

# Compressive MUSIC with Optimized Partial Support for Joint Sparse Recovery

Jong Min Kim, Ok Kyun Lee and Jong Chul Ye

Department of Bio and Brain Engineering, KAIST, Guseongdong, Daejeon, 305-701, Korea

Email: franzkim@gmail.com, jong.ye@kaist.ac.kr

**Abstract**—The multiple measurement vector (MMV) problem addresses the identification of unknown input vectors that share common sparse support. The MMV problem has been traditionally addressed either by sensor array signal processing or compressive sensing. However, recent breakthroughs in this area such as compressive MUSIC (CS-MUSIC) or subspace-augmented MUSIC (SA-MUSIC) optimally combine the compressive sensing (CS) and array signal processing such that  $k-r$  supports are first found by CS and the remaining  $r$  supports are determined by a generalized MUSIC criterion, where  $k$  and  $r$  denote the sparsity and the number of independent snapshots, respectively. Even though such a hybrid approach significantly outperforms the conventional algorithms, its performance heavily depends on the correct identification of  $k-r$  partial support by the compressive sensing step, which often deteriorates the overall performance. The main contribution of this paper is, therefore, to show that as long as  $k-r+1$  correct supports are included in any  $k$ -sparse CS solution, the optimal  $k-r$  partial support can be found using a subspace fitting criterion, significantly improving the overall performance of CS-MUSIC. Furthermore, unlike the single measurement CS counterpart that requires infinite SNR for a perfect support recovery, we can derive an information theoretic sufficient condition for the perfect recovery using CS-MUSIC under a *finite* SNR scenario.

## I. INTRODUCTION

One of important areas of compressed sensing research is the so-called multiple measurement vector problem (MMV) [1–3, 5–9]. The MMV problem addresses the recovery of a set of sparse signal vectors that share common non-zero support. In MMV, thanks to the common sparse support, it is quite predictable that the number of recoverable sparsity levels may increase with the increasing number of measurement vectors. However, the performance of the existing MMV compressive sensing algorithms are not generally satisfactory even for a noiseless case when a finite number of snapshots are available.

A recent breakthrough in this area has created a new class of algorithms such as compressive MUSIC (CS-MUSIC) proposed by our group [5] or subspace-augmented MUSIC (SA-MUSIC) proposed independently [8]. Specifically, when the number of targets is  $k$ , and  $r$  independent snapshots are available, compressive MUSIC finds  $k-r$  targets using a compressive sensing algorithm such as S-OMP or  $p$ -thresholding, and the remaining  $r$  targets are recovered using a generalized MUSIC criterion [5]. This hybridization significantly improves the performance of estimating jointly sparse signals and achieves the  $l_0$  sparse recovery bound using a finite number of snapshots. Furthermore, even if the sparsity level is not known *a priori*, compressive MUSIC can

accurately estimate the sparsity level using the generalized MUSIC criterion. In spite of its success, one of the main shortcomings of CS-MUSIC or SA-MUSIC is that the overall performance is heavily dependent upon the success of the first  $k-r$  support estimation. This is especially problematic when the measurement is so noisy or the RIP condition for the sensing matrix is so bad that the greedy  $k-r$  update step may produce incorrect support estimate.

One of the main contributions of this paper is, therefore, to relax this stringent requirement. In particular, the new algorithm requires that  $k-r+1$  supports (not in sequential order) out of  $k$  support estimation is correct rather than  $k-r$  consecutive support estimate are correct. The location of the unknown  $k-r$  true support can be then readily estimated using a subspace fitting criterion. Such optimized partial support estimates can significantly improve the accuracy of the generalized MUSIC step, hence overall performance of compressive MUSIC.

The paradigm shift from early termination of CS algorithm after  $k-r$  step to selecting the correct  $k-r$  supports out of  $k$ -sparse solution by any CS algorithm is much more significant and fundamental than just regarded as algorithmic improvement. In particular, by converting the problem as a partial support recovery problem, we can adapt rich information theoretical analysis tools that have been developed for single measurement vector CS (SMV-CS) [12]. In particular, we can derive an information theoretic sufficient condition for the perfect recovery of CS-MUSIC under a *finite* SNR scenario, which was considered not feasible in the SMV-CS [12].

## II. PROBLEM FORMULATION AND MATHEMATICAL PRELIMINARIES

Throughout the paper,  $\mathbf{x}^i$  and  $\mathbf{x}_j$  correspond to the  $i$ -th row and the  $j$ -th column of matrix  $X$ , respectively. When  $S$  is an index set,  $X^S$ ,  $A_S$  corresponds to a submatrix collecting corresponding rows of  $X$  and columns of  $A$ , respectively. The following canonical MMV formulation is very useful for our analysis.

*Definition 2.1 (Canonical form MMV [5]):* Let  $m$ ,  $n$  and  $r$  be a positive integers ( $m < n$ ) that represents the number of sensor elements, the ambient space dimension, and the number of snapshots, respectively. Suppose that we are given a multiple-measurement vector  $B \in \mathbb{R}^{m \times r}$ ,  $X = [\mathbf{x}_1, \dots, \mathbf{x}_r] \in \mathbb{R}^{n \times r}$ , and a sensing matrix  $A \in \mathbb{R}^{m \times n}$ . A canonical form

MMV problem is given by the following optimization problem:

$$\begin{aligned} & \text{minimize} \quad \|X\|_0 \\ & \text{subject to} \quad B = AX, \end{aligned} \quad (1)$$

where  $\|X\|_0 = |\text{supp}X|$ ,  $\text{supp}X = \{1 \leq i \leq n : \mathbf{x}^i \neq 0\}$ ,  $\mathbf{x}^i$  is the  $i$ -th row of  $X$ , and the measurement matrix  $B$  is full rank, i.e.  $\text{rank}(B) = r \leq \|X\|_0$ .

Note that the canonical form MMV has the additional constraint that  $\text{rank}(B) = r \leq \|X\|_0$ . This is not problematic at all since every MMV problem can be converted into canonical form using the singular value decomposition [5]. Now, the following theorem provides the  $l_0$  sparse recovery bound from noiseless measurements.

*Theorem 2.1 ( $l_0$  Bound):* [1–3] Let  $\text{spark}(A)$  denote the smallest number of linearly dependent columns of  $A$ . Then,  $X \in \mathbb{R}^{n \times r}$  is the unique solution of  $AX = B$  if and only if

$$\|X\|_0 < \frac{\text{spark}(A) + \text{rank}(B) - 1}{2} \leq \text{spark}(A) - 1. \quad (2)$$

### III. COMPRESSIVE MUSIC

Consider a canonical form MMV problem. Suppose, furthermore, that the columns of a sensing matrix  $A \in \mathbb{R}^{m \times n}$  are in general position; that is, any collection of  $m$  columns of  $A$  are linearly independent. Then, according to [3, 13], for any  $j \in \{1, \dots, n\}$ ,  $j \in \text{supp}X$  if and only if

$$Q^* \mathbf{a}_j = 0, \quad (3)$$

where  $Q \in \mathbb{R}^{m \times (m-r)}$  consists of orthonormal columns such that  $Q^*B = 0$  so that  $R(Q)^\perp = R(B)$ , which is often called “noise subspace”.

