathcal{J} \in \mathcal{S}\setminus\mathcal{I}$, define a quadratic random variable

$$Z_{\mathcal{J}} := \sum_{s=1}^{S}\left\|\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{J}}\right)\mathbf{y}^s\right\|_2^2. \quad (56)$$

Then, $\mathbb{E}\left[\mathrm{Z}_{\mathcal{J}}\right] = \mathrm{tr}\left[\mathbf{R}_{\mathcal{J}}\right]$, $\mathbb{V}\left[\mathrm{Z}_{\mathcal{J}}\right] = 2\mathrm{tr}\left[\mathbf{R}^T_{\mathcal{J}}\mathbf{R}_{\mathcal{J}}\right]$ and for any t,

$$\mathbb{E}\left[\exp\left(t\mathrm{Z}_{\mathcal{J}}\right)\right] = \prod_{i=1}^{S(M-K)}\left(1-2t\lambda_i\left(\mathbf{R}_{\mathcal{J}}\right)\right)^{-1/2},$$

where $\mathbf{R}_{\mathcal{J}}$ is given in (38).

Proof: Similar to the proof of Lemma 3,

$$\mathbf{Q}\left(\mathbf{F}^s_{\mathcal{J}}\right) = \mathbf{U}^s_{\mathcal{J}}\mathbf{D}^s\left(\mathbf{U}^s_{\mathcal{J}}\right)^T$$

where $\mathbf{D}^s$ is a diagonal matrix, whose first $M-K$ diagonals are ones and the remains are zeros, and $\mathbf{U}^s_{\mathcal{J}}$ is a unitary matrix. Then,

$$Z_{\mathcal{J}} = \sum_{s=1}^{S} \left\| \mathbf{Q}\left(\mathbf{F}^s_{\mathcal{J}}\right) \mathbf{y}^s \right\|_2^2 = \sum_{s=1}^{S} \left\| \mathbf{Q}\left(\mathbf{F}^s_{\mathcal{J}}\right) \mathbf{c}^s \right\|_2^2 = \sum_{s=1}^{S} \left\| \mathbf{D}^s \underbrace{\left(\mathbf{U}^s_{\mathcal{J}}\right)^T \mathbf{c}^s}_{=:\mathbf{b}^s} \right\|_2^2 = \sum_{s=1}^{S} \left\| \mathbf{D}^s \; \mathbf{b}^s \right\|_2^2 \quad (57)$$

where $\mathbf{b}^s$ is a multivariate Gaussian vector with mean $\mathbf{0}_M$ and

$$\mathbb{V}\left[\mathbf{b}^s\right] = \left(\sigma^2 + \left\| \mathbf{x}^s_{\mathcal{I}\backslash\mathcal{J}} \right\|_2^2\right) \mathbf{I}_M$$

and $\mathbf{c}^s = \mathbf{n}^s + \sum_{u \in \mathcal{I}\backslash\mathcal{J}} \mathbf{f}^s_u x^s(u)$. Since the first $M-K$ diagonal elements of each diagonal matrix are ones, we have

$$\begin{aligned} Z_{\mathcal{J}} &= \sum_{s=1}^{S} \left\| \mathbf{D}^s \; \mathbf{b}^s \right\|_2^2 = \sum_{s=1}^{S} \sum_{i=1}^{M-K} \left| b^s(i) \right|^2 \\ &= \sum_{s=1}^{S} \left(\mathbf{b}^s_{\mathcal{P}}\right)^T \mathbf{b}^s_{\mathcal{P}} = \mathbf{b}^T \mathbf{b} \end{aligned} \quad (58)$$

which is quadratic, where

$$\mathbf{b}^s_{\mathcal{P}} = \left[\, b^s(1) \quad b^s(2) \quad \cdots \quad b^s(M-K) \,\right]^T$$

and

$$\mathbf{b} = \left[\, \left(\mathbf{b}^1_{\mathcal{P}}\right)^T \quad \left(\mathbf{b}^2_{\mathcal{P}}\right)^T \quad \cdots \quad \left(\mathbf{b}^S_{\mathcal{P}}\right)^T \,\right]^T.$$

