Noise Robust Joint Sparse Recovery using Compressive Subspace Fitting

Jong Min Kim, Ok Kyun Lee and Jong Chul Ye

Abstract

We study a multiple measurement vector (MMV) problem where multiple signals share a common sparse support set and are sampled by a common sensing matrix. Although we can expect that joint sparsity can improve the recovery performance over a single measurement vector (SMV) problem, compressive sensing (CS) algorithms for MMV exhibit performance saturation as the number of multiple signals increases. Recently, to overcome these drawbacks of CS approaches, hybrid algorithms that optimally combine CS with sensor array signal processing using a generalized MUSIC criterion have been proposed. While these hybrid algorithms are optimal for critically sampled cases, they are not efficient in exploiting the redundant sampling to improve noise robustness. Hence, in this work, we introduce a novel subspace fitting criterion that extends the generalized MUSIC criterion so that it exhibits near-optimal behaviors for various sampling conditions. In addition, the subspace fitting criterion leads to two alternative forms of compressive subspace fitting (CSF) algorithms with forward and backward support selection, which significantly improve the noise robustness. Numerical simulations show that the proposed algorithms can nearly achieve the optimum.

Index Terms

Compressed sensing, joint sparsity, multiple measurement vector, subspace fitting, compressive MUSIC, compressive subspace fitting

Correspondence to:
Jong Chul Ye, Ph.D.
Associate Professor
Dept. of Bio and Brain Engineering, KAIST
373-1 Guseong-dong Yuseong-gu, Daejon 305-701, Korea
Email: jong.ye@kaist.ac.kr
Tel: 82-42-350-4320
Fax: 82-42-350-4310
I. INTRODUCTION

We studied a multiple measurement vector (MMV) problem where multiple signals share the same common sparse support set, and each signal is measured by multiplying it by a measurement matrix. Specifically, let \( m, n \) and \( r \) be positive integers \((m < n)\) that represent the number of sensor elements, the ambient space dimension, and the number of snapshots, respectively. Suppose that we are given a multiple-measurement vector \( B \in \mathbb{R}^{m \times r}, \ X = [x_1, \cdots, x_r] \in \mathbb{R}^{n \times r}, \) and a sensing matrix \( A \in \mathbb{R}^{m \times n}. \) Then, an MMV problem is formulated by the following optimization problem [1]:

\[
\begin{align*}
\text{minimize} & \quad \|X\|_0 \\
\text{subject to} & \quad B = AX,
\end{align*}
\]

where \( \|X\|_0 = |\text{supp}X| = k, \ \text{supp}X = \{1 \leq i \leq n : x^i \neq 0\}. \) In Eq. (1), if the measurement matrix \( B \) is full rank, i.e. \( \text{rank}(B) = r \leq \|X\|_0, \) we call this problem a canonical MMV model [1]. Recall that every MMV problem can be converted to a canonical form MMV by using a singular value decomposition and dimension reduction as described in [1]. Hence, we assume that the MMV problem in this paper assumes the canonical form.

The MMV problem is one way in which multiple correlated signals can appear in a signal ensemble, and also has many important applications [2], [3], [4], [5], [6]. The central theme in these studies has been that joint sparsity within signal ensembles enables a further reduction in the number of measurements required [2], [7], where the number of measurements required per sensor must account for the minimal features unique to that sensor [1], [8], [9], [10], [11], [12]. We can easily expect that the diversity due to the joint sparsity can improve the recovery performance over single measurement vector (SMV) compressed sensing, i.e. \( r = 1. \) Indeed, Chen and Huo [8], Feng and Bresler [13] and recently Davies and Elder [14] showed that \( X \in \mathbb{R}^{n \times r} \) is the unique solution of \( AX = B \) if and only if

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

Note that we can expect \( \text{rank}(B)/2 \) gains over SMV thanks to the MMV diversity. Furthermore, Feng and Bresler [13] showed that the noiseless \( l_0 \) bound in Eq. (2) is achievable using the MUSIC algorithm as long as \( r = \text{rank}(B) = k. \) More specifically, suppose that the columns of a sensing
matrix \( A \in \mathbb{R}^{m \times n} \) are in general position. Then, according to [13], [15], for any \( j \in \{1, \cdots, n\} \), \( j \in \text{supp}X \) if and only if

\[
a_j^*Q^*a_j = 0,
\]

where \( a_j \) denotes the \( j \)-th column of \( A \), \( Q \in \mathbb{R}^{m \times (m-r)} \) consists of orthonormal columns such that \( Q^*B = 0 \) so that \( R(Q)^\perp = R(B) \), which is often called “noise subspace”. Using the compressive sensing terminology, Eq. (3) implies that the recoverable sparsity level by MUSIC (with a probability 1 for the noiseless measurement case) is given by

\[
\|X\|_0 < m = \text{spark}(A) - 1,
\]

where the last equality comes from the definition of the spark. Therefore, the \( l_0 \) bound (2) can be achieved by MUSIC bound in (4) when \( r = k \) [13].