Note that the MUSIC criterion (3) holds for all  $m \geq k + 1$  if the columns of  $A$  are in general position. Using the compressive sensing terminology, this implies that the recoverable sparsity level by MUSIC (with a probability 1 for the noiseless measurement case) is given by

$$\|X\|_0 < m = \text{spark}(A) - 1, \quad (4)$$

where the last equality comes from the definition of the spark. Therefore, the  $l_0$  bound (2) can be achieved by MUSIC when  $r = k$ . However, for any  $r < k$ , the MUSIC condition (3) does not hold. This is a major drawback of MUSIC compared to the compressive sensing algorithms that allows perfect reconstruction with extremely large probability by increasing the sensor elements  $m$ . This drawback of the conventional MUSIC can be overcome by the following generalized MUSIC criterion [5].

*Theorem 3.1:* [5] Assume that  $A \in \mathbb{R}^{m \times n}$ ,  $X \in \mathbb{R}^{n \times r}$ , and  $B \in \mathbb{R}^{m \times r}$  satisfy  $AX = B$ . Furthermore, we assume that  $\|X\|_0 = k$  and  $A$  satisfies the RIP condition with the left RIP constant  $\delta_{2k-r+1}^L < 1$ . If we are given  $I_{k-r} \subset \text{supp}X$  with  $|I_{k-r}| = k - r$  and  $A_{I_{k-r}} \in \mathbb{R}^{m \times (k-r)}$ , which consists of columns whose indices are in  $I_{k-r}$ , then for any

$$j \in \{1, \dots, n\} \setminus I_{k-r},$$

$$\mathbf{a}_j^* \left[ P_{R(Q)} - P_{R(P_{R(Q)} A_{I_{k-r}})} \right] \mathbf{a}_j = 0 \quad (5)$$

if and only if  $j \in \text{supp}X$ .

Note that when  $r = k$ , the condition (5) is the same as the MUSIC criterion (3). By Theorem 3.1, we can develop the Compressive MUSIC algorithm, which can be executed by these processes.

- Step 1: Find  $k - r$  indices of  $\text{supp}X$  by any MMV compressive sensing algorithms such as 2-thresholding or SOMP.
- Step 2: Let  $I_{k-r}$  be the set of indices which are taken in Step 1 and  $S = I_{k-r}$ .
- Step 3: For  $j \in \{1, \dots, n\} \setminus I_{k-r}$ , calculate the quantities  $\eta(j) = \mathbf{a}_j^* [P_{R(Q)} - P_{R(P_{R(Q)} A_{I_{k-r}})}] \mathbf{a}_j$  for all  $j \notin I_{k-r}$ .
- Step 4: Make an ascending ordering of  $\eta(j)$ ,  $j \notin I_{k-r}$  and choose indices that correspond to the first  $r$  elements and put these indices into  $S$ .

In compressive MUSIC, we determine  $k - r$  indices of  $\text{supp}X$  with CS-based algorithms such as 2-thresholding or S-OMP, where the exact identification of  $k - r$  indices is a probabilistic matter. After that process, we recover remaining  $r$  indices of  $\text{supp}X$  with a generalized MUSIC criterion, which is given in Theorem 3.1, and this reconstruction process is deterministic. This hybridization makes the compressive MUSIC applicable for all ranges of  $r$ , outperforming all the existing methods.

So far, we introduced the compressive MUSIC algorithm. To analyze the performance of the compressive MUSIC, we find the number of measurements with which we can identify the support of  $X$  by using compressive MUSIC with S-OMP. For this purpose, we consider the large system limit so that we assume the following conditions.

- Let  $\rho := \lim_{n \rightarrow \infty} m(n)/n$  exist. Then we call  $\rho$  as the asymptotic under-sampling rate.
- Let  $\epsilon := \lim_{n \rightarrow \infty} k(n)/n$  exist. Then we call  $\epsilon$  as the asymptotic sparsity.
- Let  $\alpha := \lim_{n \rightarrow \infty} r(n)/k(n)$  exist and  $\gamma := \lim_{n \rightarrow \infty} \sqrt{k/m}$ .

Now, we may consider two cases according to the number of multiple measurement vectors. First, we consider the case when the number of multiple measurement vectors are finite fixed number. Conventional compressive sensing (SMV problem) is a kind of this case. Second, we consider the case when  $r$  is proportional to  $n$ . This case includes the conventional MUSIC case. To analyze S-OMP, we assume that each element of  $A$  is i.i.d. Gaussian random variable  $\mathcal{N}(0, 1/m)$ . In analyzing S-OMP, rather than analyzing the distribution of  $\|\mathbf{a}_j^* P_{R(A_{I_t})}^\perp B\|_F^2$  where  $I_t$  denotes the set of indices which are chosen in the first  $t$  step of S-OMP, we consider the following version of subspace S-OMP due to its better performance [2, 8].

- Step 1 : Initialize  $t = 0$  and  $I_0 = \emptyset$ .
- Step 2 : Compute  $P_{R(A_{I_t})}^\perp$  which is the projection operator onto the orthogonal complement of the span of

- $\{\mathbf{a}_j : j \in I_t\}$ .
- Step 3 : Compute  $P_{R(A_{I_t})}^\perp B$  and for all  $j = 1, \dots, n$ , compute  $\rho(t, j) = \|\mathbf{a}_j^* P_{R(A_{I_t})}^\perp B\|_F^2$ .
- Step 4 : Take  $j_t = \arg \max_{j=1, \dots, n} \rho(t, j)$  and  $I_{t+1} = I_t \cup \{j_t\}$  and if  $t < k - r$  return to Step 2.
- Step 5 : The final estimate of the  $k - r$  elements of support is  $I_{k-r}$ .

*Theorem 3.2:* Assume that we have multiple measurements  $B = AX$  where each element of  $A$  is generated from i.i.d.  $\mathcal{N}(0, 1/m)$  and  $N$  is an additive noise. Then, in the large system limit, with probability 1, we can identify  $k - r$  elements of the support of  $X$  with subspace S-OMP if we have one of the following conditions :

1.  $r$  is a fixed finite number and

$$m > k(1 + \delta) \frac{2 \log(n - k)}{r}$$

for some  $\delta > 0$ .

2.  $r$  satisfies  $\lim_{n \rightarrow \infty} (\log n)/r = 0$ ,  $\lim_{n \rightarrow \infty} r/k = \alpha$  and

$$m > k(1 + \delta)^2 [2 - F(\alpha)]^2$$

for some  $\delta > 0$  where

$$F(\alpha) = \frac{1}{\alpha} \int_0^{4t_1(\alpha)^2} x d\lambda_1(x),$$

$d\lambda_1(x) = (\sqrt{(4-x)x})/(2\pi x)$  is the probability measure with support  $[0, 4]$ ,  $0 \leq t_1(\alpha) \leq 1$  satisfies  $\int_0^{4t_1(\alpha)^2} d\lambda_1(x) = \alpha$ .

Here,  $F(\alpha)$  is an increasing function on  $(0, 1]$  such that  $F(1) = 1$  and  $\lim_{\alpha \rightarrow 0^+} F(\alpha) = 0$ .

*Proof:* See Appendix A. ■

By above theorem, the number of measurements for S-OMP shows some different characteristics according to the number of the measurement vectors. First, if we have small number of multiple measurement vectors, then the number of samples for S-OMP is reciprocally proportional to the number of multiple measurement vectors. On the other hand, we have sufficiently large number of snapshots such that  $\lim_{n \rightarrow \infty} (\log n)/r$  is close to 0, then the number of measurements for S-OMP varies from  $4k$  to  $k$  according to the ratio of  $r$  and  $k$  so that the  $\log n$  is not necessary. In particular, if the number of snapshots approaches the sparsity  $k$ , then we can identify the indices of  $\text{supp}X$  with only  $(1 + \delta)k$  where  $\delta$  is any small positive number, which is equivalent to the required number of multiple measurement vectors for the success of conventional MUSIC algorithm.