In (57), $\mathbf{b}^s$ is determined by $\mathbf{U}^s_{\mathcal{J}}$, $\mathbf{n}^s$ and $\left\{\mathbf{f}^s_u : u \in \mathcal{I}\backslash\mathcal{J}\right\}$. Since the elements of $\mathbf{U}^s_{\mathcal{J}}$, $\mathbf{n}^s$ and $\left\{\mathbf{f}^s_u : u \in \mathcal{I}\backslash\mathcal{J}\right\}$ are independent, $\mathbf{b}^i$ and $\mathbf{b}^j$ are mutually independent for any $1 \le i \ne j \le S$. The covariance matrix of $\mathbf{b}$ is diagonal as shown in (38). Thus, applying the Scharf's theorem to $Z_{\mathcal{J}}$ completes the proof. ■

Appendix C
Proofs of Theorems 1, 2 and 3

A. Proof of Theorem 1

It is clear that K goes to infinity as N goes to infinity in the linear sparsity regime. Then, let $M = cK$ where $c > 1$. From (32),

$$\log \mathbb{P}\left\{\mathcal{E}^c_{\mathcal{I}}\right\} \le 2^{-1} SK(c-1) \underbrace{\left(\log(1+d_1) - d_1\right)}_{=:A} + \log 2$$

where $A < 0$ due to (33). Thus,

$$\lim_{N\to\infty} \mathbb{P}\left\{\mathcal{E}^c_{\mathcal{I}}\right\} \le \lim_{K\to\infty} \exp\left(2^{-1} SK(c-1)A + \log 2\right) = 0$$

implying that the probability that the correct support set is not δ jointly typical with all the measurement vectors vanishes.

Next, from (40),

$$\begin{aligned} \log \sum_{\mathcal{J}\in\mathcal{S}\backslash\mathcal{I}} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} &\le \log\left(\binom{N}{K} p\left(d_{2,\alpha^*} - 1\right)\right) \\ &= \log\binom{N}{K} + 2^{-1} SK(c-1) \underbrace{\left(\log(1-t) + t\right)}_{=:\gamma} \\ &\le K \underbrace{\left(1 + \log\frac{N}{K} + 2^{-1} S(c_1 - 1)\gamma\right)}_{=:\eta} \end{aligned} \quad (59)$$

where the last inequality is due to

$$\binom{N}{K} \le \exp\left(K \log \frac{Ne}{K}\right). \quad (60)$$

In (59), $\gamma < 0$ for any t where

$$t = \frac{1-\rho^{-1}}{1+\mathrm{SNR}^{-1}_{\min}} \in (0,1). \quad (61)$$

If $c > 1 + S^{-1} v_1$, then $\eta < 0$, which yields

$$\lim_{N\to\infty} \sum_{\mathcal{J}\in\mathcal{S}\backslash\mathcal{I}} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} \le \lim_{K\to\infty} \exp(K\eta) = 0$$

implying that the probability that all incorrect support sets are δ jointly typical with all the measurement vectors vanishes. Thus the failure probability p_{err} defined in (8) converges to zero if M satisfies (9).

Next, the remain is to derive (11) in the sublinear sparsity regime. Similarly, let $M = K + cK\log\frac{N}{K}$ wher $c > 1$. From (32),

$$\log \mathbb{P}\left\{\mathcal{E}^c_{\mathcal{I}}\right\} \le 2^{-1} ScK \log\frac{N}{K} \underbrace{\left(\log(1+d_1) - d_1\right)}_{=:A} + \log 2$$

where $A < 0$ due to (33). Thus,

$$\lim_{N\to\infty} \mathbb{P}\left\{\mathcal{E}^c_{\mathcal{I}}\right\} \le \lim_{N\to\infty} \exp\left(2^{-1} ScKA \log\frac{N}{K} + \log 2\right) = 0$$

implying that the probability that the correct support set is not δ jointly typical with all the measurement vectors vanishes.

