Compressive MUSIC: A Missing Link between Compressive Sensing and Array Signal Processing

Jong Min Kim, Ok Kyun Lee, *Student Member, IEEE* , and Jong Chul Ye, *Member, IEEE*

***Abstract*— The multiple measurement vector (MMV) problem addresses the identification of unknown input vectors that share common sparse support. Even though MMV problems have been traditionally addressed within the context of sensor array signal processing, the recent trend is to apply compressive sensing (CS) due to its capability to estimate sparse support even with an insufficient number of snapshots, in which case classical array signal processing fails. However, CS guarantees the accurate recovery in a probabilistic manner, which often shows inferior performance in the regime where the traditional array signal processing approaches succeed. The apparent dichotomy between the *probabilistic* CS and *deterministic* sensor array signal processing has not been fully understood. The main contribution of the present article is a unified approach that unveils a missing link between CS and array signal processing. The new algorithm, which we call *compressive MUSIC*, identifies the parts of support using CS, after which the remaining supports are estimated using a novel generalized MUSIC criterion. Using a large system MMV model, we show that our compressive MUSIC requires a smaller number of sensor elements for accurate support recovery than the existing CS methods and that it can approach the optimal l_0-bound with finite number of snapshots.**

***Index Terms*—Compressive sensing, multiple measurement vector problem, joint sparsity, MUSIC, S-OMP, thresholding**

I. INTRODUCTION

Compressive sensing (CS) theory [1–3] addresses the accurate recovery of unknown sparse signals from underdetermined linear measurements and has become one of the main research topics in the signal processing area with lots of applications [4–9]. Most of the compressive sensing theories have been developed to address the single measurement vector (SMV) problem [1–3]. More specifically, let m and n be positive integers such that $m < n$. Then, the SMV compressive sensing problem is given by

$$(P0): \quad \text{minimize} \quad \|\mathbf{x}\|_0 \qquad \text{subject to} \quad \mathbf{b} = A\mathbf{x}, \tag{I.1}$$

where $\mathbf{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\|\mathbf{x}\|_0$ denotes the number of non-zero elements in the vector $\mathbf{x}$. Since (P0) requires a computationally expensive combinatorial optimization, greedy methods [10], reweighted norm algorithms [11, 12], convex relaxation using l_1 norm [2, 13], or Bayesian approaches [14, 15] have been widely investigated as alternatives. One of the important theoretical tools within this context is the so-called restricted isometry property (RIP), which enables us to guarantee the robust recovery of certain input signals [3]. More specifically, a sensing matrix $A \in \mathbb{R}^{m\times n}$ is said to have a k-restricted isometry property(RIP) if there is a constant $0 \leq \delta_k < 1$ such that

$$(1-\delta_k)\|\mathbf{x}\|^2 \leq \|A\mathbf{x}\|^2 \leq (1+\delta_k)\|\mathbf{x}\|^2$$

for all $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{x}\|_0 \leq k$. It has been demonstrated that $\delta_{2k} < \sqrt{2}-1$ is sufficient for l_1/l_0 equivalence [2]. For many classes of random matrices, the RIP condition is satisfied with extremely high probability if the number of measurements satisfies $m \geq ck\log(n/k)$ for some constant $c > 0$ [3].

Another important area of compressive sensing research is the so-called multiple measurement vector problem (MMV) [16–19]. The MMV problem addresses the recovery of a set of sparse signal vectors that share common non-zero support. More specifically, let m, n and r be positive integers such that $m < n$. In the MMV context, m and r denote the number of sensor elements and snapshots, respectively. For a given observation matrix $B \in \mathbb{R}^{m\times r}$, a sensing matrix $A \in \mathbb{R}^{m\times n}$ such that $B = AX_*$ for some $X_* \in \mathbb{R}^{n\times r}$, the multiple measurement vector (MMV) problem is formulated as:

$$\text{minimize} \quad \|X\|_0 \qquad \text{subject to} \quad B = AX, \tag{I.2}$$

where $X = [\mathbf{x}_1, \cdots, \mathbf{x}_r] \in \mathbb{R}^{n\times r}$ and $\|X\|_0 = |\text{supp}X|$, where $\text{supp}X = \{1 \leq i \leq n : \mathbf{x}^i \neq 0\}$ and $\mathbf{x}^i$ is the i-th row of X. The MMV problem also has many important applications [20–24]. Currently, greedy algorithms such as S-OMP (simultaneous orthogonal matching pursuit) [16, 25], convex relaxation methods using mixed norm [26, 27], M-FOCUSS [17], M-SBL (Multiple Sparse Bayesian Learning)

J. M. Kim and O.K Lee are with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Guseong-dong Yuseong-gu, Daejon 305-701, Korea (e-mail:franzkim@gmail.com, okkyun2@kaist.ac.kr)

J.C. Ye is with the Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Guseong-dong Yuseong-gu, Daejon 305-701, Korea (phone:82-42-350-4320; fax:82-42-350-4310; e-mail:jong.ye@kaist.ac.kr)

Parts of this work were presented at the SIAM Conference on Imaging Science, Chicago, USA, 2010, with the title "Multiple measurement vector problem with subspace-based algorithm". This research was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded in part by the Korean government (MEST) (No. 2009-0081089) and in part by the Industrial Strategic Technology Program of the Ministry of Knowledge Economy (KI001889).

[28], randomized algorithms such as REduce MMV and BOost (ReMBo)[18], and model-based compressive sensing using block-sparsity [29, 30] have also been applied to the MMV problem within the context of compressive sensing.

In MMV, thanks to the common sparse support, it is quite predictable that the recoverable sparsity level may increase with the increasing number of measurement vectors. More specifically, given a sensing matrix A, let $\mathrm{spark}(A)$ denote the smallest number of linearly dependent columns of A. Then, according to Chen and Huo [16], Feng and Bresler [31], if $X \in \mathbb{R}^{n\times r}$ satisfies $AX = B$ and

$$\|X\|_0 < \frac{\mathrm{spark}(A) + \mathrm{rank}(B) - 1}{2} \leq \mathrm{spark}(A) - 1, \quad \text{(I.3)}$$

then X is the unique solution of (I.2). In (I.3), the last inequality comes from the observation that $\mathrm{rank}(B) \leq \|X_*\|_0 := |\mathrm{supp}X_*|$. Recently, Davies and Eldar showed that (I.3) is indeed a necessary codition for X to be a unique solution for $AX = B$ [32]. Compared to the SMV case ($\mathrm{rank}(B) = 1$), (I.3) informs us that the recoverable sparsity level increases with the number of measurement vectors. However, the performance of the aforementioned MMV compressive sensing algorithms are not generally satisfactory, and significant performance gaps still exist from (I.3) even for a noiseless case when only a finite number of snapshots is available.

On the other hand, before the advance of compressive sensing, the MMV problem (I.2), which was often termed as direction-of-arrival (DOA) or the bearing estimation problem, had been addressed using sensor array signal processing techniques [21]. One of the most popular and successful DOA estimation algorithms is the so-called the MUSIC (MUltiple SIgnal Classification) algorithm [33]. The MUSIC estimator has been proven to be a large snapshot (for $r \gg 1$) realization of the maximum likelihood estimator for any $m > k$, if and only if the signals are uncorrelated [34]. As will be shown later when $\mathrm{rank}(B) = k$ and the row vectors X are in general position, the maximum sparsity level that is uniquely recoverable using the MUSIC approach is

$$\|X\|_0 < \mathrm{spark}(A) - 1 \ , \quad \text{(I.4)}$$

which implies that the MUSIC algorithm achieves the l_0 bound (I.3) of the MMV when $\mathrm{rank}(B) = k$. However, one of the main limitations of the MUSIC algorithm is its failure when $\mathrm{rank}(B) < k$. This problem is often called the "coherent source" problem within the sensor array signal processing context [21].

To the best of our knowledge, this apparent "missing link" between compressive sensing and sensor array signal processing for the MMV problem has not yet been discussed. The main contribution of the present article is, therefore, to provide a new class of algorithms that unveils the missing link. The new algorithm, termed *compressive MUSIC* (CS-MUSIC), can be regarded as a deterministic extension of compressive sensing to achieve the l_0 optimality, or as a generalization of the MUSIC algorithm using a probabilistic setup to address the difficult problem of the coherent sources estimation. This generalization is due to our novel discovery of a *generalized MUSIC criterion*, which tells us that an unknown support of size $\mathrm{rank}(B)$ can be estimated *deterministically* as long as a $k - \mathrm{rank}(B)$ support can be estimated with any compressive sensing algorithm such as S-OMP or thresholding. Therefore, as $\mathrm{rank}(B)$ approaches k, our compressive MUSIC approaches the classical MUSIC estimator; whereas, as $\mathrm{rank}(B)$ becomes 1, the algorithm approaches to a classical SMV compressive sensing algorithm. Furthermore, even if the sparsity level is not known *a priori*, compressive MUSIC can accurately estimate the sparsity level using the generalized MUSIC criterion. This emphasizes the practical usefulness of the new algorithm. Since the fraction of the support that should be estimated probabilistically is reduced from k to $k - \mathrm{rank}(B)$, one can conject that the required number of sensor elements for compressive MUSIC is significantly smaller than that for conventional compressive sensing. Using the large system MMV model, we derive explicit expressions for the minimum number of sensor elements, which confirms our conjecture. Furthermore, we derive an explicit expression of the minimum SNR to guarantee the success of compressive MUSIC. Numerical experiments confirm out theoretical findings.

The remainder of the paper is organized as follows. We provide the problem formulation and mathematical preliminaries in Section II, followed by a review of existing MMV algorithms in Section III. Section IV gives a detailed presentation of the generalized MUSIC criterion, and the required number of sensor elements in CS-MUSIC is calculated in Section V. Numerical solutions are given in Section VI, followed by the discussion and conclusion in Section VII and VIII, respectively.

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 noiseless version of the canonical MMV formulation is very useful for our analysis.

Definition 2.1 (Canonical form noiseless MMV): Let m, n and r be positive integers ($r \leq m < n$) that represent the number of sensor elements, the ambient space dimension, and the number of snapshots, respectively. Suppose that we are given a sensing matrix $A \in \mathbb{R}^{m\times n}$ and an observation matrix $B \in \mathbb{R}^{m\times r}$ such that $B = AX_*$ for some $X_* \in \mathbb{R}^{n\times r}$ and $\|X_*\| = |\mathrm{supp}X| = k$. A canonical form noiseless multiple measurement vector (MMV) problem is given as the estimation problem of k-sparse vectors $X \in \mathbb{R}^{n\times r}$ using the following formulation:

$$\begin{aligned} \text{minimize} \quad & \|X\|_0 \\ \text{subject to} \quad & B = AX, \end{aligned} \tag{II.5}$$

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 observation matrix B is full rank, i.e. $\text{rank}(B) = r \leq k$.

Compared to (I.2), the canonical form MMV has the additional constraint that $\text{rank}(B) = r \leq \|X\|_0$. This is not problematic though since every MMV problem can be converted into a canonical form using the following dimension reduction.

- Suppose we are given the following linear sensor observations: $B = AX$ where $A \in \mathbb{R}^{m\times n}$ and $X \in \mathbb{R}^{n\times l}$ satisfies $\|X\|_0 = k$.
- Compute the SVD as $B = UD_rV^*$, where D_r is an $r\times r$ diagonal matrix, $V \in \mathbb{C}^{l\times r}$ consists of right singular vectors, and $r = \text{rank}(B)$, respectively.
- Reduce the dimension as $B_{SV} = BV$ and $X_{SV} = XV$.
- The resulting canonical form MMV becomes $B_{SV} = AX_{SV}$.

We can easily show that $\text{rank}(B_{SV}) = r \leq k$ and the sparsity $k := \|X\|_0 = \|X_{SV}\|_0$ with probability 1. Therefore, without loss of generality, the canonical form of the MMV in Definition 2.1 is assumed throughout the paper.

The following definitions are used throughout this paper.

Definition 2.2: [35] The rows (or columns) in $\mathbb{R}^n$ are in general position if any n collection of rows (or columns) are linearly independent.

If $A \in \mathbb{R}^{m\times n}$, where $m < n$, the columns of A are in general position if and only if $\text{spark}(A) = m + 1$. Also, it is equivalent to $K\text{-rank}(A) = m$ where $K-$rank denotes the Kruscal rank, where a Kruscal rank of A is the maximal number q such that every collection of q columns of A is linearly independent [18].

Definition 2.3 (Mutual coherence): For a sensing matrix $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$, the mutual coherence $\mu(A)$ is given by

$$\mu = \max_{1\leq j<k\leq n} \frac{|\mathbf{a}_j^*\mathbf{a}_k|}{\|\mathbf{a}_j\|\|\mathbf{a}_k\|},$$

where the superscript * denotes the Hermitian transpose.

Definition 2.4 (Restricted Isometry Property (RIP)): A sensing matrix $A \in \mathbb{R}^{m\times n}$ is said to have a k-restricted isometry property (RIP) if there exist left and right RIP constants $0 \leq \delta_k^L, \delta_k^R < 1$ such that

$$(1-\delta_k^L)\|\mathbf{x}\|^2 \leq \|A\mathbf{x}\|^2 \leq (1+\delta_k^R)\|\mathbf{x}\|^2$$

for all $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{x}\|_0 \leq k$. A single RIP constant $\delta_k = \max\{\delta_k^L, \delta_k^R\}$ is often referred to as the RIP constant.

Note that the condition for the left RIP constant $0 \leq \delta_{2k}^L < 1$ is sufficient for the uniqueness of any k-sparse vector $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}$ for any k-sparse vector $\mathbf{x}$, but the condition $\delta_{2k} < 1$ is often too restrictive.

III. Conventional MMV Algorithms

In this section, we review the conventional algorithms for the MMV problem and analyze their limitations. This survey is useful in order to understand the necessity of developing a new class of algorithms. Except for the MUSIC and cumulant MUSIC algorithm, all other algorithms have been developed in the context of compressive sensing. We will show that all the existing methods have their own disadvantages. In particular, the maximum sparsity levels that can be resolved by these algorithms are limited in achieving the maximum gain from joint sparse recovery.

A. Simultaneous Orthogonal Matching Pursuit (S-OMP)[16, 25]

The S-OMP algorithm is a greedy algorithm that performs the following procedure:

- at the first iteration, set $B_0 = B$ and $S_0 = \emptyset$,
- after J iterations, $S_j = \{l_j\}_{j=1}^J$ and $B_J = (I - P_{S_J})B$, where P_{S_J} is the orthogonal projection onto $\text{span}\{\mathbf{a}_{l_j}\}_{j=1}^J$,
- select l_{J+1} such that $\|\mathbf{a}_{l_{J+1}}^* B_J\|_2 = \max_{1\leq l\leq N} \|\mathbf{a}_l^* B_J\|_2$ and set $S_{J+1} = S_J \cup \{l_{J+1}\}$.

Worst case analysis of S-OMP [36] shows that a sufficient condition for S-OMP to succeed is

$$\max_{j\in \text{supp}X} \|A_S^\dagger \mathbf{a}_j\|_1 < 1, \tag{III.1}$$

where $S = \text{supp}X$. An explicit form of recoverable sparsity level is then given by

$$\|X\|_0 \quad < \quad \frac{1}{2}\left(\frac{1}{\mu}+1\right) . \tag{III.2}$$

Note that these conditions are exactly the same as Tropp's exact recovery conditions for the SMV problem [37], implying that the sufficient condition for the maximum sparsity level is not improved with an increasing number of snapshots even in the noiseless case. In order to resolve this issue, the authors in [36] and [38] performed an average case analysis for S-OMP, and showed that S-OMP can recover the input signals for the MMV problem with higher probability when the number of snapshots increases. However, the simulation results in [36] and [38] suggest that S-OMP performance is saturated after some number of snapshots, even with noiseless measurements, and S-OMP never achieves the l_0 bound with a finite number of snapshots.

B. 2-Thresholding [36]

In 2-thresholding, we select a set S with $|S| = k$ such that

$$\|\mathbf{a}_l^* B\|_2 \geq \|\mathbf{a}_j^* B\|_2, \quad \text{for all } l \in S, \; j \notin S.$$

If we estimate the $\text{supp}X$ by the above criterion, we can recover the nonzero component of X by the equation $X^S = A_S^\dagger Y$. In [36], the authors demonstrated that the performance of 2-thresholding is often not as good as that of S-OMP, which suggests that 2-thresholding never achieves the l_0-bound (I.3) with finite snapshots even if the measurements are noiseless.

C. ReMBO algorithm [18]

Reduce MMV and Boost (ReMBo) by Mishali and Eldar [18] addresses the MMV problem by reducing it to a series of SMV problems based on the following.

Theorem 3.1: [18] Suppose that X satisfies $\|X\|_0 = k$ and $AX = B$ with $k < \text{spark}(A)/2$. Let $\mathbf{v} \in \mathbb{R}^r$ be a random vector with an absolutely continuous distribution and define $\mathbf{b} = A\mathbf{v}$ and $\overline{\mathbf{x}} = X\mathbf{v}$. Then, for a random SMV system $A\mathbf{x} = \mathbf{b}$ and $\mathbf{b} = B\mathbf{v}$, we have

(a) For every $\mathbf{v}$, the vector $\overline{\mathbf{x}}$ is the unique k-sparse solution.
(b) $\text{Prob}(\text{supp}(\mathbf{x}) = \text{supp}(\overline{\mathbf{x}})) = 1$.