Furthermore, in [5], we developed the analysis for the noisy setting, where we showed that the required SNR for the success of support recovery decreases when the asymptotic ratio of the number of snapshots and the sparsity level (that is,  $\lim_{n \rightarrow \infty} r/k$ ) increases, in the large system limit. This is one of the important advantages of MMV over SMV.

#### IV. OPTIMIZED PARTIAL SUPPORT SELECTION

As discussed before, we can easily expect that the performance of the compressive MUSIC is very dependent on the

selection of  $k - r$  correct indices of the support of  $X$ . Note that this is a very stringent condition. In practice, even though the consecutive  $k - r$  steps of S-OMP may not be correct, there are chances that among the  $k$ -sparse solution of S-OMP, part of the supports can be correct. Hence, if the estimate of the support of  $X$  has at least  $k - r$  indices of the support of  $X$  and we can identify them, then we can expect that the performance of the compressive MUSIC will be improved. When  $\binom{k}{k-r}$  is small, we may apply the exhaustive search, but if both  $k - r$  and  $r$  are not small, then the exhaustive search is hard to apply so that we have to find some alternative method to identify the correct indices from the estimate of  $\text{supp}X$ . Indeed, the following subspace fitting criterion can address the problem.

*Theorem 4.1:* Assume that we have a canonical MMV model  $AX = B$  where  $A \in \mathbb{R}^{m \times n}$ ,  $X \in \mathbb{R}^{n \times r}$ ,  $\|X\|_0 = k$  and  $r < k < m < n$ . If there is an index set  $I_k \subset \{1, \dots, n\}$  such that  $|I_k| = \min\{k, \text{spark}(A) - r\}$  and  $|I_k \cap \text{supp}X| \geq k - r + 1$ , then for any  $j \in I_k$ ,  $j \in \text{supp}X$  if and only if

$$P_{Q_{k,j}} \mathbf{a}_j = \mathbf{0}, \quad (6)$$

where  $Q_{k,j}$  is the orthogonal complement for  $R([B \ A_{I_k \setminus \{j\}}])$ ,  $A_{I_k \setminus \{j\}}$  consists of columns of  $A$  whose index belongs to  $I_k \setminus \{j\}$  and  $P_{R([B \ A_{I_k \setminus \{j\}}])}^\perp$  is the orthogonal projection on  $R([B \ A_{I_k \setminus \{j\}}])^\perp$ .

*Proof:* Assume that  $j \in I_k \cap \text{supp}X$ . Then  $|(I_k \setminus \{j\}) \cap \text{supp}X| \geq k - r$  so that

$$R([B \ A_{I_k \setminus \{j\}}]) \supseteq R([B \ A_{J_{j,k-r}}]) \cap R(A_S) = R(A_S)$$

where  $J_{j,k-r} \subset (I_k \setminus \{j\}) \cap S$ ,  $|J_{j,k-r}| = k - r$  and  $S = \text{supp}X$ . Since  $\mathbf{a}_j \in R(A_S)$ , (6) holds for  $j \in I_k \cap \text{supp}X$ .

To show the converse, assume that (6) holds for some  $j \in I_k$ . Then we have  $\mathbf{a}_j \in R([B \ A_{I_k \setminus \{j\}}])$ , that is, there some  $\mathbf{p} \in \mathbb{R}^r$  and  $\mathbf{q} \in \mathbb{R}^{|I_k| - 1}$  such that

$$\mathbf{a}_j = B\mathbf{p} + A_{I_k \setminus \{j\}}\mathbf{q} = AX\mathbf{p} + A_{I_k \setminus \{j\}}\mathbf{q}.$$

Since  $|( \text{supp}X \cup I_k )| \leq k + |I_k| - (k - r + 1) \leq k + \text{spark}(A) - r - (k - r + 1) = \text{spark}(A) - 1$ , if  $j \notin \text{supp}X$ , then there is an  $\mathbf{r} \in \mathbb{R}^n \setminus \{0\}$  such that  $\|\mathbf{r}\|_0 < \text{spark}(A)$  and  $A\mathbf{r} = \mathbf{0}$  since  $j \notin \text{supp}X \cup (I_k \setminus \{j\})$ . Then, by the definition of  $\text{spark}(A)$ , that is a contradiction so that  $j \in \text{supp}X$  if (6) holds. ■

In particular, if the columns of  $A$  are in general position, then we can take index set  $I_k$  with  $|I_k| = \min\{k, m - r + 1\}$ . Also, if  $A$  has an RIP condition with  $\delta_{2k} < 1$ , then we can take  $|I_k| = k$  since  $r \leq k$ . Then, Theorem 4.1 informs us that we only require the partial support recovery rather than  $k - r$  consecutive correct CS step [5]. Accordingly, the compressive MUSIC with optimized partial support is then performed by following procedure.

- step 1 : Let  $S = \emptyset$ .
  - If  $r < k$ , estimate  $k$  indices of  $\text{supp}X$  by MMV compressive sensing algorithm.
  - If  $r = k$ , goto step 5.
- step 2 : Let  $I_k$  be the set of indices which are taken in step 1.

- step 3 : For  $j \in I_k$ , calculate the quantities  $\zeta(j) = \|P_{Q_{k,j}} \mathbf{a}_j\|^2$ .
- step 4 : Make an ascending ordering of  $\zeta(j)$ ,  $j \in I_k$  and choose indices that corresponds the first  $k - r$  elements and put these indices into  $S$ .
- step 5 : For  $j \in \{1, \dots, n\} \setminus S$ , calculate the quantities  $\eta(j) = \mathbf{g}_j^* P_{G_{I_k-r}}^\perp \mathbf{g}_j$ .
- step 6 : Make an ascending ordering of  $\eta(j)$ ,  $j \notin S$  and choose indices that correspond to the first  $r$  elements and put these indices into  $S$ .

In above algorithm, we require partial correctness of support estimation instead of exactness of  $k - r$  consecutive support estimation. Moreover, the step 1 in above algorithm need not to be greedy so that we can also apply the convex optimization algorithm such as  $l_{2,1}$  minimization [9] or belief propagation [7].

So far, we have assumed that the measurement  $B$  is without noise. For the case of noisy measurement,  $B$  is corrupted so that the optimized partial support selection is affected by noise. Although we do not discuss the noise sensitivity in this paper, this issue will be investigated in the future works.

#### V. INFORMATION THEORETIC ANALYSIS FOR PARTIAL SUPPORT RECOVERY FOR MMV

From above section, we know that compressive MUSIC with optimized partial support can bear with the fractional distortion of support estimate error less than  $\alpha$  to guarantee the exact recovery in the large system limit. Therefore, in this section, we are interested in finding a sufficient condition such that we can find the estimate for the support with fractional distortion less than  $\alpha$  in an MMV step. Here, we consider the linear model in which the multiple measurement  $Y \in \mathbb{R}^{m \times r}$  is given as

$$Y = AX + N$$

where  $A \in \mathbb{R}^{m \times n}$  is a sensing matrix and  $N \in \mathbb{R}^{m \times r}$  is additive noise whose columns are i.i.d. and have the distribution  $\mathcal{N}(0, \sigma_w^2 I)$ . Also we assume that  $X$  has  $k$  nonzero rows which are indexed by the set  $S$  and that  $S$  is distributed uniformly over the  $\binom{n}{k}$  possibilities. Again, we assume that the distributions of each column of  $X$  are identical and independent. Furthermore, we assume that the elements of sensing matrix  $A$  are randomly given with i.i.d.  $\mathcal{N}(0, 1/n)$ . Here we consider the large system limit. Also, we use the following definition for SNR.