Then, from (40),

$$\begin{aligned} \log \sum_{\mathcal{J}\in\mathcal{S}\backslash\mathcal{I}} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} &\le \log\left(\binom{N}{K} p\left(d_{2,\alpha^*} - 1\right)\right) \\ &= \log\binom{N}{K} + 2^{-1} ScK \underbrace{\left(\log(1-t) + t\right)}_{=:\gamma} \log\frac{N}{K} \\ &\le K \underbrace{\left(1 + 2^{-1} Sc\gamma\right)}_{=:\eta} \log\frac{N}{K} + K \end{aligned}$$

where the last inequality is due to the bound in (60) and $\gamma < 0$ for any t in (61). If $c > S^{-1} v_2$, then $\eta < 0$, which yields

$$\lim_{N\to\infty} \sum_{\mathcal{J}\in\mathcal{S}\backslash\mathcal{I}} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} \le \lim_{N\to\infty} \exp\left(K\eta \log\frac{N}{K} + K\right) = 0$$

implying that the probability that all incorrect support sets are δ jointly typical with all the measurement vectors vanishes. Thus, the failure probability p_{err} defined in (8) converges to zero if M satisfies (11), which completes the proof. ■

B. Proof of Theorem 2

From Lemma 1,

$$\mathbb{P}\left\{\mathcal{E}^c_{\mathcal{I}}\right\} \le 2 \left(\underbrace{\exp\left(-\frac{M-K}{2} d_1\right) (1+d_1)^{\frac{M-K}{2}}}_{=:\mu_{\mathcal{I}}} \right)^S. \quad (62)$$

If $M \geq K+1$, we have

$$\log \mu_{\mathcal{I}} = 2^{-1}(M-K)(\log(1+d_1)-d_1) < 0 \tag{63}$$

due to (33), which implies $\mu_{\mathcal{I}} < 1$. From Lemma 2,

$$\mathbb{P}\{\mathcal{E}_{\mathcal{J}}\} \leq \Bigg(\underbrace{\exp\left(-\frac{M-K}{2}\left(d_{2,\alpha^*}-1\right)\right) d_{2,\alpha^*}^{\frac{M-K}{2}}}_{=:\mu_{\mathcal{J}}}\Bigg)^S. \tag{64}$$

Similarly, if $M \geq K+1$, we have

$$\log \mu_{\mathcal{J}} = 2^{-1}(M-K)(\log(1-t)+t) < 0 \tag{65}$$

due to (61), which implies $\mu_{\mathcal{J}} < 1$. Thus, we conclude

$$\lim_{S\to\infty} p_{err} \leq 2\lim_{S\to\infty}\mu_{\mathcal{I}}^S + \binom{N}{K}\lim_{S\to\infty}\mu_{\mathcal{J}}^S = 0$$

for $M \geq K+1$ which completes the proof. ■

C. Proof of Theorem 3

The mutual information in (24) is bounded by

$$\begin{aligned}\mathbb{I}\left(\mathbf{x};\mathbf{y}|\tilde{\mathbf{F}}\right) &= h\left(\mathbf{y}|\tilde{\mathbf{F}}\right) - h\left(\mathbf{y}|\mathbf{x},\tilde{\mathbf{F}}\right) \leq h(\mathbf{y}) - h(\mathbf{n})\\ &\leq \sum\nolimits_{i=1}^{SM} h(y_i) - h(\mathbf{n})\\ &\leq 2^{-1}SM\left(\log\left(2\pi e\left(Kx_{\min}^2+\sigma^2\right)\right)-\log\left(2\pi e\sigma^2\right)\right)\\ &= 2^{-1}SM\log(1+K\times \mathrm{SNR}_{\min})\end{aligned}$$

where $h(\mathbf{x})$ is the differential entropy of $\mathbf{x}$, and $h(\mathbf{x}|\mathbf{y})$ is the conditional entropy of $\mathbf{x}$ given $\mathbf{y}$. The last inequality is due to that the Gaussian distribution maximizes the differential entropy. The denominator in (24) is bounded by

$$\log\left(\left|\mathcal{X}_{\{x_{\min}\}}\right|-1\right) = \log\left(\binom{N}{K}-1\right) > K\log\frac{N}{K}$$