Employing the above theorem, Mishali and Eldar [18] proposed the ReMBo algorithm which performs the following procedure:

- set the maximum number of iterations as MaxIters, set $i = 1$ and Flag = F,
- while $i \leq$ MaxIters and Flag = F, generate a random SMV problem as in Theorem 3.1,
  - if the SMV problem has a k-sparse solution, then we let S be the support of the solution vector, and let Flag = T
  - otherwise, increase i by 1
- if Flag = T, find the nonzero components of X by the equation $X^S = A_S^\dagger B$.

In order to achieve the l_0 bound (I.4) by ReMBO without any combinatorial SMV solver, an uncountable number of random vectors $\mathbf{v}$ are required. With a finite number of choices of $\mathbf{v}$, the performance of ReMBo is therefore dependent on randomly chosen input and the solvability of a randomly generated SMV problem so that it is difficult to achieve the theoretical l^0-bound even with noiseless measurements.

D. Mixed norm approach [27]

The mixed norm approach is an extension of the convex relaxation method in SMV [10] to the MMV problem. Rather than solving the original MMV problem (II.5), the mixed norm approaches solve the following convex optimization problem:

$$\begin{aligned} \text{minimize} \quad & \|X\|_{p,q} \qquad 1 \leq p, q \leq 2 \\ \text{subject to} \quad & B = AX, \end{aligned} \tag{III.3}$$

where $\|X\|_{p,q} = (\sum_{i=1}^n \|\mathbf{x}^i\|_p^q)^{\frac{1}{q}}$. The optimization problem can be formulated as an SOCP (second order cone program) [26], homotopy continuation [39], and so on. Worst case bounds for the mixed norm approach were derived in [16], which shows no improvement with the increasing number of measurement. Instead, Eldar *et al* [38] considered the average case analysis when $p = 2$ and $q = 1$ and showed that if

$$\max_{j \notin S} \|A_S^\dagger \mathbf{a}_j\|_2 \leq \alpha < 1,$$

where $S = \text{supp}X$, then the probability success recovery of joint sparsity increases with the number of snapshots. However, it is not clear whether this convex relaxtion can achieve the l_0 bound.

E. Block sparsity approaches [29]

Block sparse signals have been extensively studied by Eldar et al using the uncertainty relation for the block-sparse signal and block coherence concept. Eldar et al. [29] showed that the block sparse signal can be efficiently recovered using a fewer number of measurements by exploiting the block sparsity pattern as described in the following theorem:

Theorem 3.2: [29] Let positive integers L, n, N and $D = [\mathbf{D}[1], \cdots, \mathbf{D}[n]] \in \mathbb{R}^{L \times N}$ be given, where $L < N$, $N = nr$ for some positive integer r and for each $1 \leq j \leq n$, $\mathbf{D}[j] \in \mathbb{R}^{L \times r}$. Let μ_B be the block-coherence which is defined by

$$\mu_B = \max_{1 \leq j < k \leq n} \frac{1}{r} \rho(\mathbf{D}[j]^* \mathbf{D}[k])$$

where ρ denotes the spectral radius, ν be the sub-coherence of the sensing matrix A which is defined by

$$\nu = \max_l \max_{i \neq j} |\mathbf{d}_i^* \mathbf{d}_j|, \quad \mathbf{d}_i, \mathbf{d}_j \in \mathbf{D}[l],$$

and r be the block size. Then, the block OMP and block mixed l_2/l_1 optimization program successfully recover the k-block sparse signal if

$$kr < \frac{1}{2} \left(\mu_B^{-1} + r - (r-1) \frac{\nu}{\mu_B} \right). \tag{III.4}$$

Note that we can transform $B = AX$ into an SMV system $\text{vec}(B^T) = (A \bigotimes I_r) \text{vec}(X^T)$, where $\text{vec}(X^T)$ is block-k sparse with length r and $\bigotimes$ denotes the Kronecker product of matrices. Therefore, one may think that we can use the block OMP or block l_2/l_1 optimization problem to solve the MMV problem. However, the following theorem shows that this is pessimistic.

Theorem 3.3: For the canonical MMV problem in Definition 2.1, the sufficient condition (III.4) for recovery using block-sparsity is equivalent to

$$k < \frac{1}{2} \left(\mu^{-1} + 1 \right), \tag{III.5}$$

where μ denotes the mutual coherence of the sensing matrix $A \in \mathbb{R}^{m \times n}$.

Proof: Since $A \bigotimes I_r = [a_{i,j} I_r]_{i,j=1}^{m,n}$, if we let $A \bigotimes I_r = [\mathbf{D}[1], \cdots, \mathbf{D}[n]]$, we have $\nu = 0$ due to the diagonality, and

$$\begin{aligned} \mu_B &= \max_{1 \leq j < k \leq n} \frac{1}{r} \rho \left(\mathbf{D}[j]^* \mathbf{D}[k] \right) \\ &= \max_{1 \leq j < k \leq n} \frac{1}{r} \rho \left(\sum_{i=1}^{m} a_{ij}^* a_{ik} I \right) = \frac{\mu}{r} \end{aligned}$$

by the definition of mutual coherence. Applying (III.4) with $\nu = 0$ and $\mu_B = \mu/r$, we obtain (III.5). ■

Note that (III.5) is the same as that of OMP for SMV. The main reason for the failure of the block sparse approach for the MMV problem is that the block sparsity model does not exploit the diversity of unknown matrix X. For example, the block sparse model cannot differentiate a rank-one input matrix X and full-rank matrix X.

F. M-SBL [28]

M-SBL (Sparse Bayesian Learning) by Wipf and Rao [40] is a Bayesian compressive sensing algorithm to address the l_0 minimization problem. M-SBL is based on the ARD (automatic relevance determination) and utilizes an empirical Bayesian prior thereby enforcing a joint sparsity. Specifically, the M-SBL performs the following procedure:

(a) initialize γ and $\Gamma := \mathrm{diag}(\gamma) \in \mathbb{R}^{n \times n}$.

(b) compute the posterior variance Σ and mean $\hat{X}$ as follows:

$$\begin{aligned} \Sigma &:= \Gamma - \Gamma A^* (A \Gamma A^* + \lambda I)^{-1} A \Gamma \\ \hat{X} &:= \Gamma A^* (A \Gamma A^* + \lambda I)^{-1} B, \end{aligned}$$

where $\lambda > 0$ denotes a regularization parameter.

(c) update γ by

$$\gamma_j^{(new)} = \frac{\|\mu_j\|^2}{r} \frac{1}{1 - \gamma_j^{-1} \Sigma_{jj}}, \quad 1 \leq j \leq n$$

(d) repeat (b) and (c) until γ converges to some fixed point γ^*.

Wipf and Rao [40] showed that increasing the number of snapshots in SBL reduces the number of local minimizers so that the possibility of recovering input signals increases from joint sparsity. Furthermore, in the noiseless setting, if we have k linearly independent measurements and the nonzero rows of X are *orthogonal*, there is a unique fixed point γ^* so that we can correctly recover the k-sparse input vectors. To the best of our knowledge, M-SBL is the only compressive sensing algorithm that achieves the same l_0-bound as MUSIC when $r = k$. However, the orthogonality condition for the input vector X that achieves the maximal sparsity level is more restricted than that of MUSIC. Furthermore, no explicit expression for the maximum sparsity level was provided for the range $\mathrm{rank}(B) < k$.

G. The MUSIC Algorithm [31, 33]

The MUSIC algorithm was originally developed to estimate the continuous parameters such as bearing angle or DOA. However, the MUSIC criterion can be still modified to identify the support set from the finite index set as follows.

Theorem 3.4: [31, 33](MUSIC Criterion) Assume that we have r linearly independent measurements $B \in \mathbb{R}^{m \times r}$ such that $B = AX_*$ for $X_* \in \mathbb{R}^{n \times r}$ and $r = \|X_*\|_0 =: k < m$. Also, we assume 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, for any $j \in \{1, \cdots, n\}$, $j \in \mathrm{supp} X_*$ if and only if

$$Q^* \mathbf{a}_j = 0, \tag{III.6}$$

or equivalently

$$\mathbf{a}_j^* P_{R(Q)} \mathbf{a}_j = 0 \tag{III.7}$$

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”. Here, for matrix A, $R(A)$ denotes the range space of A.

Proof: By the assumption, the matrix of multiple measurements B can be factored as a product $B = A_S X_*^S$ where $A_S \in \mathbb{R}^{m \times k}$ and $X_*^S \in \mathbb{R}^{k \times k}$, where $S = \mathrm{supp} X_*$, A_S is the matrix which consists of columns whose indices are in S and X_*^S is the matrix that consists of rows whose indices are in S. Since A_S has full column rank and X_*^S has full row rank, $R(B) = R(A_S)$. Then we can obtain a singular value decomposition as

$$B = [U \ Q] \mathrm{diag}[\sigma_1, \cdots, \sigma_k, 0, \cdots, 0] V^*,$$

where $R(U) = R(A_S) = R(Q)^\perp$. Then, $Q^* \mathbf{a}_j = 0$ if and only if $\mathbf{a}_j \in R(Q)^\perp = R(A_S)$ so that $\mathbf{a}_j$ can be expressed as a linear combination of $\{\mathbf{a}_k\}_{k \in S}$. Since the columns of A are in general position, $Q^* \mathbf{a}_j = 0$ if and only if $j \in \mathrm{supp} X_*$. ■

Note that the MUSIC criterion (III.6) 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

$$\|X\|_0 < m = \mathrm{spark}(A) - 1, \tag{III.8}$$

where the last equality comes from the definition of the spark. Therefore, the l_0 bound (I.3) can be achieved by MUSIC when $r = k$. However, for any $r < k$, the MUSIC condition (III.6) 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 number of sensor elements, m.

H. Cumulant MUSIC

The fourth-order cumulant or higher order MUSIC was proposed by Porat and Friedlander [41] and Cardoso [42] to improve the number of resolvable resolvable sources over the conventional second-order MUSIC. Specifically, the cumulant MUSIC derives a MUSIC-type subspace criterion from the cumulant of the observation matrix. It has been shown that the cumulant MUSIC can resolve more sources than conventional MUSIC for specific array geometries [43]. However, a significant increase in the variance of the target estimate of a weak source in the presence of stronger sources has been reported, which was not observed for second order MUSIC [44]. This increase often prohibit the use of fourth-order methods, even for large SNR, when the dynamic range of the sources is important [44]. Furthermore, for general array geometries, the performance of the cumulant MUSIC is not clear. Therefore, we need to develop a new type of algorithm that can overcome these drawbacks.

I. Main Contributions of Compressive MUSIC

Note that the existing MMV compressive sensing approaches are based on a probabilistic guarantee, whereas array signal processing provides a deterministic guarantee. Rather than taking such extreme view points to address a MMV problem, the main contribution of CS-MUSIC is to show that we should take the best of both approaches. More specifically, we show that as long as $k - \mathrm{rank}(B)$ partial support can be estimated with any compressive sensing algorithms, the remaining unknown support of $\mathrm{rank}(B)$ can be estimated deterministically using a novel generalized MUSIC criterion. By allowing such hybridization, our CS-MUSIC can overcome the drawbacks of the all existing approaches and achieves the superior recovery performance that had not been achievable by any of the aforementioned MMV algorithms. Hence, the following sections discuss what conditions are required for the generalized MUSIC and partial support support recovery to succeed, and how CS-MUSIC outperforms existing methods.

IV. Generalized MUSIC criterion for compressive MUSIC

This section derives an important component of compressive MUSIC, which we call the generalized MUSIC criterion. This extends the MUSIC criterion (III.6) for $r \leq k$. Recall that when we obtain k linearly independent measurement vectors, we can determine the support of multiple signals with the condition that $Q^*\mathbf{a}_j = 0$ if and only if $j \in \mathrm{supp}X$. In general, if we have r linearly independent measurement vectors, where $r \leq k$, we have the following.

Theorem 4.1 (Generalized MUSIC criterion): Let m, n and r be positive integers such that $r \leq m < n$. Suppose that we are given a sensing matrix $A \in \mathbb{R}^{m\times n}$ and an observation matrix $B \in \mathbb{R}^{m\times r}$. Assume that the MMV problem is in canonical form, that is, $\mathrm{rank}(B) = r \leq k$. Then, the following holds:
(a) $\mathrm{spark}(Q^*A) \leq k - r + 1$.
(b) If the k nonzero rows of X are in general position (i.e., any collection of r nonzero rows are linearly independent) and A satisfies the RIP condition with $0 \leq \delta^L_{2k-r+1}(A) < 1$, then

$$\mathrm{spark}(Q^*A) = k - r + 1.$$

Proof: See Appendix A. ■

Note that, unlike the classical MUSIC criterion, a condition for the left RIP constant $0 \leq \delta^L_{2k-r+1}(A) < 1$ is required in Theorem 4.1 (b). This condition has the following very interesting implication.

Lemma 4.2: For the canonical form MMV, $A \in \mathbb{R}^{m\times n}$ satisfies RIP with $0 \leq \delta^L_{2k-r+1} < 1$ if and only if

$$k < \frac{\mathrm{spark}(A) + \mathrm{rank}(B) - 1}{2}. \quad \text{(IV.1)}$$

Proof: Since $A \in \mathbb{R}^{m\times n}$ has the left RIP condition $0 \leq \delta^L_{2k-r+1} < 1$, any collection of $2k - r + 1$ columns of A are linearly independent so that $\mathrm{spark}(A) > 2k - r + 1$. Hence,

$$k < \frac{\mathrm{spark}(A) + r - 1}{2} = \frac{\mathrm{spark}(A) + \mathrm{rank}(B) - 1}{2}$$

since $r = \mathrm{rank}(B)$. For the converse, assume the condition (IV.1). Then we have $2k - r + 1 < \mathrm{spark}(A)$ which implies $0 \leq \delta^L_{2k-r+1} < 1$. ■

Hence, if A satisfies RIP with $0 \leq \delta^L_{2k-r+1} < 1$ and if we have k-sparse coefficient matrix X that satisfies $AX = B$, then X is the unique solution of the MMV. In other words, under the above RIP assumption, for noiseless case we can achieve the l_0-uniqueness bound, which is the same as the theoretical limit (I.3). Note that when $k = r$, we have $\mathrm{spark}(Q^*A) = 1$, which is equivalent to there being some j's such that $Q^*\mathbf{a}_j = 0$, which is equivalent to the classical MUSIC criterion. By the above lemma, we can obtain a *generalized MUSIC criterion* for the case $r \leq k$ in the following theorem.

Theorem 4.3: 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 $AX = B$ and the conditions in Theorem 4.1 (b). If $I_{k-r} \subset \mathrm{supp}X$ with $|I_{k-r}| = k - r$ and $A_{I_{k-r}} \in \mathbb{R}^{m\times(k-r)}$, which consists of columns of A, whose indices are in I_{k-r}, then for any $j \in \{1, \cdots, n\} \setminus I_{k-r}$,

$$\mathrm{rank}(Q^*[A_{I_{k-r}}, \mathbf{a}_j]) = k - r \quad \text{(IV.2)}$$

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

Proof: See Appendix B. ■

When $r = k$, $A_{I_{k-r}} = \emptyset$ and (IV.2) is the same as the classic MUSIC criterion (III.6) since $\mathrm{rank}(Q^*\mathbf{a}_j) = 0 \iff Q^*\mathbf{a}_j = 0$. However, the generalized MUSIC criterion (IV.2) for $r < k$ is based on the rank of the matrix, which is prone to error under an incorrect estimate of noise subspace Q when the measurements are corrupted by additive noise. Hence, rather than using (IV.2), the following equivalent criterion is more practical.

Corollary 4.4: Assume that $A \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{n\times r}$, $B \in \mathbb{R}^{m\times r}$, $I_{k-r} \subset \mathrm{supp}X$, and $A_{I_{k-r}}$ are the same as in Theorem 4.3. Then,

$$\mathbf{a}_j^*\left[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}\right]\mathbf{a}_j = 0 \quad \text{(IV.3)}$$

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

Proof: See Appendix B. ■

Note that $P_{R(Q)} = QQ^*$ in MUSIC criterion (III.7) is now replaced by $P_{R(Q)} - P_{R[P_{R(Q)}A_{I_{k-r}}]}$ where $I_{k-r} \subset \text{supp}X$. The following theorem shows that $P_{R(Q)} - P_{R[P_{R(Q)}A_{I_{k-r}}]}$ has very important geometrical meaning.

Theorem 4.5: Assume that we are given a noiseless MMV problem which is in canonical form. Also, suppose that A and X satisfy the conditions as in Theorem 4.1 (b). Let $U \in \mathbb{R}^{m\times r}$ and $Q \in \mathbb{R}^{m\times(m-r)}$ consist of orthonormal columns such that $R(U) = R(B)$ and $R(Q)^\perp = R(B)$. Then the following properties hold :

(a) $UU^* + P_{R(QQ^*A_{I_{k-r}})}$ is equal to the orthogonal projection onto $R(B) + R(QQ^*A_{I_{k-r}})$.

(b) $QQ^* - P_{R(QQ^*A_{I_{k-r}})}$ is equal to the orthogonal projection onto $R(Q) \cap R(QQ^*A_{I_{k-r}})^\perp$.