*Definition 5.1:* For a given multiple signal  $X$ , the SNR is given by

$$\text{SNR}(X) = \frac{\mathbb{E}[\|AX\|_F^2]}{\mathbb{E}[\|N\|_F^2]} = \frac{\|X\|_F^2}{rn\sigma_w^2}.$$

Also, for a stochastic signal class  $\mathcal{X}$ ,  $\text{SNR}(\mathcal{X})$  is called an asymptotic lower bound on  $\text{SNR}(X)$  if there exists a constant  $c > 0$  such that

$$\mathbb{P}\{\text{SNR}(\mathcal{X}(n)) \leq \text{SNR}(X)\} > 1 - e^{-nc}.$$

The analysis for partial support recovery use an information theoretic approach which was used in [12] so that we define

the following function.

*Definition 5.2:* For  $p \in [0, 1]$  and  $u \in [0, 1 - \epsilon]$ , we define

$$h(\epsilon, \alpha) = \epsilon h(\alpha) + (1 - \epsilon) h\left(\frac{\alpha}{1/\epsilon - 1}\right),$$

where  $h(p) = -p \log p - (1 - p) \log(1 - p)$  is the binary entropy function.

For a fractional distortion  $\alpha > 0$ , we define the fractional partial recovery with distortion rate  $\alpha$  by the requirement  $d(S, \hat{S})/k \leq \alpha$  where  $\hat{S}$  is the estimate for the support of  $X$  such that  $|\hat{S}| = k$  and  $d(S, \hat{S}) = |S \setminus \hat{S}|$ . If  $\alpha > 1 - \epsilon$ , the random guessing estimator  $\hat{K}_{RG}$  is asymptotically reliable so that we assume that  $\alpha \leq 1 - \epsilon$  [12].

For the analysis, we consider the maximum likelihood (ML) estimator which is given by

$$\hat{S}_{ML}(Y) = \arg \min_{|U|=k} \|P_{R(A_U)}^\perp Y\|_F^2$$

where  $P_{R(A_U)}^\perp$  is the projection operator onto the orthogonal complement of  $R(A_U)$ . For  $k$ -sparse multiple input signal  $X \in \mathbb{R}^{n \times r}$ , we introduce the following term.

*Definition 5.3:* Let  $\mathbf{Z}$  correspond to the nonzero rows of  $X$  and satisfy  $\|\mathbf{z}^1\|_2 \leq \|\mathbf{z}^2\|_2 \leq \dots \leq \|\mathbf{z}^k\|_2$ . Then, for some  $\alpha \leq 1$ , we let

$$g(\alpha, X) = \frac{1}{\alpha \|X\|_F^2} \sum_{i=1}^{\lfloor \alpha k \rfloor} \|\mathbf{z}^i\|_2^2.$$

Also, for a stochastic signal class  $\mathcal{X}$ , let  $g(\alpha, \mathcal{X})$  be the asymptotic lower bound on  $g(\alpha, X)$  if there is a constant  $c > 0$  such that

$$\mathbb{P}\{g(\alpha, \mathcal{X}(n)) \leq g(\alpha, X)\} > 1 - e^{-nc}.$$

In [12], Reeves and Gastpar gave sufficient conditions for partial support recovery for SMV problem using ML estimator. We can extend those results to the MMV problem as the following theorem.

*Theorem 5.1:* For a given signal class  $\mathcal{X}$ , sparsity  $\epsilon \in (0, 1)$ , undersampling ratio  $\rho < 1$ , the fractional distortion  $\alpha \in (0, 1 - \epsilon)$ , the estimator  $\hat{S}_{ML}$  is asymptotically reliable if

$$\text{SNR}(\mathcal{X}) > \frac{1}{\alpha g(\alpha, \mathcal{X})} \quad (7)$$

and

$$\rho > \epsilon + \frac{1}{r} \max_{u \in [\alpha, 1 - \epsilon]} \frac{2h(\epsilon, u)}{\log(\gamma(u, \mathcal{X})) + \gamma(u, \mathcal{X})^{-1} - 1} \quad (8)$$

where  $\gamma(u, \mathcal{X}) = \text{SNR}(\mathcal{X}) u g(u, \mathcal{X})$ .

*Proof:* See Appendix B. ■

Note that if  $\alpha > 1 - \epsilon$ , the random guessing estimator is asymptotically reliable so that we can identify the support with distortion less than  $\alpha$  with large probability in the large system case, by augmenting randomly chosen  $k - r$  support in generalized MUSIC step. Moreover, in this case, the sufficient condition becomes  $\rho > \epsilon$ , which is equivalent to the MUSIC for the full rank measurement.

In addition, in [12], Reeves and Gastpar gave necessary conditions for partial support recovery for SMV problem. The counterpart for MMV can be given by the following theorem. For the proof, see Appendix C.

*Theorem 5.2:* For a given stochastic signal class  $\mathcal{X}$ , sparsity  $\epsilon \in (0, 1)$ , sampling rate  $\rho < 1$  and fractional distortion  $\alpha \in (0, 1 - \epsilon)$ , a necessary condition for asymptotically reliable recovery is

$$\rho > \frac{h(\epsilon) - h(\epsilon, \alpha) + I(X; Y|S)/n}{\sum_{l=1}^r \frac{1}{2} \log \left( 1 + \frac{1}{\sigma_w^2} \kappa_l(\mathcal{X}) \right)},$$

where  $I(X; Y|K)$  is the mutual information between  $X$  and  $Y$  conditioned on  $S$ , and  $\kappa_l(\mathcal{X})$  is the asymptotic upper bound for the  $l$ -th largest eigenvalue of  $X^*X$ , where  $X \in \mathcal{X}$ .

## VI. NUMERICAL SIMULATION

We compared the performance of compressive MUSIC with optimized partial support (proposed algorithm), compressive MUSIC (CS-MUSIC), subspace-augmented MUSIC (SA-MUSIC) and S-OMP. We used S-OMP as a MMV compressive sensing algorithm for various hybrid MMV algorithms. In order to quantify the performance of each algorithms, the empirical recovery ratio is calculated which is defined as the percentage of correct identification of all supports, and the ratio are averaged for 5000 simulation results. The simulation parameters are as following:  $m = 40$ ,  $n = 100$ , the number of measurement vectors is  $r = 9$ , and  $k = 1, 2, \dots, 20$ , respectively. Each component of the sensing matrix  $A$  is generated by i.i.d. Gaussian random variable  $\frac{1}{\sqrt{m}}\mathcal{N}(0, 1)$  or  $\frac{1}{\sqrt{m}}\mathcal{N}(1, 1)$  to see the effect of RIP in each algorithms. Gaussian noise of SNR = 40dB is added to the measurement vector  $B$ . In Figure 1, we can observe that the proposed method shows significantly better performance than the original version of compressive MUSIC, SA-MUSIC and S-OMP. In particular, the proposed method is more robust to bad RIP of the sensing matrices such that the performance gain is more prominent.

## VII. CONCLUSION

This paper proposed a mathematical framework for optimized partial support selection to improve the performance of compressive MUSIC for joint sparse recovery. We first discussed about the original compressive MUSIC algorithm, and derived the sharp bound for the number of measurement for exact recovery using subspace S-OMP for partial support recovery. Then, we discussed that the requirement of the correct  $k - r$  step S-OMP can be relaxed such that as long as  $k - r + 1$  supports from  $k$ -support estimate are correct, subspace fitting criterion can identify the correct  $k - r$  support to improve the robustness of the compressive MUSIC algorithm. Information theoretical analysis was also provided to obtain a sufficient condition for MMV joint sparse recovery using compressive MUSIC algorithm. As a future work, we will derive the SNR condition for the success of subspace fitting step.