for sufficiently large N. Then,

$$\begin{aligned}p_{err} &= \mathbb{E}_{\tilde{\mathbf{F}}}\sup_{\mathbf{x}\in\mathcal{X}_{x_{\min}}}\mathbb{P}\left\{\hat{\mathcal{I}}\neq\mathcal{I}\middle|\mathbf{x},\tilde{\mathbf{F}}\right\}\\ &\geq \mathbb{E}_{\tilde{\mathbf{F}}}\min_{\hat{\mathbf{x}}\in\mathcal{X}_{\{x_{\min}\}}}\max_{\mathbf{x}\in\mathcal{X}_{\{x_{\min}\}}}\mathbb{P}\left\{\hat{\mathbf{x}}\neq\mathbf{x}\middle|\mathbf{x},\tilde{\mathbf{F}}\right\}\\ &> 1-\frac{2^{-1}SM\log(1+K\times\mathrm{SNR}_{\min})+\log 2}{K\log\frac{N}{K}}.\end{aligned} \tag{66}$$

From (66), the failure probability is bounded away from zero by zero if (25) is satisfied, which completes the proof. ■

Appendix D
Proofs of Corollaries 1, 2 and 3

A. Proof of Corollary 1

From the inequality $\log(1+x) \leq \frac{2x}{2+x}$ for $x \in (-1,0]$,

$$v_2 = -\frac{2}{\log(1-t)+t} < \frac{4-2t}{t^2} < \frac{4}{t^2} \tag{67}$$

where t is defined in (61). Then,

$$\frac{v_2}{S} < \frac{4}{St^2}. \tag{68}$$

From (61),

$$\sqrt{S}t = \frac{1-\rho^{-1}}{\sqrt[-1]{S}+\left(\sqrt{S}\times\mathrm{SNR}_{\min}\right)^{-1}}. \tag{69}$$

Combining (11), (68) and (69) leads to (13). This approach is used to get (14) using the following equality

$$v_1 = v_2(1-\log\beta) \tag{70}$$

where $\lim_{N\to\infty}\frac{K}{N} = \beta \in (0,1/2)$, which completes the proof. ■

B. Proof of Corollary 2

Substituting $\alpha = \frac{2}{3}$ in (15), and rearranging the result with respect to t can yield $\frac{2}{3} \leq t < 1$, where t is defined in (61). Then from (67), a simple computation yields that

$$v_2 < \frac{4-2t}{t^2} \leq \frac{4}{t}$$

which immediately yields that

$$\frac{v_2}{S} < \frac{4}{St}. \tag{71}$$

where

$$St = \frac{1-\rho^{-1}}{S^{-1}+(S\times\mathrm{SNR}_{\min})^{-1}}. \tag{72}$$

Combining (11), (71) and (72) leads to (16). This approach is used to get (17) using (70), which completes the proof. ■

C. Proof of Corollary 3

We assume that $\mu_{\mathcal{I}} \geq \mu_{\mathcal{J}}$ and

$$p_{err} \leq \mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} + \sum_{\mathcal{J}\in\mathcal{S}\setminus\mathcal{I}}\mathbb{P}\{\mathcal{E}_{\mathcal{J}}\} \leq \left(\binom{N}{K}+2\right)\mu_{\mathcal{I}}^S < \varepsilon < 1. \tag{73}$$

Then, if the number of measurement vectors satisfies

$$S > \frac{\log\varepsilon - \log\left(\binom{N}{K}+2\right)}{\log\mu_{\mathcal{I}}} > 0, \tag{74}$$

(73) is achieved for small ε, and hence, reliable support set reconstruction is possible. If $\mu_{\mathcal{I}} < \mu_{\mathcal{J}}$, we obtain inequalities similar to (73) and (74) by replacing $\mu_{\mathcal{I}}$ by $\mu_{\mathcal{J}}$, where

$$S > \frac{\log\varepsilon - \log\left(\binom{N}{K}+2\right)}{\log\mu_{\mathcal{J}}} > 0. \tag{75}$$

Combining (74) and (75) yields (18).