(c) $QQ^* - P_{R(QQ^*A_{I_{k-r}})}$ is equal to the orthogonal complement of $R([U\ A_{I_{k-r}}])$ or $R([B\ A_{I_{k-r}}])$.

Proof: See Appendix C. ■

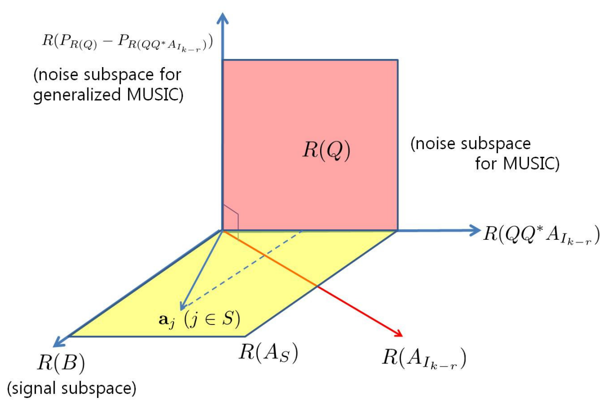


Fig. 1. Geometric view for the generalized MUSIC criterion : the dashed line corresponds to the conventional MUSIC criterion, where the squared norm of the projection of $\mathbf{a}_j (j \in \text{supp}X)$ onto the noise subspace $R(Q)$ may not be zero. $\mathbf{a}_j (j \in \text{supp}X)$ is orthogonal to the subspace $R(P_{R(Q)} - P_{R(QQ^*A_{I_{k-r}})})$ so that we can identify the indices of the support of X with the generalized MUSIC criterion.

Figure 1 illustrates the geometry of corresponding subspaces. Unlike the MUSIC, the orthogonality of the $\mathbf{a}_j$, $j \in \text{supp}X$ need to be checked with respect to $R(Q) \cap R(QQ^*A_{I_{k-r}})^\perp$. Based on the geometry, we can obtain following algorithms for support detection.

(Algorithm 1: Original form)

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

(Algorithm 2: Signal subspace form)

Alternatively, we can also use the signal subspace form to identify the support of X:

1) Find $k-r$ indices of $\text{supp}X$ by any MMV compressive sensing algorithms such as 2-thresholding or SOMP.
2) Let I_{k-r} be the set of indices which are taken in Step 1 and $S = I_{k-r}$.
3) For $j \in \{1, \cdots, n\} \setminus I_{k-r}$, calculate the quantities $\eta(j) = \mathbf{a}_j^*[P_{R(U)} + P_{R(P^\perp_{R(U)}A_{I_{k-r}})}]\mathbf{a}_j$ for all $j \notin I_{k-r}$.
4) Make a descending ordering of $\eta(j)$, $j \notin I_{k-r}$, 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 reconstruction 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 4.3 or Corollary 4.4, 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 have discussed about the recovery of the support of the multiple input vectors assuming that we know about the size of the support. One of the disadvantages of the existing MUSIC-type algorithms is that if the sparsity level is overestimated, spurious peaks are often observed. However, in CS-MUSIC when we do not know about the correct size of the support, we can still apply the following lemma to estimate the size of the support.

Lemma 4.6: 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 $AX_* = B$ and the conditions in theorem 4.1 (b), and k denotes the true sparsity level, i.e. $k = \|X_*\|_0$. Also, assume that $r < \hat{k} \le k + r$ and we are given $I_{\hat{k}-r} \subset \text{supp}X$ with $|I_{\hat{k}-r}| = \hat{k} - r$, where $I_{\hat{k}-r}$ is the partial support of size $\hat{k} - r$ estimated by any MMV compressive sensing algorithm. Also, we let $\eta(j) := \mathbf{a}_j^*[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}]\mathbf{a}_j$. Then, $\hat{k} = k = \|X_*\|_0$ if and only if

$$C(\hat{k}) := \min_{J\cap I_{\hat{k}-r}=\emptyset, |J|=r} \sum_{j\in J} \eta(j) = 0. \qquad \text{(IV.4)}$$

Proof: Necessity is trivial by Corollary 4.4 so we only need to show sufficiency of (IV.4) assuming the contrary. We divide the proof into two parts.

(i) $r < \hat{k} < k$: By the Lemma 4.1, for any $j \in \{1, \cdots, n\} \setminus I_{\hat{k}-r}$,

$$\text{rank}[Q^*_{I_{\hat{k}-r}}[A_{\hat{k}-r}, \mathbf{a}_j]] = \hat{k} - r + 1.$$

As in the proof of Corollary 4.4, this implies $\eta(j) > 0$ for any $j \in \{1, \cdots, n\} \setminus I_{\hat{k}-r}$, so that we have $C(\hat{k}) > 0$ for $\hat{k} < k$.

(ii) $k < \hat{k} \le k + r$: Here, we have already chosen at least $k-r+1$ indices of the support of X. By Corollary 4.4, (IV.3) holds only for, at most, $r-1$ elements of $\{1, \cdots, n\} \setminus I_{\hat{k}-r}$ since $I_{\hat{k}-r} \subset \text{supp}X$. Hence, $C(\hat{k}) > 0$ for $\hat{k} > k$. ■

The minimization in (IV.4) is over all index sets J of size r that include elements from $\{1, \cdots, n\}$ and no elements form $I_{\hat{k}-r}$. For fixed $\hat{k}$ and $I_{\hat{k}-r}$, this minimization can be performed by first computing the summands for all $j \in \{1, \cdots, n\} \setminus I_{\hat{k}-r}$ and then selecting the r of smallest magnitude. Lemma 4.6 also tells us that if we calculate $C(\hat{k})$ by increasing $\hat{k}$ from r, then the first $\hat{k}$ such that $C(\hat{k}) = 0$ corresponds to the unknown sparsity level. For noisy measurements, we can choose the first local minimizer of $C(\hat{k})$ by increasing $\hat{k}$.

V. Sufficient Conditions for Sparse Recovery using Compressive MUSIC

A. Large system MMV model

Note that the recovery performance of compressive MUSIC relies entirely on the correct identification of $k - r$ partial support in $\mathrm{supp}X$ via compressive sensing approaches and the remaining r indices using the generalized MUSIC criterion. In practice, the measurements are noisy, so the theory we derived for noiseless measurement should be modified. In this section, we derive sufficient conditions for the minimum number of sensor elements (the number of rows in each measurement vector) that guarantee the correct support recovery by compressive MUSIC. Note that for the success of compressive MUSIC, both CS step and the generalized MUSIC step should succeed. Hence, this section derives separate conditions for each step, which is required for the success of compressive MUSIC.

For SMV compressive sensing, Fletcher, Rangan and Goyal [45] derived an explicit expression for the minimum number of sensor elements for the 2-thresholding algorithm to find the correct support set. Also, Fletcher and Rangan [46] derived a sufficient condition for S-OMP to recover X. Even though their derivation is based on a large system model with a Gaussian sensing matrix, it has provided very useful insight into the SMV compressive sensing problem. Therefore, we employed a large system model to derive a sufficient condition for compressive MUSIC.

Definition 5.1: A large system noisy canonical MMV model, $\mathrm{LSMMV}(m, n, k, r; \epsilon)$, is defined as an estimation problem of k-sparse vectors $X \in \mathbb{R}^{n\times r}$ that shares a common sparsity pattern through multiple noisy snapshots $Y = AX + N$ using the following formulation:

$$\begin{aligned} \text{minimize} \quad & \|X\|_0 \\ \text{subject to} \quad & Y = AX + N, \end{aligned} \tag{V.1}$$

where $A \in \mathbb{R}^{m\times n}$ is a random matrix with i.i.d. $\mathcal{N}(0, 1/m)$ entries, $N = [\mathbf{n}_1, \cdots, \mathbf{n}_r] \in \mathbb{R}^{m\times r}$ is an additive noise matrix, $m = m(n) \to \infty$, $k = k(n) \to \infty$ as $n \to \infty$ and $\mathrm{rank}(AX) = r \leq k = \|X\|_0$ where $r = r(n)$ is a fixed number or is proportionally increasing with respect to n. Here, we assume that $\rho := \lim_{n\to\infty} m(n)/n > 0$ and $\alpha = \lim_{n\to\infty} r(n)/k(n) \geq 0$ exist.

Note that the conditions $k/m < 1 - \epsilon$, and $r/k < 1 - \epsilon$ are technical conditions that prevent m, k, and r from reaching equivalent values when $n \to \infty$.

B. Sufficient condition for generalized MUSIC

For the case of a noisy measurement, Y is corrupted and the corresponding noise subspace estimate Q is not correct. However, the following theorem shows that if the $I_{k-r} \subset \mathrm{supp}X$, then the generalized MUSIC estimate is consistent and achieves the correct estimation of the remaining r-indices for sufficiently large SNR.

Theorem 5.1: For a $\mathrm{LSMMV}(m, n, k, r; \epsilon)$, if we have $I_{k-r} \subset \mathrm{supp}X$, then we can find remaining r indices of $\mathrm{supp}X$ with the generalized MUSIC criterion if

$$m \geq \max\left\{ \frac{k(1+\delta)}{\left[1 - \frac{4(\kappa(B)+1)}{\mathsf{SNR}_{\min}(Y)-1}\right]}, (1+\delta)(2k - r + 1) \right\} \tag{V.2}$$

for some $\delta > 0$ provided that $\mathsf{SNR}_{\min}(Y) := \sigma_{\min}(B)/\|N\| > 1 + 4(\kappa(B) + 1)$, where $\kappa(B)$ denotes the condition number and $\sigma_{\min}(B)$ denotes the smallest singular value of B.

Proof: See Appendix D. ■

Note that for $\mathsf{SNR}_{\min}(Y) \to \infty$, the condition becomes $m \geq (1 + \delta)(2k - r + 1)$ for some $\delta > 0$. However, as $\mathsf{SNR}_{\min}(Y)$ decreases, the first term dominates and we need more sensor elements.

C. Sufficient condition for partial support recovery using 2-thresholding

Now, define the thresholding estimate as $I_t = \{p_i\}_{i=1}^{k-r}$ where

$$\rho(j) = \|\mathbf{a}_j^* Y\|_F^2.$$

Now, we derive sufficient conditions for the success of 2-thresholding in detecting $k-r$ support when r is a small fixed number or when r is proportionally increasing with respect to k.

Theorem 5.2: For a $\mathrm{LSMMV}(m, n, k, r; \epsilon)$, suppose $\mathsf{MSR}_{min}^{(k-r)}$ are deterministic sequences and

$$\mathsf{SNR}_{\min}(Y) > \frac{2\kappa(B) + \sqrt{4\kappa(B)^2 + \frac{2r\mathsf{MSR}_{\min}^{(k-r)}}{\sigma_{\min}^2(B)}}}{r\mathsf{MSR}_{\min}^{(k-r)}/(\sigma_{\min}^2(B))}, \tag{V.3}$$

$$m > 2(1+\delta)\frac{\left(\frac{\|X\|_F}{\sqrt{r}}\sqrt{\log(k-r)} + \sqrt{\frac{B(n,k,r)}{r}}\right)^2}{\left(\sqrt{\mathsf{MSR}_{\min}^{(k-r)}} - \sqrt{\frac{2(2\|B\|+\|N\|)\|N\|}{r}}\right)^2} \tag{V.4}$$

where

$$B(n,k,r) = \begin{cases} \sigma_{\min}^2(B)\log(n-k) + (\|B\|_F^2 - r\sigma_{\min}^2(B))\log((n-k)r), \\ \qquad \text{if } r \text{ is a fixed positive integer} \\ \\ \sigma_{\min}^2(B)\frac{r}{2} + (\|B\|_F^2 - r\sigma_{\min}^2(B))\log((n-k)r), \\ \qquad \text{if } \alpha := \lim_{n\to\infty} r/k > 0 \end{cases} \tag{V.5}$$

where

$$\mathsf{MSR}_{\min}^{k-r} = \frac{\|X\|_{(k-r)}^2}{r},$$

and $\|X\|_{(k-r)}^2$ is the $(k-r)$-th value if we are ordering the values of $\|\mathbf{x}^i\|^2$ for $1 \le i \le n$ with descending order. Then, 2-thresholding asymptotically finds a $k-r$ sparsity pattern.

Proof: See Appendix F. ∎

- For noiseless single measurement vector (SMV) case, i.e. $r=1$, if $\mathsf{SNR}_{\min}(Y)\to\infty$, this becomes
$$m > \frac{2(1+\delta)}{\min_{j\in I_t}|x_j|^2}\left(\|\mathbf{x}\|\sqrt{\log(k-1)} + \|\mathbf{b}\|\sqrt{\log(n-k)}\right)^2.$$
Using Lemma E.2 in Appendix E, we have
$$\lim_{n\to\infty}\frac{\|\mathbf{b}\|^2}{\|\mathbf{x}\|^2} = 1. \tag{V.6}$$
Hence, we have
$$m \ge 2(1+\delta)\frac{\|\mathbf{x}\|^2}{\min_{j\in I_t}|x_j|^2}\left(\sqrt{\log(k-1)} + \sqrt{\log(n-k)}\right)^2$$
for some $\delta > 0$, as the sufficient condition for 2-thresholding in SMV cases. Compared to the result in [45] as
$$m \ge 2(1+\delta)\frac{\|\mathbf{x}\|^2}{\min_{j\in \mathrm{supp}X}|x_j|^2}(\sqrt{\log k} + \sqrt{\log(n-k)})^2,$$
our bound has a slight gain due to $\sqrt{\log(k-1)}$ and $\min_{j\in I_t}|x_j|^2$, where $|I_t| = k-1$. This is because even for the SMV problem, the one remaining index can be estimated using the generalized MUSIC criterion.
- If $\|B\|_F^2 = r\sigma_{\min}^2(B)$, r is a fixed number and $\mathsf{SNR}_{\min}(Y)\to\infty$, by using Lemma E.2 in Appendix E, our bound can be simplified as
$$m \ge \frac{2(1+\delta)\|X\|_F^2}{\|X\|_{(k-r)}^2}\left(\sqrt{\log(k-r)} + \sqrt{\frac{\log(n-k)}{r}}\right)^2. \tag{V.7}$$
Therefore, the MMV gain over SMV mainly comes from $\sqrt{(\log(n-k))/r}$.
- If $\|B\|_F^2 = r\sigma_{\min}^2(B)$ and $\lim_{n\to\infty} r/k = \alpha > 0$, then under the condition (V.6) we have
$$m \ge 2(1+\delta)\frac{\|X\|_F^2}{\|X\|_{(k-r)}^2}\left(\sqrt{\log(k-r)} + \frac{1}{\sqrt{2}}\right)^2.$$
Therefore, the $\log(n-k)$ factor disappears, which provides more MMV gain compared to (V.7).

D. Sufficient condition for partial support recovery using subspace S-OMP

Next, we consider the minimum number of measurements for compressive MUSIC with S-OMP. 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 superior performance [32, 47].

1) Initialize $t=0$ and $I_0=\emptyset$.
2) Compute $P_{R(A_{I_t})}^\perp$ which is the projection operator onto the orthogonal complement of the span of $\{\mathbf{a}_j : j \in I_t\}$.
3) Compute $P_{R(A_{I_t})}^\perp B$ and for all $j=1,\cdots,n$, compute $\rho(t,j) = \|\mathbf{a}_j^* P_{R(P_{R(A_{I_t})}^\perp B)}\|_F^2$.
4) Take $j_t = \arg\max_{j=1,\cdots,n}\rho(t,j)$ and $I_{t+1} = I_t \cup \{j_t\}$. If $t<k$ return to Step 2.
5) The final estimate of the sparsity pattern is I_k.

Now, we also consider two cases according to the number of multiple measurement vectors. First, we consider the case when the number of multiple measurement vectors is a finite fixed number. Conventional compressive sensing (the SMV problem) is this kind of case. Second, we consider the case when r is proportional to n. This case includes the conventional MUSIC case.

Theorem 5.3: For $\mathrm{LSMMV}(m,n,k,r;\epsilon)$, let $\mathsf{SNR}_{\min}(Y) = \sigma_{\min}(AX)/\|N\|$ and suppose the following conditions hold:
(a) r is a fixed finite number.
(b) Let $\mathsf{SNR}_{\min}(Y)$ satisfy

$$\mathsf{SNR}_{\min}(Y) > 1 + \frac{4k}{r}(\kappa(B)+1). \tag{V.8}$$

If we have

$$m > k(1+\delta)\left[1 - \frac{4k}{r}\frac{(\kappa(B)+1)}{\mathsf{SNR}_{\min}(Y)-1}\right]^{-1}\frac{2\log(n-k)}{r}, \tag{V.9}$$

then we can find $k-r$ correct indices of $\mathrm{supp}X$ by applying subspace S-OMP.

Proof: See Appendix G. ∎

- As a simple corollary of Theorem 5.3, when $\mathsf{SNR}_{\min}(Y)\to\infty$, we can easily show that the number of sensor elements required for the conventional OMP to find the all k-support indices in SMV problem is given by
$$m > 2(1+\delta)k\log(n-k), \tag{V.10}$$

for a small $\delta > 0$. This is equivalent to the result in [45].