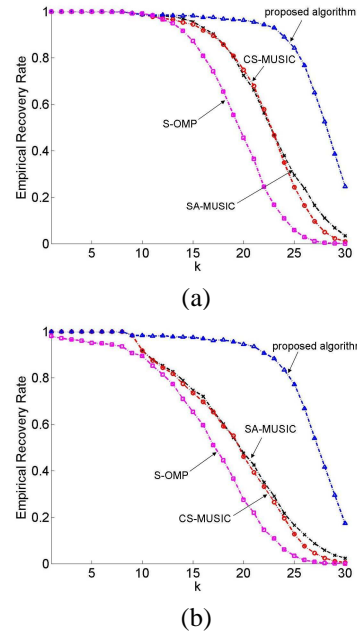


Fig. 1. Recovery rates when  $m = 40$ ,  $r = 9$ , SNR = 40dB and  $A$  is generated from (a)  $\frac{1}{\sqrt{m}}\mathcal{N}(0, 1)$  and (b)  $\frac{1}{\sqrt{m}}\mathcal{N}(1, 1)$ .

## ACKNOWLEDGMENT

This research was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MEST) (No. 2009-0081089).

## REFERENCES

- [1] J. Chen and X. Huo. Theoretical results on sparse representations of multiple measurement vectors. *IEEE Trans. on Signal Processing*, 54(12):4634–4643, 2006.
- [2] M.E. Davies and Y.C. Eldar. Rank awareness for joint sparse recovery. *arXiv:1004.4529*, 2010.
- [3] P. Feng. *Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals*. Dissertation, University of Illinois, Urbana-Champaign, 1997.
- [4] A.K. Fletcher, S. Rangan, and V.K. Goyal. Necessary and sufficient conditions for sparsity pattern recovery. *IEEE Trans. on Information Theory*, 55(12):5758–5772, 2009.
- [5] J.M. Kim, O.K. Lee, and J.C. Ye. Compressive MUSIC: a missing link between compressive sensing and array signal processing. *arXiv:1004.4398*, 2010.
- [6] Jong Min Kim, Ok Kyun Lee, and Jong Chul Ye. Compressive MUSIC with optimized partial support for joint sparse recovery. *arXiv:1102.3288*, 2011.
- [7] Jongmin Kim, Woohyuk Chang, Bangchul Jung, Dror Baron, and Jong Chul Ye. Belief propagation for joint sparse recovery. *arXiv:1102.3289*, 2011.
- [8] K. Lee and Y. Bresler. Subspace-augmented MUSIC for joint sparse recovery. *arXiv:1004.4371*, 2010.
- [9] D. Malioutov, M. Cetin, and A.S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. on Signal Processing*, 53(8):3010–3022, 2005.
- [10] V.A. Marčenko and L.A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics*, 1(4):457–483, 1967.
- [11] Galen Reeves. Sparse signal sampling using noisy linear projections. *Technical Report No. UCB/EECS-2008-3*, 2008.
- [12] Galen Reeves and Michael Gastpar. Sampling bounds for sparse support recovery in the presence of noise. In *IEEE International Symposium on Information Theory*, pages 2187–2191, Toronto, Canada, July 6–11 2008.
- [13] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. on Antennas and Propagation*, 34(3):276–280, 1986.

APPENDIX A: NUMBER OF MEASUREMENTS FOR  
COMPRESSIVE MUSIC WITH SUBSPACE S-OMP

In this section, we assume the large system limit so that we will assume that  $\rho$ ,  $\epsilon$ ,  $\alpha$  and  $\gamma$  exist. In this section, we will use the following theorem, which gives us the asymptotic distribution of singular values for Gaussian random matrices.

*Theorem 7.1:* [10] Suppose that each entry of  $A \in \mathbb{R}^{m \times k}$  is generated from i.i.d. Gaussian random variable  $\mathcal{N}(0, 1/m)$ . Then the probability density of squared singular value of  $A$  is given by

$$d\lambda_\gamma(x) := \frac{1}{2\pi\gamma^2} \frac{\sqrt{((1+\gamma)^2 - x)(x - (1-\gamma)^2)}}{x} dx. \quad (9)$$

By using above theorem, we prove the Theorem 3.2.

*Proof of Theorem 3.2:* First, for  $j \notin \text{supp}X$ , since  $\mathbf{a}_j$  is statistically independent of  $P_{R(A_{I_t})}^\perp B$ . For  $t \leq k - r$ , the dimension of  $P_{R(A_{I_t})}^\perp B$  is  $r$  so that  $m\|\mathbf{a}_j P_{R(P_{R(A_{I_t})}^\perp B)}\|^2$  is of chi-squared distribution of degree of freedom  $r$ .

On the other hand, for  $j \in \text{supp}X$ , we have

$$\begin{aligned} \max_{j \in \text{supp}X} \|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|^2 &\geq \frac{1}{k} \|A_S^* P_{R(P_{R(A_{I_t})}^\perp B)} A_S\|_F^2 \\ &\geq \frac{\sum_{j=1}^r \sigma_j^2(A_S)}{k} \end{aligned}$$

since  $R(P_{R(A_{I_t})}^\perp B) \subset R(A_S)$ , where  $A_S$  have singular values  $0 < \sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_k$ . Then by (9), we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^r \sigma_j^2(A_S)}{k} = \alpha \frac{\int_{(1-\gamma)^2}^{(1-\gamma+2\gamma t_\gamma(\alpha))^2} x d\lambda_\gamma(x)}{\alpha} \quad (10)$$

where  $0 \leq t_\gamma(\alpha) \leq 1$  is the value satisfying

$$\int_{1-\gamma}^{1-\gamma+2\gamma t_\gamma(\alpha)} ds_\gamma(x) = \int_{(1-\gamma)^2}^{(1-\gamma+2\gamma t_\gamma(\alpha))^2} d\lambda_\gamma(x) = \alpha.$$

If we let

$$d\lambda_{0,\gamma}(x) = \frac{1}{\pi} \frac{\sqrt{4-s}\sqrt{s}}{\gamma s + (1-\gamma)^2} dx, \quad (11)$$

then we have for any  $0 \leq t \leq 4$ ,

$$\int_0^t d\lambda_1(x) \leq \int_0^t d\lambda_{0,\gamma}(x) \quad (12)$$

then by substitution with  $s = (x - (1-\gamma)^2)/\gamma$ , we have

$$\begin{aligned} &\int_{(1-\gamma)^2}^{(1-\gamma+2\gamma t_\gamma(\alpha))^2} x d\lambda_\gamma(x) \\ &= \int_0^{\frac{(1-\gamma+2\gamma t_\gamma(\alpha))^2 - (1-\gamma)^2}{\gamma}} \frac{((1-\gamma)^2 + \gamma s)}{r} [(1-\gamma)^2 + \gamma s] d\lambda_{0,\gamma}(s) \\ &\geq \int_0^{4t_1(\alpha)^2} [(1-\gamma)^2 + \gamma s] d\lambda_1(s) \\ &= (1-\gamma)^2 \alpha + \gamma \int_0^{4t_1(\alpha)^2} s d\lambda_1(s), \end{aligned} \quad (13)$$

where the inequality comes from Lemma 7.2. Substituting (13) into (10), we have

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^r \sigma_j^2(A_S)}{k} \\ &\geq \alpha \left[ (1-\gamma)^2 + \gamma \frac{\int_0^{4t_1(\alpha)^2} s d\lambda_1(s)}{\alpha} \right] \\ &= \frac{r}{m} (1/\gamma - 1)^2 + \alpha \gamma F(\alpha) \end{aligned}$$

where

$$F(\alpha) := (1/\alpha) \int_0^{4t_1(\alpha)^2} s d\lambda_1(s) = \frac{\int_0^{4t_1(\alpha)^2} s d\lambda_1(s)}{\int_0^{4t_1(\alpha)^2} d\lambda_1(s)}$$

is an increasing function with respect to  $\alpha$  such that  $\lim_{\alpha \rightarrow 0} F(\alpha) = 0$  and  $F(1) = 1$ .