Next, a simple computation yields that for any d_1 in (33),

$$\frac{\partial\log\mu_{\mathcal{I}}}{\partial d_1} = -\frac{d_1}{2(1+d_1)} < 0$$

where $\log\mu_{\mathcal{I}}$ is given in (63). From (33), we see $d_1 \propto \mathrm{SNR}_{\min}$ that leads to $\log\mu_{\mathcal{I}} \propto \mathrm{SNR}_{\min}^{-1}$. Also, for any t in (61),

$$\frac{\partial\log\mu_{\mathcal{J}}}{\partial t} = -\frac{t}{2(1-t)} < 0$$

where $\log\mu_{\mathcal{J}}$ is given in (65). From (61), we see $t \propto \mathrm{SNR}_{\min}$ that leads to $\log\mu_{\mathcal{J}} \propto \mathrm{SNR}_{\min}^{-1}$. Hence, the sufficient condition on S in (18) turns out to be a decreasing function with respect to $\mathrm{SNR}_{\min}$, which completes the proof. ■

APPENDIX E
PROOFS OF PROPOSITIONS 1 AND 2

First of all, we introduce the exponential inequalities [32], and use them in the proofs of Propositions 1 and 2.

A. The Exponential Inequalities [32]

Let Y_i, $i = 1, 2, \cdots, D$ be i.i.d. Gaussian variables with a zero mean and a unit variance. Then, let α_i, $i = 1, 2 \cdots, D$ be non-negative. We set

$$|\alpha|_\infty = \sup |\alpha_i|, \ |\alpha|_2^2 = \sum_{i=1}^{D} \alpha_i^2$$

and let

$$Y = \sum_{i=1}^{D} \alpha_i \left(Y_i^2 - 1\right). \tag{76}$$

Then, the following inequalities hold for any positive x

$$\mathbb{P}\left\{Y \geq 2|\alpha|_2 \sqrt{x} + 2|\alpha|_\infty x\right\} \leq \exp(-x) \tag{77}$$

$$\mathbb{P}\left\{Y \leq -2|\alpha|_2 \sqrt{x}\right\} \leq \exp(-x). \tag{78}$$

B. Proof of Proposition 1

In the proof of Lemma 3, $Z_{\mathcal{I}}$ is represented by

$$Z_{\mathcal{I}} = \sum_{s=1}^{S} \sum_{i=1}^{M-K} w^s(i)^2$$

where $w^s(i)$ is Gaussian with a zero mean and a unit variance. Define a random variable Y as

$$\begin{aligned} Y &= Z_{\mathcal{I}} - S(M-K) \\ &= \sum_{s=1}^{S} \sum_{i=1}^{M-K} \left(w^s(i)^2 - 1\right) \end{aligned}$$

which is of the form of (76). Then,

$$\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} = \underbrace{\mathbb{P}\left\{Y \leq -SM\delta/\sigma^2\right\}}_{=:A} + \underbrace{\mathbb{P}\left\{Y \geq SM\delta/\sigma^2\right\}}_{=:B}.$$

Combining A with (78) gives

$$\begin{aligned} \mathbb{P}\left\{Y \leq -SM\delta/\sigma^2\right\} &= \mathbb{P}\left\{Y \leq -2\sqrt{S(M-K)x}\right\} \\ &\leq \underbrace{\exp\left(-\frac{SM^2\delta^2}{4(M-K)\sigma^4}\right)}_{=:C} \end{aligned}$$

and combining B with (77) gives

$$\begin{aligned} \mathbb{P}\left\{Y \geq SM\delta/\sigma^2\right\} &= \mathbb{P}\left\{Y \geq 2\sqrt{S(M-K)x} + 2x\right\} \\ &\leq p_{1,\exp} \end{aligned}$$

where $p_{1,\exp}$ is defined in (28). It is readily seen that $p_{1,\exp} \geq C$, which leads to $\mathbb{P}\left\{\mathcal{E}_{\mathcal{I}}^c\right\} \leq 2p_{1,\exp}$.