- When $\mathsf{SNR}_{\min}(Y) \to \infty$, then the number of sensor elements for subspace S-OMP is
$$m > 2(1+\delta)\frac{k}{r}\log(n-k)$$
for some $\delta > 0$. Hence, the sampling ratio is the reciprocal of the number of multiple measurement vectors.
- Since $k \to \infty$ in our large system model, (V.8) tells us that the required $\mathsf{SNR}_{\min}(Y)$ should increase to infinity.

Next, we consider the case that r is proportionally increasing with respect k. In this case, we have the following theorem.

Theorem 5.4: For $\mathrm{LSMMV}(m, n, k, r; \epsilon)$, let $\mathsf{SNR}_{\min}(Y) = \sigma_{\min}(AX)/\|N\|$ and suppose the following conditions hold.
(a) r is proportionally increasing with respect to k so that $\alpha := \lim_{n\to\infty} r(n)/k(n) > 0$ exist.
(b) Let $\mathsf{SNR}_{\min}(Y)$ satisfy
$$\mathsf{SNR}_{\min}(Y) > 1 + \frac{4}{\alpha}(\kappa(B)+1). \quad \text{(V.11)}$$
Then if we have
$$m > k(1+\delta)^2 \frac{1}{\left[1 - \frac{4}{\alpha}\frac{\kappa(B)+1}{\mathsf{SNR}_{\min}(Y)-1}\right]^2}\left[2 - F(\alpha)\right]^2, \quad \text{(V.12)}$$
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 $[0,4]$, $0 \le t_1(\alpha) \le 1$ satisfies $\int_0^{2t_1(\alpha)} ds_1(x) = \alpha$ and $ds_1(x) = (1/\pi)\sqrt{4-x^2}$ is a probability measure with support $[0,2]$. Here, $F(\alpha)$ is an increasing function such that $F(1) = 1$ and $\lim_{\alpha\to 0^+} F(\alpha) = 0$. Then we can find $k-r$ correct indices of $\mathrm{supp}X$ by applying subspace S-OMP.

Proof: See Appendix G. ∎

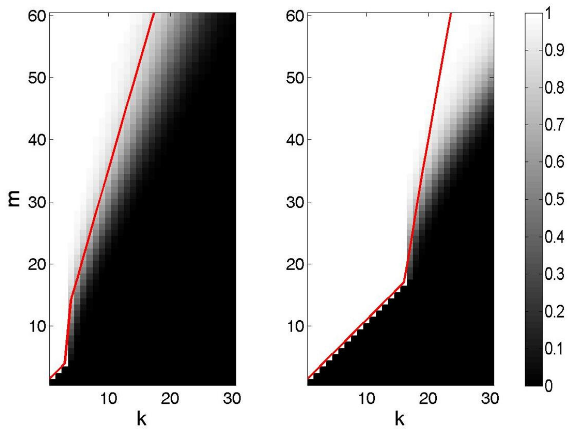
- As a corollary of Theorem 5.4, when $r(n)/k(n) \to 1$ and $\mathsf{SNR}_{\min}(Y) \to \infty$, we can see that the number of sensor elements required for subspace S-OMP to find $k-r$ support indices is given by
$$m > (1+\delta)k,$$
for a small $\delta > 0$, which is the same as the number of sensor elements required for MUSIC.
- We can expect that the number of sensor elements required for subspace S-OMP to find $k-r$ support indices is at most $4(1+\delta)k$ in the noiseless case, where $\delta > 0$ is an arbitrary small number. Hence, the $\log n$ factor is not necessary.
- Unlike the case in Theorem 5.3, the SNR condition is now lower bounded by a finite number $1+(4/\alpha)(\kappa(B)+1)$. This implies that we don't need infinite SNR for support recovery, in contrast to SMV or Theorem 5.3. This is one of the important advantages of MMV over SMV.

VI. Numerical Results

In this section, we demonstrate the performance of compressive MUSIC. This new algorithm is compared to the conventional MMV algorithms, especially 2-SOMP, 2-thresholding and $l_{2,1}$ mixed-norm approach [26]. We do not compare the new algorithm with the classical MUSIC algorithm since it fails when $r < k$. We declared the algorithm as a success if the estimated support is the same as the true $\mathrm{supp}X$, and the success rates were averaged for 5000 experiments. The simulation parameters were as follows: $m \in \{1, 2, \ldots, 60\}$, $n = 200$, $k \in \{1, 2, \ldots, 30\}$, and $r \in \{1, 3, 8, 16\}$, respectively. Elements of sensing matrix A were generated from a Gaussian distribution having zero mean and variance of $1/m$, and the $\mathrm{supp}X$ were chosen randomly. The maximum iteration was set to k for the S-OMP algorithm.

According to (V.2), (V.9) and (V.12), for noiseless measurements, piece-wise continuous boundaries exist for the phase transition of CS-MUSIC with subspace S-OMP:
$$m > \begin{cases} k+1, & r = k; \\ 2k - r + 1, & r < k; \\ (2k\log(n-k))/r & r \ll k. \\ k[2-F(\alpha)]^2, & \lim_{n\to\infty} r/k > 0. \end{cases} \quad \text{(VI.1)}$$
Note that in our canonical MMV model, $r = k$ includes many MMV problems in which the number of snapshots is larger than the sparsity level since our canonical MMV model reduces the effective snapshot r as $r \ge k$. Figure 2(a) shows a typical phase transition map of our compressive MUSIC with subspace S-OMP for noiseless measurements when $n = 200$ and $r = 3$ and $\|\mathbf{x}^i\|$ is constant for all $i = 1, \cdots, n$. Even though the simulation step is not in the large system regime, but r is quite small, so that we can expect that $(2k\log(n-k))/r$ is a boundary for phase transition. Figure 2(b) corresponds to the case when $r = 16$ and $\|\mathbf{x}^i\|$ is constant for all $i = 1, \cdots, n$. Since in this setup r is comparable to k, we use the $k[2-F(\alpha)]^2$ as a boundary. The results clearly indicates the tightness of our sufficient condition.

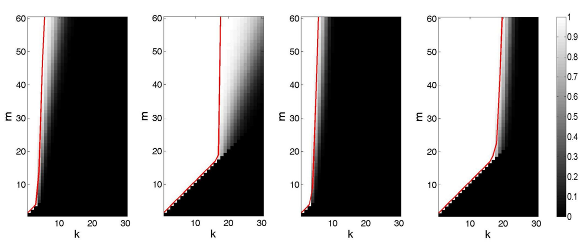
Similarly, multiple piecewise continuous boundaries exist for the phase transition map for compressive MUSIC with 2-thresholding:
$$m > \begin{cases} k+1, & r = k; \\ 2k - r + 1, & r < k; \\ 2\frac{\left(\|X\|_F\frac{\sqrt{\log(n-k)}}{\sqrt{r}} + \sqrt{\frac{B(n,k,r)}{r}}\right)^2}{\mathsf{MSR}_{\min}^{k-r}}, & r < k. \end{cases} \quad \text{(VI.2)}$$
Since the phase transition boundary depends on the unknown joint sparse signal X through $\|X\|_F$ and $\mathsf{MSR}_{\min}^{k-r}$, we investigate this effect. Figure 3(a) and (b) show a typical phase transition map of our compressive MUSIC with 2-thresholding



(a) (b)

Fig. 2. Phase transition map for compressive MUSIC with subspace S-OMP when $n = 200$, $\mathsf{SNR} = \infty$, $\|\mathbf{x}^i\|^2$ is constant for all $i = 1, \cdots, n$, and (a) $r = 3$, and (b) $r = 16$. The overlayed curves are calculated based on (VI.1).

when $r = 3$ and $r = 16$, respectively, for noiseless measurements and $\|\mathbf{x}^i\|$ are constant for all i; Figure 3(c) and (d) corresponds to the same case except $\|\mathbf{x}^i\|^2 = (0.7)^i$. We overlayed theoretically calculated phase boundaries over the phase transition diagram. The empirical phase transition diagram clearly revealed the effect of the distribution X. Still, the theoretically calculated boundary clearly indicates the tightness of our sufficient condition.



(a) (b) (c) (d)

Fig. 3. Phase transition map for compressive MUSIC with 2-thresholding when $n = 200$, $\mathsf{SNR} = \infty$, and (a) $r = 3$, (b) $r = 16$ when $\|\mathbf{x}^i\|^2$ is constant for all $i = 1, \cdots, n$, and (c) $r = 3$, (d) $r = 16$ when $\|\mathbf{x}^i\|^2 = 0.7^{i-1}$. The overlayed curves are calculated based on (VI.2).

Fig. 4 shows the success rate of S-OMP, 2-thresholding, and compressive MUSIC with subspace S-OMP and 2-thresholding for 40dB noisy measurement when $\|\mathbf{x}^i\|$ is constant for all $i = 1, \cdots, n$. When $r = 1$, the performance level of the compressive MUSIC algorithm is basically the same as that of a compressive sensing algorithm such as 2-thresholding and S-OMP. When $r = 8$, the recovery rate of the compressive MUSIC algorithm is higher than the case $r = 1$, and the compressive MUSIC algorithm outperforms the conventional compressive sensing algorithms. If we increase r to 16, the success of the compressive MUSIC algorithm becomes nearly deterministic and approaches the l_0 bound, whereas conventional compressive sensing algorithms do not.

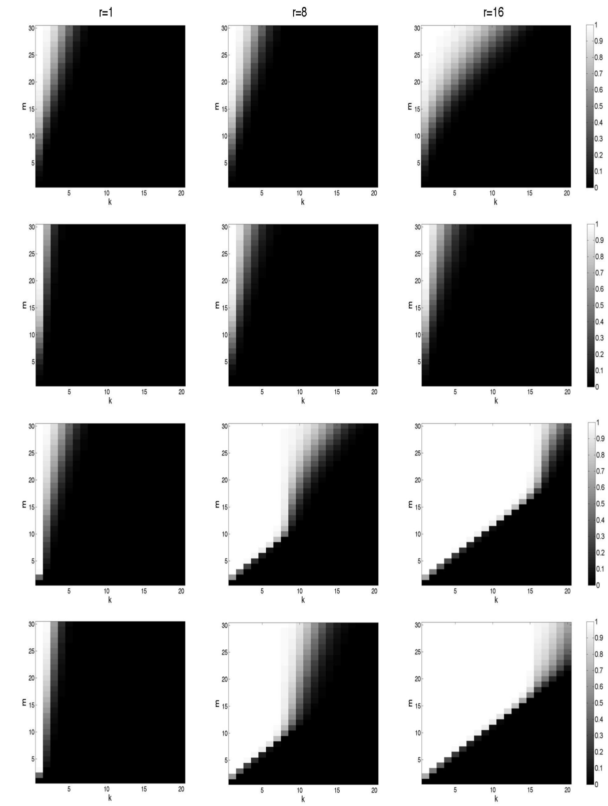
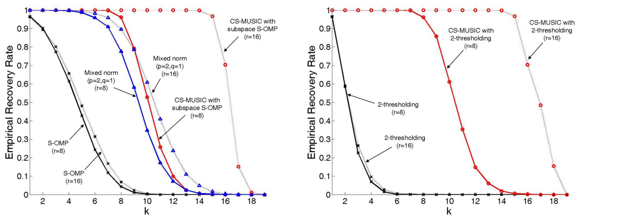


Fig. 4. Recovery rates for various m and k when SNR =40dB and nonzero rows of $\|\mathbf{x}^i\|$ are constant for all i. Each row (from top to bottom) indicates the recovery rates by S-OMP, 2-thresholding, and compressive MUSIC with subspace S-OMP and 2-thresholding. Each column (from left to right) indicates $r = 1, 8$ and 16, respectively.

In order to compare compressive MUSIC with other methods more clearly the recovery rates of various algorithms are plotted in Fig. 5(a) for S-OMP, compressive MUSIC with subspace S-OMP, and the mixed norm approach when $p = 2, q = 1$; and in Fig. 5(b) for 2-thresholding and compressive MUSIC with 2-thresholding, when $n = 200$, $m = 20$ and $r = 8, 16$, $\|\mathbf{x}^i\|$ is constant, and $\mathsf{SNR} = 40$dB. Note that compressive MUSIC outperforms the existing methods.

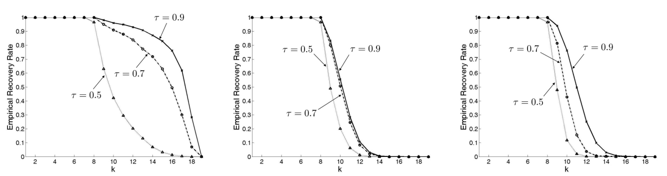
To show the relationship between the recovery performance in the noisy setting and the condition number of matrices X, we performed the simulation on the recovery results for three different types of the source model X. More specifically, the singular values of X are set to be exponentially decaying with (i) $\tau = 0.9$, (ii) $\tau = 0.7$ and (iii) $\tau = 0.5$ respectively, i.e. the singular values of X are given by $\sigma_j = \tau^{j-1}$ for $j = 1, \cdots, \mathrm{rank}(X)$. In this simulation, we are using noisy samples that are corrupted by additive Gaussian noise of SNR =40dB. Figure 6(a) shows the results when $k - r$ entries of the support are known *a priori* by an "oracle"



(a) (b)

Fig. 5. Recovery rates by various MMV algorithms for a uniform source when $n = 200,\ m = 20,\ r = 8$, and 16 and SNR =40dB: (a) recovery rate for S-OMP, compressive MUSIC with subspace S-OMP, and mixed norm approach when $p = 2, q = 1$ and (b) recovery rate for 2-thresholding and compressive MUSIC with 2-thresholding.

algorithm, whereas $k-r$ entries of the support are determined by subspace S-OMP in Fig. 6(b) and by thresholding in Fig. 6(c). The results provide evidence of the significant impact of the condition number of X.



(a) (b) (c)

Fig. 6. Recovery rates by compressive MUSIC when $k-r$ nonzero supports are estimated by (a) an "oracle" algorithm, (b) subspace S-OMP, and (c) 2-thresholding. Here, X is given with $\tau = 0.9$, $\tau = 0.7$ and $\tau = 0.5$. Smaller τ provides larger condition number $\kappa(X)$. The measurements are corrupted by additive Gaussian noise of SNR =40dB and $n = 200,\ m = 20,\ r = 8$.

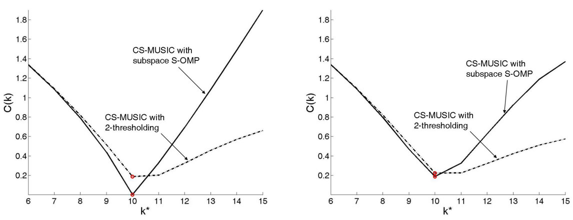
Figure 7 illustrates the cost function to estimate the unknown sparsity level, which confirms that compressive MUSIC can accurately estimate the unknown sparsity level k as described in Lemma 4.6. In this simulation, $n = 200,\ m = 40$ and $r = 5$. The correct support size k is marked as circle. Note that $C(\hat{k})$ has the smallest value at that point for the noiseless measurement cases, as shown Fig. 7(a), confirming our theory. For the 40dB noisy measurement case, we can still easily find the correct k since it corresponds to the first local minimizer as $\hat{k}$ increases, as shown in Fig. 7(b).

VII. Discussion

A. Comparison with subspace-augmented MUSIC [47]

Recently, Lee and Bresler [47] independently developed a hybrid MMV algorithm called as subspace-augmented MUSIC (SA-MUSIC). The SA-MUSIC performs the following procedure.

1) Find $k-r$ indices of $\text{supp}X$ by applying SOMP to the MMV problem $U = AX$ where the set of columns of U is an orthonormal basis for $R(B)$.
2) Let I_{k-r} be the set of indices which are taken in Step 1 and $S = I_{k-r}$.



(a) (b)

Fig. 7. Cost function for sparsity estimation when $n = 200,\ m = 40,\ r = 5,\ k = 10$, and the measurements are (a) noiseless and (b) corrupted by additive Gaussian noise of SNR =40dB. The circles illustrate the local minima, whose position corresponds to the true sparsity level.

3) For $j \in \{1, \cdots, n\} \setminus I_{k-r}$, compute $\eta(j) = \|\tilde{Q}^* \mathbf{a}_j\|^2$ where $\tilde{Q} \in \mathbb{R}^{m \times (m-k)}$ consists of orthonormal columns such that $\tilde{Q}^*[U \quad A_{I_{k-r}}] = 0$.
4) Make an ascending ordering of $\eta(j)$, $j \notin I_{k-r}$, choose indices that correspond to the first r elements, and put these indices into S.

By Theorem 4.5(c), we can see that the subspace-augmented MUSIC is equivalent to compressive MUSIC, sinces the MUSIC criterion in subspace-augmented measurement $[U\ A_{I_{k-r}}]$ and the generalized MUSIC criterion in compressive MUSIC are equivalent. Therefore, we can expect that the performance of both algorithm should be similar except for the following differences. First, the subspace S-OMP in Lee and Bresler [47] is applying the subspace decomposition once for the data matrix Y whereas our analysis for the subspace S-OMP is based on subspace decomposition for the residual matrix at each step. Hence, our subspace S-OMP is more similar to that of [32]. However, based on our experiments the two versions of the subspace S-OMP provide similar performance when combined with the generalized MUSIC criterion. Second, the theoretical analysis of SA-MUSIC is based on the RIP condition whereas ours is based on large system limit model. One of the advantage of RIP based analysis is its generality for any type of sensing matrices. However, our large system analysis can provide explicit bounds for the number of required sensor elements and SNR requirement thanks to the Gaussian nature of sensing matrix.