Then we consider two limiting cases according to the number of measurement vectors.

**(Case 1)** For  $t \leq k - r$ ,  $\{m\|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|^2 : j \notin \text{supp}X\}$  are independent chi-squared random variables of degree of freedom  $r$  so that by Lemma 7.4, we have

$$\lim_{n \rightarrow \infty} \max_{j \notin \text{supp}X} \frac{m\|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|^2}{2 \log(n-k)} = 1. \quad (14)$$

Here we assume that

$$m > k \frac{2(1+\delta) \log(n-k)}{r}. \quad (15)$$

Then by Marcenko-Pastur theorem [10],

$$\begin{aligned} \lim_{n \rightarrow \infty} \sigma_{\min}(A_S) &= \lim_{n \rightarrow \infty} (1 - \sqrt{k/m})^2 \\ &\geq \lim_{n \rightarrow \infty} \left(1 - \sqrt{r/(2 \log(n-k))}\right)^2 = 1 \end{aligned}$$

so that

$$\begin{aligned} &\liminf_{n \rightarrow \infty} \max_{j \in \text{supp}X} \frac{m\|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|^2}{2 \log(n-k)} \\ &\geq \liminf_{n \rightarrow \infty} \frac{m}{2 \log(n-k)} \frac{\sum_{j=1}^r \sigma_{k-j+1}^2(A_S)}{k} \\ &\geq \liminf_{n \rightarrow \infty} \frac{r}{2 \log(n-k)} \left(\frac{1}{\gamma}\right)^2 \geq 1 + \delta. \end{aligned} \quad (16)$$

Hence, when  $r$  is a fixed number, if we have (15), then we can identify  $k - r$  correct indices of  $\text{supp}X$  with subspace S-OMP, in the large system limit.

**(Case 2)** Similarly as in the previous case, for  $t < k - r$ ,  $\{m\|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|^2 : j \notin \text{supp}X\}$  are independent chi-squared distribution. Since  $\lim_{n \rightarrow \infty} (\log n)/r = 0$ , by Lemma 3 in [4], we have

$$\lim_{n \rightarrow \infty} \max_{j \notin \text{supp}X} \frac{m\|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|^2}{r} = 1. \quad (17)$$

On the other hand, for  $j \in \text{supp}X$ , we have

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \max_{j \in \text{supp}X} \frac{m \|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp)^\perp} Y\|^2}{r} \\ & \geq \left(\frac{1}{\gamma} - 1\right)^2 + F(\alpha). \end{aligned} \quad (18)$$

We let

$$m > k(1 + \delta)^2 [2 - F(\alpha)]^2 \quad (19)$$

for some  $\delta > 0$ . Note that (19) is equivalent to

$$\frac{1}{\gamma} > (1 + \delta)[2 - F(\alpha)]$$

Again we let

$$u := F(\alpha) \text{ and } v := \frac{4}{\alpha} \frac{\kappa(B) + 1}{\text{SNR}_{\min}(B) - 1}.$$

Then for a quadratic function  $Q(x) = (x - 1)^2 + ux$ , if  $x > (1 + \delta)(2 - u)$ , then we have

$$\begin{aligned} Q(x) &= x^2 - (2 - u)x + 1 = x[x - (2 - u)] + 1 \\ &> \delta(1 + \delta)(2 - u)^2 + 1 \geq 1 + \delta(1 + \delta) \end{aligned} \quad (20)$$

since  $0 \leq u \leq 1$ . Combining (18) and (20), we have for  $0 \leq t < k - r$  and  $j \in \text{supp}X$ , we have

$$\liminf_{n \rightarrow \infty} \max_{j \in \text{supp}X} \frac{m \|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp)^\perp} Y\|^2}{r} \geq 1 + \delta(1 + \delta)$$

for some  $\delta > 0$ . Hence, in the case of  $\lim_{n \rightarrow \infty} r/k = \alpha > 0$ , we can identify the correct indices of  $\text{supp}X$  if we have (19).

**Lemma 7.2:** For  $0 \leq \gamma \leq 1$  and  $0 \leq \alpha \leq 1$ , we let  $0 \leq t_\gamma(\alpha) \leq 1$  which satisfies

$$\int_{(1-\gamma)^2}^{(1-\gamma+2\gamma t_\gamma(\alpha))^2} ds_\gamma(x) = \alpha$$

where  $d\lambda_\gamma(x)$  is the probability measure which is given by

$$d\lambda_\gamma(x) := \frac{1}{\pi\gamma^2} \frac{\sqrt{((1+\gamma)^2 - x)(x - (1-\gamma)^2)}}{x}.$$

Furthermore, we let  $d\lambda_{0,\gamma}(x)$  is the probability measure which is given by

$$d\lambda_{0,\gamma}(x) = \frac{1}{\pi} \frac{\sqrt{4-x}\sqrt{x}}{\gamma x + (1-\gamma)^2} dx.$$

Then we have

$$\begin{aligned} & \int_0^{\frac{(1-\gamma+2\gamma t_\gamma(\alpha))^2 - (1-\gamma)^2}{\gamma}} [(1-\gamma)^2 + \gamma x] d\lambda_{0,\gamma}(x) \\ & \geq \int_0^{4t_1(\alpha)^2} [(1-\gamma)^2 + \gamma x] d\lambda_1(x). \end{aligned}$$

For the proof of Lemma 7.2, we need the following lemma.

**Lemma 7.3:** Let  $-\infty < a < b < \infty$ . Suppose that  $f_1(x)$  and  $f(x)$  are continuous probability density functions on

$[a, b]$  such that for any  $t \in [a, b]$ ,

$$\int_a^t f(x) dx \geq \int_a^t f_1(x) dx,$$

and satisfy that

$$f_1(x) > 0 \text{ and } f(x) > 0 \text{ on } (a, b).$$

Then for any nonnegative increasing function  $g(x)$  on  $[a, b]$  and for any  $(q_1, q) \in [a, b] \times [a, b]$  such that

$$\int_a^{q_1} f_1(x) dx = \int_a^q f(x) dx, \quad (21)$$

we have

$$\int_a^{q_1} g(x) f_1(x) dx \geq \int_a^q g(x) f(x) dx. \quad (22)$$

*Proof:* First, we define

$$F_1(x) = \int_a^x f_1(t) dt \text{ and } F(x) = \int_a^x f(t) dt.$$

Then both  $F_1(x)$  and  $F(x)$  are strictly increasing functions so that their inverse functions exist and satisfy  $F_1^{-1}(x) \geq F^{-1}(x)$  for any  $x \in [0, 1]$ . For any  $(q_1, q) \in [a, b] \times [a, b]$  which satisfies (21), there is some  $c \in [0, 1]$  such that  $F_1(q_1) = F(q) = c$ . Applying the change of variable, we have

$$\begin{aligned} & \int_a^{q_1} g(x) f_1(x) dx - \int_a^q g(x) f(x) dx \\ &= \int_0^c [g(F_1^{-1}(s)) - g(F^{-1}(s))] ds \geq 0 \end{aligned}$$

since  $F_1^{-1}(x) \geq F^{-1}(x)$  for any  $x \in [0, 1]$  and  $g(x)$  is increasing on  $[a, b]$ . ■