Next, from (32) and (28),

$$\log p(d_1) = 2^{-1} S(M-K)(\log(1+d_1) - d_1)$$

and

$$\log p_{1,\exp} = -2^{-1} S(M-K) d_1^2 (2+4d_1)^{-1}$$

where $d_1 > 0$ is defined in (33). Then, we have

$$\log \frac{p(d_1)}{p_{1,\exp}} = \frac{S(M-K)}{2} \underbrace{\left(\log(1+d_1) - d_1 + d_1^2(2+4d_1)^{-1}\right)}_{=:g(d_1)}.$$

For any $d_1 > 0$, $\frac{\partial g(d_1)}{\partial d_1} = -d_1^2(2+3d_1)(1+d_1)^{-1}(1+2d_1)^{-2} < 0$ and $\max_{d_1>0} g(d_1) = 0$. Thus, we conclude $\log \frac{p(d_1)}{p_{1,\exp}} \leq 0$, which completes the proof. ■

C. Proof of Proposition 2

In the proof of Lemma 4, $Z_{\mathcal{J}}$ is represented by

$$\begin{aligned} Z_{\mathcal{J}} &= \sum_{s=1}^{S} \sum_{i=1}^{M-K} b^s(i)^2 \\ &= \sum_{s=1}^{S} \sum_{i=1}^{M-K} \alpha_{\mathcal{J},s} g^s(i)^2 \end{aligned}$$

where $\alpha_{\mathcal{J},s}$ is defined in (39) and $g^s(i)$ is Gaussian with a zero mean and a unit variance. Define a new random variable Y as

$$\begin{aligned} Y &= Z_{\mathcal{J}} - S(M-K) \\ &= \sum_{s=1}^{S} \sum_{i=1}^{M-K} \alpha_{\mathcal{J},s} \left(g^s(i)^2 - 1\right) \end{aligned}$$

which is of the form of (76). Then, from (44)

$$\begin{aligned} \mathbb{P}\left\{\mathcal{E}_{\mathcal{J}}\right\} &\leq \mathbb{P}\left\{Y < SM\delta - (M-K)\sum_{s=1}^{S} \left\|\mathbf{x}_{\mathcal{I}\setminus\mathcal{J}}^s\right\|_2^2\right\} \\ &\leq \mathbb{P}\left\{Y < \underbrace{SM\delta - S(M-K)x_{\min,\mathcal{J}}^2}_{=:A}\right\} \\ &\leq p_{2,\mathcal{J},\exp} \end{aligned} \tag{79}$$

where $p_{2,\mathcal{J},\exp}$ is defined in (30), the last inequality is due to (78). Due to (29), A is negative. Thus the exponential inequality of (78) gives the upper bound $p_{2,\mathcal{J},\exp}$.

Next, from (40) and (30),

$$\log p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right) = 2^{-1} S(M-K)(t + \log(1-t))$$

and

$$\begin{aligned} \log p_{2,\mathcal{J},\exp} &\geq -\frac{S(M-K)}{4}\left(\frac{x_{\min,\mathcal{J}}^2 - \frac{M\delta}{M-K}}{x_{\min,\mathcal{J}}^2 + \sigma^2}\right)^2 \\ &= -4^{-1} S(M-K) t^2 \end{aligned}$$

where $t \in (0,1)$, is defined in (61) and the inequality is due to (50). Then,

$$\log \frac{p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right)}{p_{2,\mathcal{J},\exp}} \leq \frac{S(M-K)}{4} \underbrace{\left(t^2 + 2t + 2\log(1-t)\right)}_{=:g(t)}.$$

For any $t \in (0,1)$, $\frac{\partial g(t)}{\partial t} = -2t^2(1-t)^{-1} < 0$ and $\max g(t) = 0$. We conclude $\log \frac{p\left(d_{2,\lambda_{\min}(\mathbf{R}_{\mathcal{J}})} - 1\right)}{p_{2,\mathcal{J},\exp}} \leq 0$. It completes the proof. ■