B. Comparison with results of Davies and Eldar [32]

Another recent development in joint sparse recovery approaches is the rank-awareness algorithm by Davies and Eldar [32]. The algorithm is derived in the noiseless measurement setup and is basically the same as our subspace S-OMP in Section 5 except that $\mathbf{a}_j$ in step 3 is normalized after applying $R^{\perp}_{R(A_{I_t})}$ to the original dictionary A. For the full rank measurement, i.e. $r = k$, the performance of the rank-aware subspace S-OMP is equivalent to that of MUSIC. However, for $r < k$, the lack of the generalized MUSIC criterion may make the algorithm inferior since our generalized MUSIC criterion can identify r support deterministically whereas the rank-aware subspace S-OMP should estimate the remaining r support with additional error-prone greedy steps.

C. Compressive MUSIC with a mixed norm approach

Another important issue in CS-MUSIC is how to combine the general MUSIC criterion with non-greedy joint sparse recovery algorithms such as a mixed norm approach [26]. Towards this, the $k-r$ greedy step required for the analysis for CS-MUSIC should be modified. One solution to mitigate this problem is to choose a $k-r$ support from the non-zero support of the solution and use it as a partial support for generalized MUSIC criterion. However, we still need a criterion to identify a correct $k-r$ support from the solution, since the generalized MUSIC criterion only holds with a correct $k-r$ support. Recently, we showed that a correct $k-r$ partial support out of k-sparse solution can be identified using a subspace fitting criterion [48]. Accordingly, the joint sparse recovery problem can be relaxed to a problem to find a solution that has at least $k-r+1$ correct support out of k nonzero support estimate. This is a significant relaxation of CS-MUSIC in its present form that requires $k-r$ successful consecutive greedy steps. Accordingly, the new formulation was shown to significantly improve the performance of CS-MUSIC for the joint-sparse recovery [48]. However, the new results are beyond scope of this paper and will be reported separately.

D. Relation with distributed compressive sensing coding region

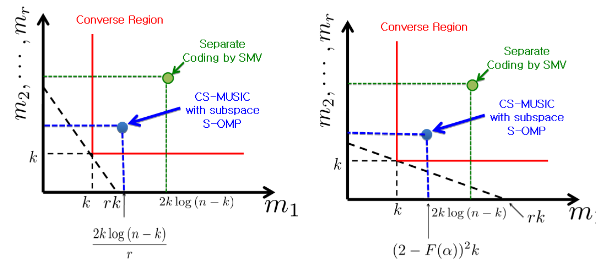
Our theoretical results as well as numerical experiments indicate that the number of resolvable sources can increase thanks to the exploitation of the noise subspace. This observation leads us to investigate whether CS-MUSIC achieves the rate region in distributed compressed sensing [20], which is analogous to Slepian-Wolf coding regions in distributed source coding [49].

Recall that the necessary condition for a maximum likelihood for SMV sparse recovery is given by [45]:

$$m > \frac{2k\log(n-k)}{\mathsf{SNR}\cdot\mathsf{MSR}^{\mathsf{k}}_{\min}} + k - 1 \longrightarrow k-1,$$

as $\mathsf{SNR}\rightarrow\infty$. Let m_i denote the number of sensor elements at the i-th measurement vector. If the total number of samples from r vectors are smaller than that of SMV-CS, i.e. $\sum_{i=1}^r m_i < k-1$, then we cannot expect a perfect recovery even from noiseless measurement vectors. Furthermore, the minimum sensor elements should be $m_i = k$ to recover the values of the i-th coefficient vector, even when the k indices of the support are correctly identified. Hence, the converse region at $\mathsf{SNR}\rightarrow\infty$ is defined by the $m_i < k, i = 1,\cdots,r$ as shown Fig. 8(a)(b).

Now, for a fixed r our analysis shows that the achievable rate by the CS-MUSIC is $m_i = 2k\log(n-k)/r$ (Fig. 8(a)). On the other hand, if $\lim_{n\rightarrow\infty} r/k = \alpha > 0$, the achievable rate by the CS-MUSIC is $m_i = (2-F(\alpha))^2k$ as shown in Fig. 8(b). Therefore, CS-MUSIC approaches the converse regin at $r=k$, whereas for the intermediate ranges of r there exists a performance gap from the converse region. However, even in this case if we consider a separate SMV decoding without considering correlation structure in MMV, the required sampling rate is $m_i \geq 2k\log(n-k)$ which is significantly larger than that of CS-MUSIC. This analysis clearly reveals that CS-MUSIC is a quite efficient decoding method from distributed compressed sensing perspective.



(a) (b)

Fig. 8. Rate regions for the multiple measurement vector problem and CS-MUSIC, when (a) r is a fixed number, and (b) $\lim_{n\rightarrow\infty} r/k = \alpha > 0$.

E. Discretization

The MUSIC algorithm was originally developed for spectral estimation or direction-of-arrival (DOA) estimation problem, where the unknown target locations and bearing angle are continuously varying parameters. If we apply CS-MUSIC to this type of problems to achieve a finer resolution, the search region should be discretized more finely with a large n. The main problem of such discretization is that the mutual coherence of the dictionary A approaches to 1, which can violate the RIP condition of the CS-MUSIC. Therefore, the trade-off between the resolution and the RIP condition should be investigated; Duarte and Baraniuk recently investigated such trade-off in the context of spectral compressive sensing [50]. Since this problem is very important not only for the CS-MUSIC but for SMV compressed sensing problems that are originated from discretizing continuous problems, systematic study needs to be done in the future.

VIII. Conclusions and Future Works

In this paper, we developed a novel compressive MUSIC algorithm that outperforms the conventional MMV algorithms. The algorithm estimates $k-r$ entries of the support using conventional MMV algorithms, while the remaining r support indices are estimated using a generalized MUSIC criterion, which was derived from the RIP properties of sensing matrix. Theoretical analysis as well as numerical simulation demonstrated that our compressive MUSIC algorithm achieved the l_0 bound as r approaches the non-zero support size k. This is fundamentally different from existing information theoretic analysis [51], which requires the number of snapshots to approach infinity to achieve the l_0 bound. Furthermore, as r approaches 1, the recovery rate approaches that of the conventional SMV compressive sensing. We also provided a method that can estimate the unknown sparsity, even under noisy measurements. Theoretical analysis based on a large system MMV model showed that the required number of sensor elements for compressive MUSIC is much smaller than that of conventional MMV compressive sensing. Furthermore, we provided a closed form expression of the minimum SNR to guarantee the success of compressive MUSIC.

The compressive sensing and array signal processing produce two extreme approaches for the MMV problem: one is based on a probabilistic guarantee, the other on a deterministic guarantee. One important contribution of this paper is to abandon such extreme viewpoints and propose an optimal method to take the best of both worlds. Even though the resulting idea appears simple, we believe that this opens a new area of research. Since extensive research results are available from the array signal processing community, combining the already well-established results with compressive sensing may produce algorithms that may be superior to the compressive MUSIC algorithm in its present form. Another interesting observation is that the RIP condition $\delta^L_{2k-r+1} < 1$, which is essential for compressive MUSIC to achieve the l_0 bound, is identical to the l_0 recovery condition for the so-called modified CS [52]. In modified CS, r support indices are known *a priori* and the remaining $k-r$ are estimated using SMV compressive sensing. The duality between compressive MUSIC and the modified CS does not appear incidental and should be investigated. Rather than estimating $k-r$ indices first using MMV compressive sensing and estimating the remaining r using the generalized MUSIC criterion, there might be a new algorithm that estimates r supports indices first in a deterministic pattern, while the remaining $k-r$ are estimated using compressive sensing. This direction of research might reveal new insights about the geometry of the MMV problem.

Appendix A: Proof of Theorem 4.1

Proof: (a) First, we show that $\mathrm{spark}(Q^*A) \leq k-r+1$. Since $Q^*AX = 0$, we have $Q^*A\mathbf{x}_i = 0$ for $1 \leq i \leq r$. Take a set $P \subset \mathrm{supp}X$ with $|P| = r-1$. Then, there exists a nonzero $\mathbf{c} = [c_1, \cdots, c_r] \in \mathbb{R}^r$ such that

$$X^P\mathbf{c} = 0, \text{ where } X^P \in \mathbb{R}^{(r-1)\times r}, \tag{A.1}$$

where X^P denotes a submatrix collecting rows corresponding to the index set P.

Since the columns of X are linearly independent, $\sum_{i=1}^r c_i\mathbf{x}_i \neq 0$ but $Q^*A(\sum_{i=1}^r c_i\mathbf{x}_i) = 0$. By (A.1),

$$\|\sum_{i=1}^r c_i\mathbf{x}_i\|_0 \leq k-r+1$$

so that $\mathrm{spark}(Q^*A) \leq k-r+1$.

(b) Suppose that there is $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

$$Q^*A\mathbf{x} = 0, \ \|\mathbf{x}\|_0 \leq k-r+1 \text{ and } \mathrm{supp}(\mathbf{x}) \not\subseteq \mathrm{supp}X.$$

Since $Q^*A\mathbf{x} = 0$, $A\mathbf{x} \in R(Q)^\perp = R(B)$ so that there is a $\tilde{\mathbf{x}}$ such that $A\mathbf{x} = A\tilde{\mathbf{x}}$ and $\mathrm{supp}(\tilde{\mathbf{x}}) \subset \mathrm{supp}X$. Hence, we have

$$A(\mathbf{x}-\tilde{\mathbf{x}}) = 0, \ \|\mathbf{x}-\tilde{\mathbf{x}}\|_0 \leq 2k-r+1.$$

By the RIP condition $0 \leq \delta^L_{2k-r+1}(A) < 1$, $\mathbf{x} = \tilde{\mathbf{x}}$. It follows that whenever $\|\mathbf{x}\|_0 \leq k-r+1$ and $Q^*A\mathbf{x} = 0$, we have $\mathrm{supp}(\mathbf{x}) \subset \mathrm{supp}X$. Since $A\mathbf{x} \in R(B) = R(AX)$, there is a $\mathbf{y} \in R(X)$ such that $A\mathbf{x} = A\mathbf{y}$. Hence, if $Q^*A\mathbf{x} = 0$ and $\|\mathbf{x}\|_0 \leq k-r+1$, by the RIP condition of A, we have $\mathbf{x} \in R(X)$.

Finally, it suffices to show that for any $\mathbf{x} \in R(X) \setminus \{0\}$,

$$\|\mathbf{x}\|_0 \geq k-r+1.$$

Suppose that $\|\mathbf{x}\|_0 \leq k-r$. Then there is a set Z such that $|Z| = r$ and $Z \subset \mathrm{supp}X \setminus \mathrm{supp}(\mathbf{x})$. Then, there exists a $\mathbf{c} \in \mathbb{R}^r \setminus \{0\}$ such that

$$X^Z\mathbf{c} = 0, \text{ where } X^Z \in \mathbb{R}^{r\times r}.$$

This is impossible since the nonzero rows of X are in general position.

■

Appendix B: Proof of Theorem 4.3 and Corollary 4.4

Proof of Theorem 4.3: In order to show that (IV.2) implies $j \in \mathrm{supp}X$, let I_{k-r} be an index set with $|I_{k-r}| = k-r$ and $I_{k-r} \subset \mathrm{supp}X$. Then by Lemma 4.1 and the definition of the spark(A),

$$\mathrm{rank}[Q^*A_{I_{k-r}}] = k-r.$$

By the assumption, there is an $\mathbf{x}_{k-r} \in \mathbb{R}^{k-r}$ such that $Q^*\mathbf{a}_j = Q^*A_{I_{k-r}}\mathbf{x}_{k-r}$ so that we have

$$Q^*[\mathbf{a}_j - A_{I_{k-r}}\mathbf{x}_{k-r}] = 0.$$

Since $\mathbf{a}_j - A_{I_{k-r}}\mathbf{x}_{k-r} \in N(Q^*) = R(Q)^\perp = R(B)$, there is a $\tilde{\mathbf{x}} \in \mathbb{R}^n$ such that $\mathrm{supp}(\tilde{\mathbf{x}}) \subset \mathrm{supp}X$ and $\mathbf{a}_j - A_{I_{k-r}}\mathbf{x}_{k-r} = A\tilde{\mathbf{x}}$. Hence we have $\mathbf{y} \in \mathbb{R}^n$ such that $A\mathbf{y} = 0$ and $\mathrm{supp}(\mathbf{y}) \subset \mathrm{supp}X \cup \{j\} \cup I_{k-r}$ so that $\|\mathbf{y}\| \leq 2k-r+1$. By the RIP condition $0 \leq \delta^L_{2k-r+1}(A) < 1$, it follows that $\{j\} \cup \mathrm{supp}(\mathbf{x}_{k-r}) = \mathrm{supp}(\tilde{\mathbf{x}}) \subset \mathrm{supp}X$ since $j \notin I_{k-r}$. Hence, under the condition (IV.2), we have $j \in \mathrm{supp}X$.

In order to show that $j \in \mathrm{supp}X$ implies (IV.2), assume the contrary. Then we have

$$\mathrm{rank}(Q^*[A_{I_{k-r}}, \mathbf{a}_j]) = k-r+1,$$

where $I_{k-r} \subset \mathrm{supp}X$ with $|I_{k-r}| = k-r$. Then for any $\mathbf{x}_{k-r} \in \mathbb{R}^{k-r}$, $Q^*[\mathbf{a}_j - A_{I_{k-r}}\mathbf{x}_{k-r}] \neq 0$ so that

$\mathbf{a}_j - A_{I_{k-r}}\mathbf{x}_{k-r} \notin R(B)$. Set $P = \text{supp}X \setminus (I_{k-r} \cup \{j\})$ so that $|P| = r-1$. Then there is a $\mathbf{c} \in \mathbb{R}^r \setminus \{0\}$ such that

$$X^P\mathbf{c} = 0, \text{ where } X^P \in \mathbb{R}^{(r-1)\times r}.$$

Then we have $\|X\mathbf{c}\|_0 = k-r+1$ since the rows of X are in general position. Note that $\text{supp}(X\mathbf{c}) = \{j\} \cup I_{k-r}$. Since $AX\mathbf{c} \in R(B)$, $\mathbf{a}_j - A_{I_{k-r}}\mathbf{x}_{k-r} \in R(B)$ for some $\mathbf{x}_{k-r} \in \mathbb{R}^{k-r}$, which is a contradiction.

Proof of Corollary 4.4: Here we let $G_{I_{k-r}} := Q^*A_{I_{k-r}}$ and $\mathbf{g}_j = Q^*\mathbf{a}_j$. Since we already have $\text{rank}[G_{I_{k-r}}] = k-r$, (IV.2) holds if and only if

$$\det\,[G_{I_{k-r}}, \mathbf{g}_j]^*[G_{I_{k-r}}, \mathbf{g}_j] = 0.$$

Note that

$$\begin{bmatrix} G^*_{I_{k-r}} \\ \mathbf{g}^*_j \end{bmatrix} [G_{I_{k-r}}, \mathbf{g}_j] = \begin{bmatrix} A^*_{I_{k-r}} \\ \mathbf{a}^*_j \end{bmatrix} QQ^*[A_{I_{k-r}}, \mathbf{a}_j]$$
$$= \begin{bmatrix} A^*_{I_{k-r}}P_{R(Q)}A_{I_{k-r}} & A^*_{I_{k-r}}P_{R(Q)}\mathbf{a}_j \\ \mathbf{a}^*_jP_{R(Q)}A_{I_{k-r}} & \mathbf{a}^*_jP_{R(Q)}\mathbf{a}_j \end{bmatrix},$$

where $\det\,[A^*_{I_{k-r}}P_{R(Q)}A_{I_{k-r}}] > 0$ because of $\text{rank}[G_{I_{k-r}}] = k-r$. Since

$$\begin{aligned}
&\det \begin{bmatrix} A^*_{I_{k-r}}P_{R(Q)}A_{I_{k-r}} & A^*_{I_{k-r}}P_{R(Q)}\mathbf{a}_j \\ \mathbf{a}^*_jP_{R(Q)}A_{I_{k-r}} & \mathbf{a}^*_jP_{R(Q)}\mathbf{a}_j \end{bmatrix} \\
&= \det(A^*_{I_{k-r}}P_{R(Q)}A_{I_{k-r}})\det(\mathbf{a}^*_jP_{R(Q)}\mathbf{a}_j \\
&\quad - \mathbf{a}^*_jP_{R(Q)}A_{I_{k-r}}(A^*_{I_{k-r}}P_{R(Q)}A_{I_{k-r}})^{-1}A^*_{I_{k-r}}P_{R(Q)}\mathbf{a}_j),
\end{aligned}$$

(IV.2) is equivalent to

$$\begin{aligned}
&\mathbf{a}^*_jP_{R(Q)}\mathbf{a}_j \\
&\quad - \mathbf{a}^*_jP_{R(Q)}A_{I_{k-r}}(A^*_{I_{k-r}}P_{R(Q)}A_{I_{k-r}})^{-1}A^*_{I_{k-r}}P_{R(Q)}\mathbf{a}_j \\
&=\mathbf{a}^*_jQQ^*\mathbf{a}_j - \mathbf{a}^*_jQP_{R(Q^*A_{I_{k-r}})}Q^*\mathbf{a}_j \\
&=\mathbf{a}^*_j\left[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}\right]\mathbf{a}_j = 0,
\end{aligned}$$

where $P_{R(Q)} = QQ^*$. Hence (IV.3) holds if and only if $j \in \text{supp}X$.