*Proof of Lemma 7.2* Noting that we have

$$\alpha = \int_0^{\frac{(1-\gamma+2\gamma t_\gamma(\alpha))^2 - (1-\gamma)^2}{\gamma}} d\lambda_{0,\gamma}(s) = \int_0^{4t_1(\alpha)^2} d\lambda_1(s),$$

by Lemma 7.3, we only need to show that

$$\int_0^t d\lambda_{0,\gamma}(s) \geq \int_0^t d\lambda_1(s)$$

for any  $t \in [0, 4]$ . Let  $f_1(x)$  and  $f(x)$  be given by

$$\begin{aligned} f_1(x) &= \frac{1}{\pi} \frac{\sqrt{4-x}\sqrt{x}}{\gamma x + (1-\gamma)^2}, \\ f(x) &= \frac{1}{\pi} \frac{\sqrt{4-x}\sqrt{x}}{x}. \end{aligned}$$

Then we can see that

$$\begin{aligned} f(x) &\geq f_1(x) && \text{for } x \in (0, 1 - \gamma) \\ \text{and } f_1(x) &\geq f(x) && \text{for } x \in [1 - \gamma, 1]. \end{aligned}$$

Since  $f_1(x)$  and  $f(x)$  are probability density functions with support  $[0, 4]$  so that we can easily see that for any  $t \in [0, 4]$ ,

$$\int_0^t f(x) dx \geq \int_0^t f_1(x) dx$$

so that the claim holds.

*Lemma 7.4:* Suppose that  $r$  is a given number, and  $\{u_j^{(n)}\}_{j=1}^n$  is a set of i.i.d. chi-squared random variables with degree of freedom  $r$ . Then

$$\lim_{n \rightarrow \infty} \max_{j=1, \dots, n} \frac{u_j^{(n)}}{2 \log n} = 1$$

in probability.

*Proof:* Assume that  $Z_r$  is a chi-squared random variable of degree of  $r$ , then we have

$$P\{Z_r > x\} = \frac{\Gamma(r/2, x/2)}{\Gamma(r/2)}, \quad (23)$$

where  $\Gamma(k, z)$  denotes the upper incomplete Gamma function. Then we use the following asymptotic behavior :

$$P\{Z_r > x\} \sim \frac{1}{\Gamma(r/2)} x^{r/2-1} e^{-x/2} \text{ as } x \rightarrow \infty.$$

For  $n \rightarrow \infty$ , we consider the probability  $P(\max_{1 \leq j \leq n} u_j^{(n)} > 2(1 + \epsilon) \log n)$ . By using union bound, we see that

$$\begin{aligned} & P(\max_{1 \leq j \leq n} u_j^{(n)} > 2(1 + \epsilon) \log n) \\ & \leq n \frac{1}{\Gamma(r/2)} (2(1 + \epsilon) \log n)^{r/2-1} e^{-(1+\epsilon) \log n} \\ & \leq \frac{1}{\Gamma(r/2)} (2(1 + \epsilon) \log n)^{r/2-1} n^{-\epsilon} \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . Now, considering the probability  $P(\max_{1 \leq j \leq n} u_j^{(n)} < 2(1 - \epsilon) \log n)$ , we see that

$$\begin{aligned} & P(\max_{1 \leq j \leq n} u_j^{(n)} < 2(1 - \epsilon) \log n) \\ & \leq \left(1 - \frac{1}{\Gamma(r/2)} (2(1 - \epsilon) \log n)^{r/2-1} e^{-(1-\epsilon) \log n}\right)^n \\ & \leq \left(1 - \frac{1}{\Gamma(r/2)} (2(1 - \epsilon) \log n)^{r/2-1} \frac{1}{n^{1-\epsilon}}\right)^n \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$  so that the claim is proved.  $\blacksquare$

## APPENDIX B: PROOF OF THEOREM 5.1

The proof of Theorem 5.1 basically follows the line from [11], which provides us the information theoretic analysis for partial support recovery with maximum likelihood (ML) estimator. Let  $P_e(\alpha)$  be the error probability conditioned on the true support set  $S$  with fractional distortion  $\alpha$ . Since the sampling procedure is independent from  $S$  so that for any distribution over  $S$ , we have  $P_e(\alpha) = P_e(\alpha|S)$ . Consider the sets

$$\begin{aligned} G &= \{U : |U| = k, |U \cap S| > (1 - \alpha)k\}, \\ B &= \{U : |U| = k, |U \cap S| \leq (1 - \alpha)k\}. \end{aligned}$$

Let  $\text{err}(U) = (1/\sigma_w^2) \|P_{R(A_U)}^\perp Y\|_F^2$ . For any  $t > 0$ , we define two events

$$A_B = \{\min_{U \in B} \text{err}(U) < t\}, \quad A_G = \{U : \min_{V \in G} \text{err}(V) > t\}.$$

Then  $P_e(\alpha|S) \leq P(A_B) + P(A_G)$ .

First, if we noting that  $\min_{V \in G} \text{err}(V) \geq \text{err}(K)$ , we have  $P(A_G) \leq P(\text{err}(K) > t)$ . Since  $N$  has zero mean i.i.d. Gaussian columns, and  $P_{R(A_S)}^\perp$  is an orthogonal projection matrix with rank  $m - k$ , the random variable  $\text{err}(K) = (1/\sigma_w^2) \|P_{R(A_S)}^\perp N\|_F^2$  has a chi-squared distribution with degree of freedom  $r(m - k)$  since  $r$  columns of  $N$  are independent.

Second, we consider  $P(A_B)$ . We partition  $B$  by  $B = \cup_{a=a_*}^{a^*} B_a$  where

$$\tilde{B}(a) = \{U : |U| = k, |U \cap S| = k - a\},$$

$a_* = \lfloor \alpha k \rfloor$  and  $a^* = \lceil (1 - \epsilon)k \rceil$ . Then

$$P(A_B) \leq \sum_{a=a_*}^{a^*} P(A_{\tilde{B}(a)})$$

where

$$A_{\tilde{B}(a)} = \left\{ \min_{U \in \tilde{B}(a)} \text{err}(U) < t \right\}.$$

Then we need to quantify the distribution of  $\text{err}(U)$  for  $U \in \tilde{B}(a)$ . First, if we condition on the set  $S \setminus U$ , the magnitude of the missed components of  $X$  is given by  $\text{SNR}(X_{S \setminus U})$ . Furthermore, for any  $U$ ,  $\Lambda(U) := (1/\sigma_w^2) \|P_{R(A_U)}^\perp N\|_F^2$  is a chi-squared random variable with  $r(m - k)$  degree of freedom by the independency of each column of  $N$ . Conditioned on  $\text{SNR}(X^{S \setminus U}) = \theta$ , the random vector  $(\sigma_w^2 \theta)^{-1/2} A_{S \setminus U} \mathbf{x}_j^{S \setminus U}$  has iid zero mean Gaussian random elements with variance 1, where  $\mathbf{x}_j$  is the  $j$ -th column of  $X$ . If we also add an another condition  $\Lambda(U) = \lambda$ , then we see that

$$\frac{1}{\theta} \text{err}(U) = \frac{1}{\sigma_w^2 \theta} \|P_{R(A_S)}^\perp (A_{S \setminus U} X^{S \setminus U} + N)\|_F^2$$

is a non-central chi-squared random variable with non-centrality parameter  $\lambda/\theta$  and degree of freedom  $r(m - k)$ . This implies that

$$\begin{aligned} & P\{\text{err}(U) < t | \text{SNR}(X_{S \setminus U}) = \theta, \Lambda(U) = \lambda\} \\ & = P\{\chi_{NC}^2(r(m - k), \lambda/\theta) < t/\theta\}. \end{aligned}$$