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[5] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[6] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
[7] E. J. Candès and M. B. Wakin, "An introduction to compressive sensing," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[8] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. (2009). "Distributed compressed sensing." [Online]. Available: https://arxiv.org/pdf/0901.3403.pdf
[9] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2005, pp. 1537–1541.
[10] C. Caione, D. Brunelli, and L. Benini, "Compressive sensing optimization for signal ensembles in WSNs," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 382–392, Feb. 2014.
[11] W. Chen, R. D. Rodrigues, and I. J. Wassell, "Distributed compressive sensing reconstruction via common support discovery," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–5.
[12] J. Oliver, W.-B. Lee, and H.-N. Lee, "Filters with random transmittance for improving resolution in filter-array-based spectrometers," *Opt. Exp.*, vol. 21, no. 4, pp. 3969–3989, Feb. 2013.
[13] S. Goginneni and A. Nehorai, "Target estimation using sparse modeling for distributed MIMO radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5315–5325, Nov. 2011.
[14] Y. Wu *et al.*, "Accelerated MR diffusion tensor imaging using distributed compressed sensing," *Magn. Reson. Med.*, vol. 71, no. 2, pp. 763–772, Feb. 2014.
[15] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.
[16] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, May 2012.
[17] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, Apr. 2015.
[18] M. J. 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.
[19] 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.
[20] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, Oct. 2010.
[21] 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.
[22] J. Scarlett, J. S. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 427–439, Jan. 2013.
[23] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593–620, Jan. 2017.
[24] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.
[25] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 4634–4643, Dec. 2006.
[26] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.
[27] Y. Jin and B. D. Rao, "Support recovery of sparse signals in the presence of multiple measurement Vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3139–3157, May 2013.
[28] M. F. Duarte, M. B. Wakin, D. Braon, S. Sarvotham, and R. G. Baraniuk, "Measurement bounds for sparse signal ensembles via graphical models," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4280–4289, Jul. 2013.
[29] J. Kim, O. Lee, and J. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 278–301, Jan. 2012.
[30] J. D. Blanchard and M. E. Davies, "Recovery guarantees for rank aware pursuits," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 427–430, Jul. 2012.
[31] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York, NY, USA: Wiley, 2001.
[32] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1303–1338, Oct. 2000.
[33] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 1991.
[34] R. Rajagopalan and P. K. Varsheny, "Data-aggregation techniques in sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 8, no. 4, pp. 48–63, 4th Quart., 2006.

Sangjun Park received the B.S. degree in computer engineering from the Chungnam National University, Daejeon, Korea, in 2009. He is pursuing a Ph.D degree at the School of Electrical Engineering and Computer Science in the Gwangju Institute of Science and Technology (GIST), Gwangju, Korea. His research interests include information theory, numerical optimization and compressed sensing.

Nam Yul Yu (M'07) received the B.S. degree in electronics engineering from the Seoul National University, Seoul, Korea, in 1995 and the M.S. degree in electronic and electrical engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 2000. He received the Ph. D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 2007. From 2000 to 2003, he was with Telecommunication Research and Development Center, Samsung Electronics, Korea, where he worked on channel coding schemes for wireless communication systems. In 2007, he was a senior research engineer in LG Electronics, Korea, working on the standardization of the 3GPP-LTE. From 2008 to 2014, he has been working as an Assistant/Associate Professor in the Department of Electrical Engineering at the Lakehead University, Thunder Bay, Ontario, Canada. In 2014, he joined the Gwangju Institute of Science and Technology (GIST), Gwangju, Korea, and is currently working as an Associate Professor in the School of Electrical Engineering and Computer Science. From 2009 to 2011, he served as an Associate Editor for Sequences in the IEEE Transactions on Information Theory. His research interests include sequence design, compressed sensing, and security for wireless communications.

Heung-No Lee (SM'13) received the B.S., M.S., and Ph.D. degrees from the University of California, Los Angeles, CA, USA, in 1993, 1994, and 1999, respectively, all in electrical engineering. He then worked at HRL Laboratories, LLC, Malibu, CA, USA, as a Research Staff Member from 1999 to 2002. From 2002 to 2008, he worked as an Assistant Professor at the University of Pittsburgh, PA, USA. In 2009, he then moved to the School of Electrical Engineering and Computer Science, GIST, Korea, where he is currently affiliated. His areas of research include information theory, signal processing theory, communications/networking theory, and their application to wireless communications and networking, compressive sensing, future internet, and brain-computer interface. He has received several prestigious national awards, including the Top 100 National Research and Development Award in 2012, the Top 50 Achievements of Fundamental Researches Award in 2013, and the Science/Engineer of the Month (January 2014). He has written more than fifty international journal publications and a hundred international conferences and workshop papers. He was the Director of Electrical Engineering and Computer Science track within GIST College in 2014. In March 2015, he was appointed as the Dean of Research at GIST.