Appendix C: Proof of Theorem 4.5 and Lemma 4.6

Proof of Theorem 4.5: (a) By the definitions of U and Q, we have $U^*Q = 0$ so that

$$\begin{aligned}
&[UU^* + P_{QQ^*A_{I_{k-r}}}]^2 \\
=&UU^* \\
&+ UU^*QQ^*A_{I_{k-r}}(A^*_{I_{k-r}}QQ^*A_{I_{k-r}})^{-1}A^*_{I_{k-r}}QQ^* \\
&+ QQ^*A_{I_{k-r}}(A^*_{I_{k-r}}QQ^*A_{I_{k-r}})^{-1}A^*_{I_{k-r}}QQ^*UU^* \\
&\quad + QQ^*A_{I_{k-r}}(A^*_{I_{k-r}}QQ^*A_{I_{k-r}})^{-1}A^*_{I_{k-r}}Q \\
&\quad\quad \times Q^*A_{I_{k-r}}(A^*_{I_{k-r}}QQ^*A_{I_{k-r}})^{-1}A^*_{I_{k-r}}QQ^* \\
=&UU^* + P_{QQ^*A_{I_{k-r}}}.
\end{aligned}$$

Since $UU^* + P_{QQ^*A_{I_{k-r}}}$ is a self-adjoint matrix, it is an orthogonal projection. Next, to show that $R(UU^* + P_{QQ^*A_{I_{k-r}}}) = R(B) + R(QQ^*A_{I_{k-r}})$, we only need to show the following properties :

(i) $[UU^* + P_{QQ^*A_{I_{k-r}}}]\mathbf{b} = \mathbf{b}$ for any $\mathbf{b} \in R(B)$,
(ii) $[UU^* + P_{QQ^*A_{I_{k-r}}}]\mathbf{q}_1 = \mathbf{q}_1$ for any $\mathbf{q}_1 \in R(QQ^*A_{I_{k-r}})$,
(iii) $[UU^* + P_{QQ^*A_{I_{k-r}}}]\mathbf{q}_2 = \mathbf{0}$ for any $\mathbf{q}_2 \in R(Q) \cap R(QQ^*A_{I_{k-r}})^\perp$.

For (i), it can be easily shown by using $Q^*\mathbf{b} = \mathbf{0}$ and $UU^*\mathbf{b} = \mathbf{b}$ for any $\mathbf{b} \in R(B)$. For (ii), any $\mathbf{q}_1 \in R(QQ^*A_{I_{k-r}})$, there is a $\mathbf{w} \in \mathbb{R}^{k-r}$ such that $\mathbf{q}_1 = QQ^*A_{I_{k-r}}\mathbf{w}$. Then by using the property $U^*Q = 0$, we can see that property (ii) also holds. Finally, we can easily see that (iii) also holds by using $U^*\mathbf{q} = 0$ for any $\mathbf{q} \in R(Q)$.

(b) This is a simple consequence of (a) since $QQ^* - P_{QQ^*A_{I_{k-r}}} = I - [UU^* + P_{QQ^*A_{I_{k-r}}}]$ and $R(Q) \cap R(QQ^*A_{I_{k-r}})^\perp$ is an orthogonal complement of $R(B) + R(QQ^*A_{I_{k-r}})$.

(c) Since $\text{spark}(Q^*A) = k-r+1$ by Lemma 4.1, $[U \;\; A_{I_{k-r}}]$ has k linearly independent columns. Hence we only need to find the orthogonal complement of $R([U \;\; A_{I_{k-r}}]) = R(U) + R(A_{I_{k-r}}) = R(B) + R(A_{I_{k-r}})$. Since $R(U)^\perp = R(Q)$, we have $R(U) + R(A_{I_{k-r}}) = R(U) + R(P_QA_{I_{k-r}})$ by the projection update rule so that $(R(U) + R(QQ^*A_{I_{k-r}}))^\perp = R(Q) \cap R(QQ^*A_{I_{k-r}})^\perp$ is the noise subspace for $[U \;\; A_{I_{k-r}}]$ or $[B \;\; A_{I_{k-r}}]$.

Appendix D: Proof of Theorem 5.1

We first need to show the following lemmas.

Lemma D.1: Assume that we have noisy measurement through multiple noisy snapshots where

$$Y = AX + N,$$

where $A \in \mathbb{R}^{m\times n}$, $X \in \mathbb{R}^{n\times r}$, and $N \in \mathbb{R}^{m\times r}$ is additive noise. We also assume that $I_{k-r} \subset \text{supp}X$. Then there is a $\eta > 0$ such that for any $j \notin \text{supp}X$ and $l \in \text{supp}X$,

$$\begin{aligned}
&\mathbf{a}^*_j\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j \\
&\quad > \mathbf{a}^*_l\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_l \qquad \text{(D.1)}
\end{aligned}$$

if $\|N\| < \eta$, where $\|N\|$ is a spectral norm of N and $\hat{Q} \in \mathbb{R}^{m\times(m-k)}$ consists of orthonormal columns such that $\hat{Q}^*Y = 0$.

Proof: First, here we let $B = AX$, $\sigma_{\min}(B)$ (or $\sigma_{\max}(B)$) be the minimum (or the maximum) nonzero singular value of B. Then, $Y = B + N$ is also of full column rank if $\|N\| < \sigma_{\min}(B)$. For such an N,

$$\begin{aligned}
&\|P_{R(Y)} - P_{R(B)}\| \\
=&\ \|Y(Y^*Y)^{-1}Y^* - B(B^*B)^{-1}B^*\| \\
=&\|(B+N)[(B+N)^*(B+N)]^{-1}(B+N)^* - B(B^*B)^{-1}B^*\|
\end{aligned}$$

$$
\begin{aligned}
&\le \|N\|\|[(B+N)^*(B+N)]^{-1}(B+N)^*\| \\
&\quad + \|(B+N)[(B+N)^*(B+N)]^{-1}\| \\
&\quad\quad \times \|(B+N)^*(B+N) - B^*B\|\|(B^*B)^{-1}B^*\| \\
&\quad + \|B(B^*B)^{-1}\|\|N\| \\
&\le \|N\|\|(B+N)^\dagger\| + \|(B+N)^\dagger\|\|B^\dagger\|[2\|B\|\|N\| + \|N\|^2] \\
&\quad + \|B^\dagger\|\|N\|
\end{aligned}
$$

by the consecutive use of triangle inequality. If we have $\|N\| < \sigma_{\min}(B)$, we get

$$
\|(B+N)^\dagger\| \le (\sigma_{\min}(B) - \|N\|)^{-1}
$$

so that

$$
\begin{aligned}
&\frac{\|P_{R(Y)} - P_{R(B)}\|}{\|Y-B\|} \\
\le\ & \frac{1}{\sigma_{\min}(B) - \|N\|} \\
& + \frac{1}{\sigma_{\min}(B)(\sigma_{\min}(B) - \|N\|)}[2\|B\| + \|N\|] + \frac{1}{\sigma_{\min}(B)} \\
=\ & \frac{2(\sigma_{\max}(B) + \sigma_{\min}(B))}{\sigma_{\min}(B)(\sigma_{\min}(B) - \|N\|)} \qquad \text{(D.2)}
\end{aligned}
$$

where we use $\|B^\dagger\| = 1/(\sigma_{\min}(B))$ and $\|B\| = \sigma_{\max}(B)$. By the projection update rule, we have

$$
\begin{aligned}
P_{R([B \quad A_{I_{k-r}}])} &= P_{R(B)} + P_{R(P^\perp_{R(B)}A_{I_{k-r}})} \\
&= I - \left[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}\right] \\
&= P_{R(A_{I_{k-r}})} + P_{R(P^\perp_{R(A_{I_{k-r}})}B)}, \qquad \text{(D.3)}
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
P_{R([Y \quad A_{I_{k-r}}])} &= P_{R(Y)} + P_{R(P^\perp_{R(Y)}A_{I_{k-r}})} \\
&= I - \left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right] \\
&= P_{R(A_{I_{k-r}})} + P_{R(P^\perp_{R(A_{I_{k-r}})}Y)}. \qquad \text{(D.4)}
\end{aligned}
$$

By applying (D.3) and (D.4) as done in [47], we have

$$
\begin{aligned}
&\|[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}] - [P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}]\| \\
&= \|P_{R(P^\perp_{R(A_{I_{k-r}})}Y)} - P_{R(P^\perp_{R(A_{I_{k-r}})}B)}\| \\
&\le \|P_{R(Y)} - P_{R(B)}\|. \qquad \text{(D.5)}
\end{aligned}
$$

Then, for any $j \notin \mathrm{supp}X$ and $l \in \mathrm{supp}X$, by the generalized MUSIC criterion (IV.3) we have

$$
\begin{aligned}
& \mathbf{a}_j^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j \\
& -\mathbf{a}_l^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_l \\
=\ & \mathbf{a}_j^*\left[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}\right]\mathbf{a}_j \\
& -\mathbf{a}_l^*\left[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}\right]\mathbf{a}_l \\
& +\mathbf{a}_j^*\Big[[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}] \\
& \qquad -[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}]\Big]\mathbf{a}_j \\
& -\mathbf{a}_l^*\Big[[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}] \\
& \qquad -[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}]\Big]\mathbf{a}_l \\
\ge\ & \min_{j\notin \mathrm{supp}X} \mathbf{a}_j^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j \\
& -2\max(\|\mathbf{a}_j\|^2, \|\mathbf{a}_l\|^2)\|P_{R(Y)} - P_{R(B)}\| \\
\ge\ & \min_{j\notin \mathrm{supp}X} \mathbf{a}_j^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j \\
& -2\max_{1\le j\le n}\|\mathbf{a}_j\|^2 \frac{2(\sigma_{\max}(B) + \sigma_{\min}(B))\|N\|}{\sigma_{\min}(B)(\sigma_{\min}(B) - \|N\|)} > 0
\end{aligned}
$$

if we have

$$
\|N\| < \frac{\sigma^2_{\min}(B)\zeta}{4(\sigma_{\max}(B) + \sigma_{\min}(B)) + \sigma_{\min}(B)\zeta} \qquad \text{(D.6)}
$$

where

$$
\zeta := \frac{\min_{j\notin \mathrm{supp}X} \mathbf{a}_j^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j}{\max_{1\le j\le n}\|\mathbf{a}_j\|^2}.
$$

■

Lemma D.2: Suppose a minimum SNR is given by

$$
\mathsf{SNR}_{\min}(Y) := \frac{\sigma_{\min}(B)}{\|N\|} \ge \eta,
$$

where

$$
\eta := 1 + \frac{4(\kappa(B) + 1)}{\zeta},
$$

$\kappa(B)$ is the condition number of $B = AX$ and

$$
\zeta := \frac{\min_{j\notin \mathrm{supp}X} \mathbf{a}_j^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j}{\max_{1\le j\le n}\|\mathbf{a}_j\|^2}.
$$

Then, for any $j \notin \mathrm{supp}X$ and $l \in \mathrm{supp}X$,

$$
\begin{aligned}
\mathbf{a}_j^*\left[P_{R(\hat{Q})} \right.&- \left.P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_j \\
&> \mathbf{a}_l^*\left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right]\mathbf{a}_l.
\end{aligned}
$$

Proof: Using (D.6), the generalized MUSIC correctly estimates the r remaining indices when

$$
\|N\| < \frac{\sigma_{\min}(B)\zeta}{4(\kappa(B) + 1) + \zeta}
$$

where we use the definition of the condition number of the $B = AX$ matrix, i.e. $\kappa(AX) = \frac{\sigma_{\max}(B)}{\sigma_{\min}(B)}$. This implies that

$$
\mathsf{SNR}_{\min}(Y) > 1 + \frac{4(\kappa(B) + 1)}{\zeta}.
$$

This concludes the proof. ■

Corollary D.3: For a $\mathrm{LSMMV}(m, n, k, r; \epsilon)$, if we have

$I_{k-r} \subset \text{supp}X$ and a minimum SNR satisfies

$$\mathsf{SNR}_{\min}(Y) > 1 + \frac{4(\kappa(B)+1)}{1-\gamma^2} \geq 1 + 4(\kappa(B)+1) \quad \text{(D.7)}$$

where $\gamma = \lim_{n\to\infty} \sqrt{k(n)/m(n)}$, then we can find remaining r indices of $\text{supp}X$ with generalized MUSIC criterion.

Proof: It is enough to show that

$$\lim_{n\to\infty} \zeta(n) = 1 - \gamma^2.$$

First, for each $1 \leq j \leq n$, $m\|\mathbf{a}_j\|^2$ is a chi-square random variable with degree of freedom m so that we have by Lemma 3 in [45],

$$\lim_{n\to\infty} \frac{\max_{1\leq j\leq n} \|\mathbf{a}_j\|^2}{m} = 1$$

since $\lim_{n\to\infty}(\log n)/m = 0$. On the other hand, for any $j \notin \text{supp}X$, $\mathbf{a}_j$ is independent from $P_{R(\hat{Q})} - P_{R(P_{R(Q)}A_{I_{k-r}})}$ so that $m\mathbf{a}_j^* \left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right] \mathbf{a}_j$ is a chi-square random variable with degree of freedom $m - k$ since $P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}$ is the projection operator onto the orthogonal complement of $R[B \ A_{I_{k-r}}]$. Since $\lim_{n\to\infty}(\log(n-k))/(m-k) = 0$, again by Lemma 3 in [45], we have

$$\lim_{n\to\infty} \frac{\min_{j\notin\text{supp}X} \mathbf{a}_j^* \left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right] \mathbf{a}_j}{m-k} = 1$$

so that

$$\begin{aligned} &\lim_{n\to\infty} \frac{\min_{j\notin\text{supp}X} \mathbf{a}_j^*[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}]\mathbf{a}_j}{\max_{1\leq j\leq n} \|\mathbf{a}_j\|^2} \\ = &\lim_{n\to\infty} \left(\frac{\min_{j\notin\text{supp}X} \mathbf{a}_j^* \left[P_{R(\hat{Q})} - P_{R(P_{R(\hat{Q})}A_{I_{k-r}})}\right] \mathbf{a}_j}{m-k} \right. \\ &\left. \times \frac{m}{\max_{1\leq j\leq n} \|\mathbf{a}_j\|^2} \frac{m-k}{m} \right) = 1 - \gamma^2 \leq 1. \end{aligned}$$

■

Proof of Theorem 5.1: First, we need to show the left RIP condition $0 \leq \delta_{2k-r+1}^L < 1$ to apply the generalized MUSIC criterion. Using Marćenko-Pastur theorem [53], we have

$$\limsup_{n\to\infty} \delta_{2k-r+1}^L = 1 - \liminf_{n\to\infty} (1 - \sqrt{(2k-r+1)/m})^2 < 1.$$

Hence, we need $m \geq (1+\delta)(2k-r+1)$ to make $\limsup_{n\to\infty} \delta_{2k-r+1}^L > 0$. Second, we need to calculate the condition for the number of sensor elements for the SNR condition (D.7). Since $\gamma = \lim_{n\to\infty} \sqrt{k/m}$, we have (D.7) provided that

$$m \geq k(1+\delta) \left[1 - \frac{4(\kappa(B)+1)}{\mathsf{SNR}_{\min}(Y)}\right]^{-1}.$$

Therefore, if we have $\mathsf{SNR}_{\min}(Y) > 1 + 4(\kappa(B)+1)$ and

$$m \geq \max\left\{ \frac{k(1+\delta)}{1 - \frac{4(\kappa(B)+1)}{\mathsf{SNR}_{\min}(Y)-1}}, (1+\delta)(2k-r+1) \right\},$$

then we can identify the remaining r indices of $\text{supp}X$.

Appendix E

The following two lemmas are quite often used in this paper.

Lemma E.1: 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

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

in probability.