By the Lemma A.3 in [11], since  $\Lambda(U) \geq 0$ , we have

$$\begin{aligned} & P\{\text{err}(U) < t | \text{SNR}(X_{S \setminus U}) = \theta\} \\ & \leq P\{\chi^2(r(m - k)) < t/\theta\} \end{aligned}$$

using  $\chi_{NC}^2(r(m - k), 0) = \chi^2(r(m - k))$ . Hence we have

$$\begin{aligned} & P\{\text{err}(U) < t | \text{SNR}(X_{S \setminus U} \geq \theta)\} \\ & \leq P\{\chi^2(r(m - k)) < t/\theta\}. \end{aligned}$$

Then

$$P(A_{\tilde{B}(a)}) \leq P(\text{SNR} < \theta) + \sum_{U \in \tilde{B}(a)} P\{\chi^2(r(m - k)) < t/\theta\}.$$

By the definition of SNR and  $g(a/k, X)$ ,

$$\frac{\text{SNR}(X_{S \setminus U})}{\alpha g(a/k, X)} \geq \text{SNR}(X).$$



By the definition of  $g(a/k, \mathcal{X})$  and  $\text{SNR}(\mathcal{X})$ , there is a  $c_0 > 0$  such that

$$P\left\{\min_{U \in \tilde{B}(a)} \text{SNR}(X_{S \setminus U}) < 1/\zeta(a)\right\} < e^{-nc_0}$$

where  $\zeta(a) = [\text{SNR}(\mathcal{X})(a/k)g(a/k, \mathcal{X})]^{-1}$ . Hence

$$\begin{aligned} P(A_{\tilde{B}(a)}) &< e^{-nc_0} + \sum_{U \in \tilde{B}(a)} P\{\chi^2(r(m-k)) < \zeta(a)t\} \\ &= e^{-nc_0} + \binom{k}{a} \binom{n-k}{a} P\{\chi^2(r(m-k)) < \zeta(a)t\}. \end{aligned}$$

Reminding  $P_e(\alpha) \leq P(A_G) + P(A_B)$ , we first bound  $P(A_G)$ . For arbitrary  $\nu > 0$ , we choose  $t_\nu = (1 + \nu)r(m-k)$ . Then by Lemma 7.5, we have

$$P\{\chi^2(r(m-k)) > t_\nu\} \leq \exp(-nE_1),$$

where  $E_1 := (\rho - \epsilon)\nu^2/4$ . With  $\nu$  arbitrary close to 0, we consider the probability  $P(A_B)$ . To use Lemma 7.5, we need the condition

$$\text{SNR}(\mathcal{X}) > \max_{u \in [\alpha, 1-\epsilon]} \frac{1}{ug(u, \mathcal{X})} = \frac{1}{\alpha g(\alpha, \mathcal{X})}.$$

If this condition is satisfied, then

$$P\{\chi^2(r(m-k)) < \tau(a)t\} \leq e^{-nE_2(a)}$$

where

$$E_2(a) = \frac{\rho - \epsilon}{2} [-\log(\zeta(a)) + \zeta(a) - 1].$$

Finally noting that  $\log \binom{k}{a} \binom{n-k}{a} \rightarrow nh(\epsilon, a/k)$  as  $n \rightarrow \infty$ , we have

$$\begin{aligned} P_e(\alpha) &\leq e^{-nE_1} + \sum_{a=a_*}^{a^*} \left[ e^{-n(E_2(a) - h(\epsilon, a/k))} + e^{-nc_0} \right] \\ &< e^{-nE_1} + \max_{\alpha k \leq a \leq (1-\epsilon)k} e^{-n(E_2(a) - h(\epsilon, a/k)) + \log k} \\ &\quad + e^{-nc_0 + \log k}. \end{aligned}$$

For  $n \rightarrow \infty$ , the ML estimator is asymptotically reliable if we have SNR condition (7) and  $E_2(a) > h(\epsilon, a/k)$  for  $\alpha k \leq a \leq (1-\epsilon)k$  which holds under the condition (8).

*Lemma 7.5:* [11] For positive integer  $r$  and random variable  $Z$  which has the distribution  $\chi^2(r)$  and for any  $\epsilon > 0$  we have

$$\begin{aligned} P\{Z > (1 + \epsilon)r\} &\leq e^{-\frac{\epsilon}{4}\epsilon^2}, \\ P\{Z < (1 - \epsilon)r\} &\leq \exp\left(-\frac{r}{2} [-\log(1 - \epsilon) - \epsilon]\right). \end{aligned}$$

#### APPENDIX C: PROOF OF THEOREM 5.2

Again, the proof of Theorem 5.2 is a generalization of the result for the necessary condition in [11] to the MMV cases.

We define  $Z = X_S$ . Then the pair  $(S, Z)$  is equivalent to  $X$ . By the data processing inequality and the chain rule for mutual information, we have

$$I(B; Y) \geq I(Z, S; Y) = I(S; Y) + I(Z; Y|S) \quad (24)$$

where  $B = AX$  is the noiseless measurement. Since the noise

$N$  is i.i.d. Gaussian with covariance matrix  $\sigma_w^2 I$  and

$$E((AX + N)^*(AX + N)) = X^*X + \sigma_w^2 I$$

for a given  $X$ , we can obtain an upper bound of  $I(B; Y)$  as

$$\begin{aligned} I(B; Y) &\leq \frac{1}{2} \log [\det (I + \frac{1}{\sigma_w^2} X^*X)] \\ &= \sum_{l=1}^r \frac{1}{2} \log (1 + \frac{1}{\sigma_w^2} \lambda_l(X^*X)) \\ &\leq \sum_{l=1}^r \frac{1}{2} \log (1 + \frac{1}{\sigma_w^2} \kappa_l(\mathcal{X})) \end{aligned} \quad (25)$$

asymptotically since  $re^{-cn} \rightarrow 0$  for  $n \rightarrow \infty$ , where  $\lambda_l(X^*X)$  is the  $l$ -th largest eigenvalue of  $X^*X$  for  $1 \leq l \leq r$ .

Then, we consider the information  $I(S, Y)$ . Given that  $S$  is uniformly chosen over  $\binom{n}{k}$  possibilities, the asymptotic number of bits we need to decode  $S$  to with distortion rate  $\alpha$  is given by  $nh(\epsilon) - nh(\epsilon, \alpha)$ , where we used  $\log \binom{n}{k} = nh(k/n) + \mathcal{O}(\log n)$ . By Fano's inequality,  $P_e(\alpha) = 0$  only if

$$I(S; Y) \geq nh(\epsilon) - nh(\epsilon, \alpha). \quad (26)$$

Applying (24), (25) and (26) we have

$$m \geq \frac{nh(\epsilon) - nh(\epsilon, \alpha) + I(Z; Y|S)}{\sum_{l=1}^r \frac{1}{2} \log [1 + \frac{1}{\sigma_w^2} \kappa_l(\mathcal{X})]}$$