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

$$P\{Z_r > x\} = \frac{\Gamma(r/2, x/2)}{\Gamma(r/2)}, \quad \text{(E.1)}$$

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

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

For $n \to \infty$, 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

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

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

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

as $n \to \infty$ so that the claim is proved. ■

Lemma E.2: Let $A \in \mathbb{R}^{m\times n}$ be the Gaussian sensing matrix whose components $a_{i,j}$ are independent random variable with distribution $\mathcal{N}(0, 1/m)$. Then

$$\lim_{n\to\infty} \frac{\|AX\|_F^2}{\|X\|_F^2} = 1.$$

Proof: Because $a_{i,j} \sim \mathcal{N}(0, 1/m)$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$, $m\|\mathbf{a}_j\|^2$ is a chi-squared random variable of degree of freedom m so that by Lemma 3 in [45], we have

$$\lim_{n\to\infty} \max_{1\leq j\leq n} \|\mathbf{a}_j\|^2 = \lim_{n\to\infty} \min_{1\leq j\leq n} \|\mathbf{a}_j\|^2 = 1. \tag{E.2}$$

Since, for fixed $1 \leq j \leq n$, $\mathbf{a}_i (i \neq j)$ is m-dimensional random vector that is nonzero with a probability of 1 and independent of $\mathbf{a}_j$, the random variable $u(i,j) = \mathbf{a}_i^* \mathbf{a}_j / \|\mathbf{a}_j\|$ is a Gaussian random variable with a variance of $1/m$ by applying Lemma 2 in [45]. Since we have (E.2) and the variance of $u(i,j)$ goes to 0 as $n \to \infty$,

$$\lim_{n\to\infty} \mathbf{a}_i^* \mathbf{a}_j = 0 \tag{E.3}$$

for all $1 \leq j < k \leq n$. Then

$$\frac{\|AX\|_F^2}{\|X\|_F^2} = \frac{\text{trace}(X^*A^*AX)}{\text{trace}(X^*X)} \to 1$$

as $n \to \infty$. ∎

Appendix F: Proof of Theorem 5.2

Proof of Theorem 5.2: Let $I_t \subset \text{supp}X$ with $|I_t| = k - r$, where I_t is constructed by the first $k-r$ indices of X if we are ordering the values of $\|\mathbf{x}^i\|^2$ for $1 \leq i \leq n$ with decreasing order. Then for $i \in I_t$,

$$\mathbf{a}_i^* B = \|\mathbf{a}_i\|^2 \mathbf{x}^i + \mathbf{a}_i^* E^i$$

where $E_i = [\mathbf{e}_1^i, \cdots, \mathbf{e}_r^i] \in \mathbb{R}^{n\times r}$ and $\mathbf{e}_l^i = \mathbf{b}_l - \mathbf{a}_i x_{i,l}$. Then

$$\begin{aligned} \mathbf{a}_i^* B B^* \mathbf{a}_i &= \| \|\mathbf{a}_i\|^2 \mathbf{x}^i + \mathbf{a}_i^* E^i \|^2 \\ &\geq | \|\mathbf{a}_i\|^2 \|\mathbf{x}^i\| - \|\mathbf{a}_i^* E^i\| |^2 \\ &= \left| \sqrt{A_i} - \sqrt{\sum_{l=1}^r B_l^i Z_l^i} \right|^2 \end{aligned} \tag{F.1}$$

where

$$A_i = \|\mathbf{a}_i\|^4 \|\mathbf{x}^i\|^2, \; B_l^i = \|\mathbf{a}_i\|^2 \|\mathbf{e}_l^i\|^2, \; Z_l^i = \frac{|\mathbf{a}_i^* \mathbf{e}_l^i|^2}{\|\mathbf{a}_i\|^2 \|\mathbf{e}_l^i\|^2}.$$

First, by Lemma 3 in [45], $\lim_{n\to\infty} \sup_{i\in I_t} \|\mathbf{a}_i\|^2 = 1$ so that we have

$$\liminf_{n\to\infty} \frac{A_i}{r \mathsf{MSR}_{\min}^{k-r}} = \liminf_{n\to\infty} \frac{\|\mathbf{a}_i\|^4}{\mathsf{MSR}_{\min}^{k-r}} \frac{\|\mathbf{x}^i\|^2}{r} \geq 1 \tag{F.2}$$

by the definition of $\mathsf{MSR}_{\min}^{k-r}$ and the construction of I_t.

For B_l^i, observe that each $\mathbf{e}_l^i$ is a Gaussian m-dimensional vector with total variance

$$V_l^i := E[\|\mathbf{e}_l^i\|^2] \leq E[\|\mathbf{b}_l\|^2] = \|\mathbf{x}_l\|^2$$

and $(m/V_l^i)\|\mathbf{e}_l^i\|^2$ is a chi-squared distribution with a degree of freedom m for $i \in I_t$. Hence using Lemma 3 in [45] and

$$\frac{\log(k-r)}{m} \leq \frac{\log m}{m} \longrightarrow 0 \tag{F.3}$$

as $n \to \infty$ so that

$$\limsup_{n\to\infty} \max_{i\in I_t} \frac{\|\mathbf{a}_i\|^2 \|\mathbf{e}_l^i\|^2}{\|\mathbf{x}_l\|^2} \leq \limsup_{n\to\infty} \max_{i\in I_t} \frac{\|\mathbf{a}_i\|^2 V_i^l}{\|\mathbf{x}_l\|^2} \leq 1$$

so that we have

$$\limsup_{n\to\infty} \max_{i\in I_t} \frac{B_l^i}{\|\mathbf{x}_l\|^2} \leq 1$$

for $i \in I_t$ and $1 \leq l \leq r$. Finally,

$$Z_l^i = \frac{|\mathbf{a}_i^* \mathbf{e}_l^i|^2}{\|\mathbf{a}_i\|^2 \|\mathbf{e}_l^i\|^2}$$

follows beta distribution $\text{Beta}(1, m-1)$ as shown in [45]. Since there are $k - r$ terms in I_t, Lemma 6 in [45] and inequality (F.3) shows that

$$\limsup_{n\to\infty} \max_{i\in I_t} \frac{m}{2\log(k-r)} Z_l^i \leq 1$$

so that we have

$$\begin{aligned} &\limsup_{n\to\infty} \max_{i\in I_t} \frac{m}{2r\log(k-r)} \frac{\sum_{l=1}^r B_l^i Z_l^i}{\|X\|_F^2 / r} \\ \leq \; &\sum_{l=1}^r \limsup_{n\to\infty} \max_{i\in I_t} \frac{\|\mathbf{x}_l\|^2}{\|X\|_F^2} \frac{m}{2\log(k-r)} Z_l^i \leq 1. \end{aligned} \tag{F.4}$$

For $i \notin \text{supp}X$, we have

$$\begin{aligned} &\|\mathbf{a}_i^* B\|^2 \\ = \; &\mathbf{a}_i^* B B^* \mathbf{a}_i = \sum_{l=1}^r \sigma_l^2(B) \|\mathbf{a}_j^* \mathbf{u}_l\|^2 \\ = \; &\sigma_{\min}^2(B) \sum_{l=1}^r \|\mathbf{a}_i^* \mathbf{u}_l\|^2 + \sum_{l=1}^r (\sigma_l^2(B) - \sigma_{\min}^2(B)) \|\mathbf{a}_i^* \mathbf{u}_l\|^2 \end{aligned}$$

where $B = U\Sigma V$ is the singular value decompostion of B, $U = [\mathbf{u}_1, \cdots, \mathbf{u}_l]$ and $\Sigma = \text{diag}[\sigma_1(B), \cdots, \sigma_r(B)]$ where $\sigma_1(B) \geq \cdots \geq \sigma_r(B) = \sigma_{\min}(B) > 0$. As will be shown later, the decomposition in the second line of the above equation is necessary to deal with different asymptotic behavior of chi-square random variable of degree of freedom 1 and r. Since $\mathbf{a}_i$ is statistically independent from $\{\mathbf{u}_l\}_{l=1}^r$ for $i \notin \text{supp}X$ and $\{\mathbf{u}_l\}_{l=1}^r$ is an orthonormal set, $\sum_{l=1}^r m\|\mathbf{a}_j^* \mathbf{u}_l\|^2$ is a chi-squared random variable of degree of freedom r and each $m\|\mathbf{a}_j^* \mathbf{u}_l\|$ is a chi-squared random variable of degree of freedom 1. Also, we have

$$\begin{aligned} &\limsup_{n\to\infty} \frac{\sum_{l=1}^r (\sigma_l^2(B) - \sigma_{\min}^2(B)) m \|\mathbf{a}_i^* \mathbf{u}_l\|^2}{2r\log((n-k)r)} \\ \leq \; &\limsup_{n\to\infty} \left(\frac{\|B\|_F^2}{r} - \sigma_{\min}^2(B) \right) \end{aligned} \tag{F.5}$$

since

$$\limsup_{n\to\infty} \max_{i\notin \text{supp}X, 1\leq l\leq r} \frac{m\|\mathbf{a}_j \mathbf{u}_l\|^2}{2\log((n-k)r)} \leq 1$$

by Lemma 4 in [45]. When r is a fixed number, then by Lemma E.1, we have

$$\lim_{n\to\infty}\max_{j=1,\cdots,n}\frac{\sum_{l=1}^{r} m\|\mathbf{a}_j^*\mathbf{u}_l\|^2}{2\log(n-k)}=1. \tag{F.6}$$

On the other hand, when r is proportionally increasing with respect to k [45], then we have

$$\lim_{n\to\infty}\max_{j=1,\cdots,n}\frac{\sum_{l=1}^{r} m\|\mathbf{a}_j^*\mathbf{u}_l\|^2}{r}=1. \tag{F.7}$$

Combining (F.6), (F.7) and (F.5), we have

$$\limsup_{n\to\infty}\frac{m\|\mathbf{a}_i^*B\|^2}{2B(n,k,r)}\leq 1 \tag{F.8}$$

for $j\notin \mathrm{supp}X$, when $B(n,k,r)$ is given by (V.5).

For the noisy measurement Y, we have for all $1\leq j\leq n$,

$$\begin{aligned}&\limsup_{n\to\infty}\frac{|\|\mathbf{a}_j^*B\|^2-\|\mathbf{a}_j^*Y\|^2|}{(2\|B\|+\|N\|)\|N\|}\\ \leq\ &\limsup_{n\to\infty}\|\mathbf{a}_j\|=1.\end{aligned} \tag{F.9}$$

Let

$$\lambda:=\mathsf{MSR}_{\min}^{k-r},\qquad \mu:=\frac{\|X\|_F^2}{r}2\log(k-r)$$
$$\nu:=\frac{2(2\|B\|+\|N\|)\|N\|}{r},\qquad \xi=\frac{2B(n,k,r)}{r}.$$

Then, for $i\in\mathrm{supp}X$, combining (F.1), (F.2), (F.4) and (F.9), we have

$$\begin{aligned}&\liminf_{n\to\infty}\frac{m\|\mathbf{a}_i^*Y\|^2-m(2\|B\|+\|N\|)\|N\|}{r\xi}\\ \geq\ &\liminf_{n\to\infty}\frac{m\|\mathbf{a}_i^*B\|-2m(2\|B\|+\|N\|)\|N\|}{r\xi}\\ \geq\ &\liminf_{n\to\infty}\frac{\left([\sqrt{\lambda}\sqrt{m}-\sqrt{\mu}]^2-\nu m\right)}{\xi}.\end{aligned}$$

On the other hand, using (F.8) and (F.9), for $j\notin\mathrm{supp}X$ we have

$$\begin{aligned}&\limsup_{n\to\infty}\frac{m\|\mathbf{a}_i^*Y\|^2-m(2\|B\|+\|N\|)\|N\|}{r\xi}\\ \leq\ &\limsup_{n\to\infty}\frac{m\|\mathbf{a}_i^*B\|^2}{r\xi}\leq 1\end{aligned}$$

so that we need to show that

$$\liminf_{n\to\infty}\frac{[\sqrt{\lambda}\sqrt{m}-\sqrt{\mu}]^2-\nu m}{\xi}\geq 1+\delta \tag{F.10}$$

under the condition (V.3) and (V.4). First, note that $\lambda>\nu$ if and only if

$$r\mathsf{MSR}_{\min}^{k-r}>2(2\|B\|+\|N\|)\|N\|.$$

which is equivalent to that

$$\frac{r\mathsf{MSR}_{\min}^{k-r}}{\sigma_{\min}^2(B)}\mathsf{SNR}_{\min}^2(Y)-4\kappa(B)\mathsf{SNR}_{\min}(Y)-2>0$$

which holds under the condition (V.3), where we used the definition $\kappa(B):=\|B\|/\sigma_{\min}(B)$. Then we can see that if we have $\sqrt{m}\geq\frac{\sqrt{\mu}}{\sqrt{\lambda}-\sqrt{\nu}}$, then

$$(\sqrt{\lambda}\sqrt{m}-\sqrt{\mu})^2-\nu m\geq[(\sqrt{\lambda}-\sqrt{\nu})\sqrt{m}-\sqrt{\mu}]^2. \tag{F.11}$$

Also, if we have

$$\sqrt{m}\geq\frac{\sqrt{\mu}+\sqrt{1+\delta}\sqrt{\xi}}{\sqrt{\lambda}-\sqrt{\nu}},$$

then

$$\frac{[\sqrt{\lambda}\sqrt{m}-\sqrt{\mu}]^2-\nu m}{\xi}\geq 1+\delta. \tag{F.12}$$

Hence, by applying (F.11) and (F.12), if we assume the condition

$$\sqrt{m}\geq\sqrt{1+\delta}\frac{\sqrt{\mu}+\sqrt{\xi}}{\sqrt{\lambda}-\sqrt{\nu}}\geq\frac{\sqrt{\mu}+\sqrt{1+\delta}\sqrt{\xi}}{\sqrt{\lambda}-\sqrt{\nu}},$$

then the inequality (F.10) holds so that we can identify $I_t\subset\mathrm{supp}X$ by 2-thresholding.

Appendix G: Proof of Theorem 5.3 and 5.4

In this section, we assume the large system limit such that ρ, ϵ, α and γ exist. We first need to have the following results.

Theorem G.1: [53] 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

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

where $\gamma=\lim_{n\to\infty}\sqrt{k/m}$.

Corollary G.2: 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 singular value of A is given by

$$ds_\gamma(x):=\frac{1}{\pi\gamma^2}\frac{\sqrt{((1+\gamma)^2-x^2)(x^2-(1-\gamma)^2)}}{x}. \tag{G.2}$$

Proof: This is obtained from Theorem G.1 using a simple change of variable. ■

Lemma G.3: Let $r\leq k<m$ be positive integers and $A\in\mathbb{R}^{m\times k}$. Then for any r-dimensional subspace W of $R(A)$, we have

$$\|A^*P_W\|_F^2\geq\sum_{j=1}^{r}\sigma_{k-j+1}^2(A)$$

where $\sigma_1(A)\geq\sigma_2(A)\geq\cdots\geq\sigma_k(A)\geq 0$.

Proof: Let $A^* = \tilde{U}\tilde{\Sigma}\tilde{V}^*$ be the extended singular value decomposition of A^* where

$$\begin{aligned}\tilde{\Sigma} &= \operatorname{diag}[\sigma_1, \sigma_2, \cdots, \sigma_m],\\ \tilde{V} &= [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m]\end{aligned}$$

and $\sigma_{k+1} = \sigma_{k+2} = \cdots = \sigma_m = 0$. If we let $Z = \tilde{V}^* P_W$, then we have

$$\|Z\|_F^2 = \operatorname{trace}(P_W \tilde{V}\tilde{V}^* P_W) = \operatorname{trace}(P_W) = r \tag{G.3}$$

and

$$\|A^* P_W\|_F^2 = \|A^* \tilde{V} Z\|_F^2 = \|\tilde{U}\tilde{\Sigma} Z\|_F^2$$

If we let $Z = [\mathbf{z}_1^*, \cdots, \mathbf{z}_m^*]^*$, since W is a subspace of $R(A)$ and $R(A) = N(A^*)^\perp$, we have

$$\mathbf{z}_{k+1} = \mathbf{z}_{k+2} = \cdots = \mathbf{z}_m = \mathbf{0}$$

and

$$\sum_{l=1}^{k} \|\mathbf{z}_l\|^2 = r.$$

by (G.3). Since $0 \le \|\mathbf{z}_l\|^2 \le 1$ for $1 \le j \le k$, using $\sigma_1(A) \ge \sigma_2(A) \ge \cdots \ge \sigma_k(A)$, we have

$$\|A^* P_W\|_F^2 = \|\tilde{U}\tilde{\Sigma} Z\|_F^2 = \sum_{l=1}^{k} \sigma_l^2(A)\|\mathbf{z}_l\|^2 \ge \sum_{j=1}^{r} \sigma_{k-j+1}^2(A).$$

■

Lemma G.4: For $0 \le \gamma \le 1$ and $0 \le \alpha \le 1$, we let $0 \le t_\gamma(\alpha) \le 1$ 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

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

For the proof of Lemma G.4, we need the following lemma.

Lemma G.5: Let $-\infty < a < b < \infty$. Suppose that $f_1(x)$ and $f(x)$ are continouous probability density functions on $[a, b]$ such that for any $t \in [a, b]$,

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

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

$$\int_a^{q_1} f_1(x)dx = \int_a^q f(x)dx, \tag{G.4}$$

we have

$$\int_a^{q_1} g(x)f_1(x)dx \ge \int_a^q g(x)f(x)dx. \tag{G.5}$$

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 $F_1^{-1}(x) \ge F^{-1}(x)$ for any $x \in [0, 1]$. For any $(q_1, q) \in [a, b] \times [a, b]$ which satisfies (G.4), there is some $c \in [0, 1]$ such that $F_1(q_1) = F(q) = c$. Applying the change of variable, we have

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

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

Proof of Lemma G.4 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 G.5, we only need to show that

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

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

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

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.

Proof of Theorem 5.3 and 5.4: Note that S-OMP can find $k-r$ correct indices from $\mathrm{supp}X$ if we have

$$\max_{j\in \mathrm{supp}X} \|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2 > \max_{j\notin \mathrm{supp}X} \|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2 \tag{G.6}$$

for each $0 \leq t < k-r$, since $\|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2 = 0$ for $j \in \mathrm{supp}X \cap I_t$. Hence, it is enough to check that the condition (G.6) for $0 \leq t < k-r$.

First, for $j \notin \mathrm{supp}X$, since $\mathbf{a}_j$ is statistically independent of $P^\perp_{R(A_{I_t})}Y$. For $t \leq k-r$, the dimension of $P^\perp_{R(A_{I_t})}Y$ is r so that $m\|\mathbf{a}_j P_{R(P^\perp_{R(A_{I_t})}Y)}\|^2$ is of chi-squared distribution of degree of freedom r.

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

$$\begin{aligned}\max_{j\in \mathrm{supp}X} \|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2 &\geq \frac{1}{k}\|A_S^* P_{R(P^\perp_{R(A_{I_t})}B)}\|_F^2 \\ &\geq \frac{\sum_{j=1}^r \sigma^2_{k-j+1}(A_S)}{k}\end{aligned}$$

by using $R(P^\perp_{R(A_{I_t})}B) \subset R(A_S)$ and Lemma G.3, where A_S have singular values $0 < \sigma_k(A_S) \leq \sigma_{k-1}(A_S) \leq \cdots \leq \sigma_1(A_S)$. If we let

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

then by (G.1), we have

$$\lim_{n\to\infty} \frac{\sum_{j=1}^r \sigma^2_{k-j+1}(A_S)}{k} = \int_{(1-\gamma)^2}^{(1-\gamma+2\gamma t_\gamma(\alpha))^2} x d\lambda_\gamma(x) \tag{G.7}$$

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

$$\int_{1-\gamma}^{1-\gamma+2\gamma t_\gamma(\alpha)} ds_\gamma(x) = \alpha = \lim_{n\to\infty}\frac{r}{k}.$$

If we let

$$d\lambda_{0,\gamma}(x) = \frac{1}{\pi}\frac{\sqrt{4-s}\sqrt{s}}{\gamma s + (1-\gamma)^2}dx, \tag{G.8}$$

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

$$\int_0^t d\lambda_1(x) \leq \int_0^t d\lambda_{0,\gamma}(x) \tag{G.9}$$

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

$$\begin{aligned}&\int_{(1-\gamma)^2}^{(1-\gamma+2\gamma t_\gamma(\alpha))^2} x d\lambda_\gamma(x) \\ =\ &\int_0^{\frac{(1-\gamma+2\gamma t_\gamma(\alpha))^2-(1-\gamma)^2}{r}} [(1-\gamma)^2+\gamma s] d\lambda_{0,\gamma}(s) \\ \geq\ &\int_0^{4t_1(\alpha)^2} [(1-\gamma)^2+\gamma s] d\lambda_1(s) \\ =\ &(1-\gamma)^2\alpha + \gamma\int_0^{4t_1(\alpha)^2} s d\lambda_1(s),\end{aligned} \tag{G.10}$$

where the inequality comes from Lemma G.4. Substituting (G.10) into (G.7), we have

$$\begin{aligned}&\liminf_{n\to\infty} \max_{j\in \mathrm{supp}X} \|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2 \\ \geq\ &\liminf_{n\to\infty} \frac{\sum_{j=1}^r \sigma^2_{k-j+1}(A_S)}{k} \\ \geq\ &\alpha\left[(1-\gamma)^2 + \gamma\frac{\int_0^{4t_1(\alpha)^2} s d\lambda_1(s)}{\alpha}\right] \\ =\ &\lim_{n\to\infty}\frac{r}{m}(1/\gamma-1)^2 + \alpha\gamma F(\alpha)\end{aligned} \tag{G.11}$$

where $F(\alpha) := (1/\alpha)\int_0^{4t_1(\alpha)^2} s d\lambda_1(s)$ is an increasing function with respect to α such that $\lim_{\alpha\to 0} F(\alpha) = 0$ and $F(1) = 1$, and $\alpha\gamma^2 = (\lim_{n\to\infty} r/k)(\lim_{n\to\infty} k/m) = \lim_{n\to\infty} r/m$.

For noisy measurement Y, we have the following inequality:

$$\begin{aligned}&\left|\|P_{R(P^\perp_{R(A_{I_t})}Y)}\mathbf{a}_j\|^2 - \|P_{R(P^\perp_{R(A_{I_t})}B)}\mathbf{a}_j\|^2\right| \\ \leq\ &(\|P_{R(P^\perp_{R(A_{I_t})}Y)}\mathbf{a}_j\| + \|P_{R(P^\perp_{R(A_{I_t})}B)}\mathbf{a}_j\|) \\ &\times \left|\|P_{R(P^\perp_{R(A_{I_t})}Y)}\mathbf{a}_j\| - \|P_{R(P^\perp_{R(A_{I_t})}B)}\mathbf{a}_j\|\right| \\ \leq\ &2\|\mathbf{a}_j\|\|P_{R(P^\perp_{R(A_{I_t})}Y)}\mathbf{a}_j - P_{R(P^\perp_{R(A_{I_t})}B)}\mathbf{a}_j\| \\ \leq\ &2\|\mathbf{a}_j\|\|P_{R(Y)}\mathbf{a}_j - P_{R(B)}\mathbf{a}_j\| \\ \leq\ &2\|\mathbf{a}_j\|^2\|P_{R(Y)} - P_{R(B)}\| \\ &\to 2\|P_{R(Y)} - P_{R(B)}\| \leq \frac{4(\sigma_{\max}(B)+\sigma_{\min}(B))\|N\|}{\sigma_{\min}(B)(\sigma_{\min}(B)-\|N\|)} \\ &= \frac{4(\kappa(B)+1)}{\mathsf{SNR}_{\min}(B)-1}\end{aligned} \tag{G.12}$$

as $n \to \infty$, where $\mathsf{SNR}_{\min}(B) = \sigma_{\min}(B)/\|N\|$.

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

(Case 1 : Theorem 5.3) For $t \leq k-r$, $\{m\|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2 : j \notin \mathrm{supp}X\}$ are independent chi-squared random variables of degree of freedom r so that by Lemma E.1, we have

$$\lim_{n\to\infty}\max_{j\notin \mathrm{supp}X} \frac{m\|\mathbf{a}_j^* P_{R(P^\perp_{R(A_{I_t})}B)}\|^2}{2\log(n-k)} = 1. \tag{G.13}$$

Here we assume that

$$\mathsf{SNR}_{\min}(Y) > 1 + 4\frac{k}{r}(\kappa(B)+1) \tag{G.14}$$

and

$$m > k\left[1 - \frac{4k}{r}\frac{\kappa(B)+1}{\mathsf{SNR}_{\min}-1}\right]^{-1} 2(1+\delta)\frac{\log(n-k)}{r}. \tag{G.15}$$

Then by Marćenko-Pastur theorem [53],

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

so that

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

Combining (G.12) and (G.16), for noisy measurement Y, we have

$$\begin{aligned}&\liminf_{n\to\infty}\max_{j\in\mathrm{supp}X}\frac{m\|\mathbf{a}_j^* P_{R(P^{\perp}_{R(A_{I_t})}Y)}\|^2}{2\log(n-k)} \\ \geq\ &\liminf_{n\to\infty}\frac{r}{2\log(n-k)}\left[1-\frac{4k}{r}\frac{\kappa(B)+1}{\mathsf{SNR}_{\min}(Y)-1}\right]\frac{1}{\gamma^2} \\ \geq\ &1+\delta\end{aligned}$$

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

(Case 2: Theorem 5.4) Similarly as in the previous case, for $t < k-r$, $\{m\|\mathbf{a}_j^* P_{R(P^{\perp}_{R(A_{I_t})}B)}\|^2 : j \notin \mathrm{supp}X\}$ are independent chi-squared distribution. Since $\lim_{n\to\infty}(\log n)/r = 0$, by Lemma 3 in [45], we have

$$\lim_{n\to\infty}\max_{j\notin\mathrm{supp}X}\frac{m\|\mathbf{a}_j^* P_{R(P^{\perp}_{R(A_{I_t})}B)}\|^2}{r} = 1. \tag{G.17}$$

By using (G.12), we have

$$\begin{aligned}&\liminf_{n\to\infty}\max_{j\in\mathrm{supp}X}\frac{m\|\mathbf{a}_j^* P_{R(P^{\perp}_{R(A_{I_t})}Y)}\|^2}{r} \\ \geq&\left(\frac{1}{\gamma}-1\right)^2 + F(\alpha)\frac{1}{\gamma} - \frac{4}{\alpha}\frac{\kappa(B)+1}{\mathsf{SNR}_{\min}(B)-1}\frac{1}{\gamma^2}.\end{aligned} \tag{G.18}$$

We let

$$\mathsf{SNR}_{\min}(B) > 1 + \frac{4}{\alpha}(\kappa(B)+1) \tag{G.19}$$

and

$$m > k(1+\delta)^2\frac{1}{\left(1-\frac{4}{\alpha}\frac{(\kappa(B)+1)}{\mathsf{SNR}_{\min}-1}\right)^2}\left[2-F(\alpha)\right]^2 \tag{G.20}$$

for some $\delta > 0$. Note that (G.20) is equivalent to

$$\frac{1}{\gamma} > (1+\delta)\frac{1}{1-\frac{4}{\alpha}\frac{(\kappa(B)+1)}{\mathsf{SNR}_{\min}-1}}\left[2-F(\alpha)\right]$$

Again we let

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

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

$$\begin{aligned}Q(x) &= (1-v)x^2 - (2-u)x + 1 \\ &= (1-v)x\left[x - \frac{2-u}{(1-v)}\right] + 1 \\ &> \delta(1+\delta)\frac{(2-u)^2}{1-v} + 1 \\ &\geq 1 + \delta(1+\delta)\end{aligned} \tag{G.21}$$

since $1-v > 0$ by (G.19) and $0 \leq u \leq 1$. Combining (G.18) and (G.21), we have for $0 \leq t < k-r$ and $j \in \mathrm{supp}X$, we have

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

for some $\delta > 0$. Hence, in the case of $\lim_{n\to\infty} r/k = \alpha > 0$, we can identify the correct indices of $\mathrm{supp}X$ if we have (G.20).

Acknowledgments

This research was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded in part by the Korean government (MEST) (No. 2009-0081089) and in part by the Industrial Strategic Technology Program of the Ministry of Knowledge Economy (KI001889). The authors would like to thank Dr. Dmitry Malioutov for providing the $l_{1,2}$ mixed norm code in [26] and anonymous reviewers for their helpful comments that have improved this paper significantly.

References

[1] D.L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] H. Jung, J.C. Ye, and E.Y. Kim, "Improved k–t BLAST and k–t SENSE using FOCUSS," *Physics in Medicine and Biology*, vol. 52, pp. 3201–3226, 2007.

[5] J.C. Ye, S. Tak, Y. Han, and H.W. Park, "Projection reconstruction MR imaging using FOCUSS," *Magnetic Resonance in Medicine*, vol. 57, no. 4, pp. 764–775, 2007.

[6] H. Jung, K. Sung, K.S. Nayak, E.Y. Kim, and J.C. Ye, "k-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI," *Magnetic Resonance in Medicine*, vol. 61, no. 1, pp. 103–116, 2009.

[7] G.H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical Physics*, vol. 35, pp. 660, 2008.

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

[9] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 47, no. 10, pp. 44–51, 2008.

[10] J.A. Tropp, "Just Relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. on Information Theory*, vol. 52, no. 3, pp. 1030–1051, March 2006.

[11] I.F. Gorodnitsky and B.D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: Re-weighted minimum norm algorithm," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[12] E.J. Candès, M.B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific computing*, vol. 20, no. 1, pp. 33–61, 1999.

[14] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[15] D.P. Wipf and B.D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[16] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple measurement vectors," *IEEE Trans. on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

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

[18] M. Mishali and Y.C. Eldar, "Reduce and boost: recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. on Signal Processing*, vol. 56(10), pp. 4692–4702, 2008.

[19] E. Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. on Information Theory*, vol. 56(5), pp. 2516–2527, 2010.

[20] M.F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, and R.G. Baraniuk, "Distributed compressed sensing of jointly sparse singals," in *Asilomar Conf. Signals, Sys., Comput*, 2005, pp. 1537–1541.

[21] H. Krim and M Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[22] K.P. Pruessmann, M. Weiger, M.B. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[23] A. Joshi, W. Bangerth, K. Hwang, J.C. Rasmussen, and E.M. Sevick-Muraca, "Fully adaptive FEM based fluorescence optical tomography from time-dependent measurements with area illumination and detection," *Medical physics*, vol. 33, pp. 1299, 2006.

[24] O. Lee, J. M. Kim, Y. Bresler, and J. C. Ye, "Compressive diffuse optical tomography: non-iterative exact reconstruction using joint sparsity," *IEEE Trans. on Medical Imaging*, vol. 30, no. 5, pp. 1129–1142, 2011.

[25] J.A. Tropp, A.C. Gilbert, and M.J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[26] D. Malioutov, M. Cetin, and A.S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[27] J.A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[28] D.P. Wipf, *Bayesian methods for finding sparse representations*, Ph.D. thesis, University of California, San Diego, 2006.

[29] Y.C. Eldar, P. Kuppinger, and H. Bolcskei, "Compressed sensing of block-sparse signals: uncertainty relations and efficient recovery," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[30] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[31] P. Feng, *Universal minimum-rate sampling and spectrum-blind reconstruction for multiband signals*, Dissertation, University of Illinois, Urbana-Champaign, 1997.

[32] M.E. Davies and Y.C. Eldar, "Rank awareness for joint sparse recovery," *arXiv:1004.4529*, 2010.

[33] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[34] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[35] D.L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. National Academy of Sciences*, vol. 107(25), pp. 9446–9451, 2005.

[36] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 655–687, 2008.

[37] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[38] Y.C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. on Information Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[39] F.R. Bach, "Consistency of the group Lasso and multiple kernel learning," *The Journal of Machine Learning Research*, vol. 9, pp. 1179–1225, 2008.

[40] D.P. Wipf and B.D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. on Signal Processing*, vol. 55, no. 7 Part 2, pp. 3704–3716, 2007.

[41] B. Porat and B. Friedlander, "Direction finding algorithms based on high-order statistics," *IEEE Trans. on Signal Processing*, vol. 39, no. 9, pp. 2016–2024, 1991.

[42] J.F. Cardoso, "Source separation using higher order moments," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1989, pp. 2109–2112.

[43] E. Gonen, J.M. Mendel, and M.C. Dogan, "Applications of cumulants to array processing. IV. Direction finding in coherent signals case," *IEEE Trans. on Signal Processing*, vol. 45, no. 9, pp. 2265–2276, 1997.

[44] J.F. Cardoso and E. Moulines, "Asymptotic performance analysis of direction-finding algorithms based on fourth-order cumulants," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 214–224, 1995.

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

[46] A.K. Fletcher and S. Rangan, "Orthogonal matching pursuit from noisy random measurements: A new analysis," in *Proc. conf. Neural Information Processing Systems*, 2009.

[47] K. Lee and Y. Bresler, "Subspace-augmented MUSIC for joint sparse recovery," *arXiv:1004.4371*, 2010.

[48] J. M. Kim, O. Lee, and J. C. Ye, "Compressive MUSIC with optimized partial support for joint sparse recovery," *arXiv:1102.3288*, 2011.

[49] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[50] M. Duarte and R. Baraniuk, "Spectral compressive sensing," *Submitted for publication*, 2010.

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

[52] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT), Seoul, Korea*, 2009.

[53] V.A. Marčenko and L.A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Sbornik: Mathematics*, vol. 1, no. 4, pp. 457–483, 1967.

PLACE
PHOTO
HERE

Jong Min Kim Jong Min Kim received the B.S. degree in mathematics from KAIST, Daejeon, Republic of Korea. He also received the M.S. degree in applied mathematics in 2001, as well as Ph.D. degree in mathematical sciences in 2007, both from KAIST. He is currently a research professor at Dept. of Bio and Brain Engineering, KAIST. His research interests include compressive sensing, sampling theory, harmonic analysis and signal processing.

PLACE
PHOTO
HERE

Ok Kyum Lee Ok Kyun Lee received the B.S. degree in electrical, electronic and computer engineering in 2007 from the Hanyang University. He received the M.S. degree in eletronic and eletrical engineering in 2009 from the Korea Advanced Institute of Science and Technology (KAIST), where he is currently pursuing the Ph.D. degree. His research interrests include compressive sensing, neuroimaging, and diffuse optical tomography.

PLACE
PHOTO
HERE

Jong Chul Ye Jong Chul Ye received the B.Sc. and M.Sc. degrees with honors from Seoul National University, Korea, in 1993 and 1995, respectively, and the Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, in 1999. Before he joined KAIST in 2004, he worked as was a postdoc fellow at University of Illinois at Urbana-Champaign, and senior member research staff at Philips Research, NY, and GE Global Research Center, NY. He is currently an Associate Professor at Dept. of Bio and Brain Engineering of KAIST, and also an Adjunct Professor at Dept. of Electrical Engineering of KAIST. His current research interests include compressed sensing and statistical signal processing for various imaging modalities such as MRI, NIRS, CT, PET, optics, etc.

Dr. Ye received Korean Government Overseas Scholarship Award for his doctoral study. He served as a technical committee member and organizing committee in various conferences such as IEEE Int. Conf. on Image Processing (ICIP), IEEE Symp. Biomedical Imaging (ISBI), and etc.