However, for any \( r < k \), the MUSIC condition (3) does not hold. This is a major drawback of MUSIC compared to CS algorithms that allow perfect reconstruction with an extremely large probability by increasing the sensor elements \( m \). One the other hand, even though the conventional CS algorithms for MMV, such as simultaneous OMP (S-OMP), and \( p \)-thresholding [16], [17] have good recovery performance when \( r \ll k \), they exhibit performance saturation as \( r \) increases and never achieve the \( l_0 \) bound with finite snapshot, even in a noiseless case. Recently, Kim et al. [1] and Lee et al. [12] independently showed that this drawback of the existing approaches can be overcome by what they called the generalized MUSIC criterion [1], [12]. Using this, we developed a computationally tractable relaxation algorithm called compressive MUSIC (CS-MUSIC) that can be stated as following [1]:

- (Step 1: Compressed sensing step) Find \( k - r \) indices of \( \text{supp}X \) by any MMV compressive sensing algorithms, such as 2-thresholding or SOMP. Let \( I_{k-r} \) be a set of selected indices and \( S = I_{k-r} \);

- (Step 2: Generalized MUSIC step) For \( j \in \{1, \cdots, n\} \setminus I_{k-r} \), calculate the quantities \( \eta(j) = a_j^*[P_{R(Q)} - P_{R(P_{R(Q)}A_{I_{k-r}})}]a_j \) for all \( j \notin I_{k-r} \). Make an ascending ordering of \( \eta(j) \), \( j \notin I_{k-r} \) and choose indices that correspond to the first \( r \) elements and put these indices into \( S \).

This hybridization makes the compressive MUSIC applicable for all ranges of \( r \), outperforming all the existing methods. Similar observation was made by Lee et al. [12] in their subspace augmented
MUSIC (SA-MUSIC). Hence, due to the equivalence of the two algorithms, we use the term CS-MUSIC to represent both CS-MUSIC and SA-MUSIC.

For compressive MUSIC, due to the reduction of uncertainty from $|I_k|$ to $|I_{k-r}|$, we can expect a more relaxed sampling condition. More specifically, as shown in [1], the number of measurements for compressive MUSIC with subspace S-OMP for partial support recovery exhibits two distinct characteristics depending on the number of the measurement vectors. First, if $r$ is sufficiently small with respect to $k$, then the number of samples for S-OMP is reciprocally proportional to the number of multiple measurement vectors. On the other hand, if we have a sufficiently large number of snapshots such that \( \lim_{n \to \infty} \frac{r}{k} \) is nonzero, then the number of samples for S-OMP is bounded by \( 4k \) so that the log $n$ is not necessary. In particular, if the number of snapshots approaches $k$, then we can identify the support of $X$ with only $k$ measurements, which is equivalent to the required number of samples for the conventional MUSIC. Furthermore, when the asymptotic ratio of the number of snapshots and the sparsity level (that is, \( \lim_{n \to \infty} \frac{r}{k} \)) is nonzero in the large system limit, we can see that only finite SNR is required for the support recovery, which is a significant improvement over SMV-CS.

The performance improvement of CS-MUSIC or SA-MUSIC is quite significant in noiseless cases; however, as signal-to-noise (SNR) decreases, the performance improvement tends to be reduced. As shown in Fig 1(a)(b), even if the $k-r$ supports are known a priori by an “oracle”, the perfect recovery rate are significantly degraded as SNR decreases. Such performance degradation is especially severe for a small $r$. The results indicate that the MUSIC step is sensitive to noise. The noise sensitivity of the generalized MUSIC step has also been reported by Lee and Bresler [12] using restricted isometry property (RIP) analysis, which shows that the sensitivity is an inherent property of the generalized MUSIC step since the subspace spanned by the data matrix is smaller compared to the full rank MUSIC case. Therefore, we need a new hybrid algorithm that combines the compressed sensing and sensor array processing algorithm that is still robust to noise.

One of the main contributions of this paper is to show that such noise sensitive MUSIC step can be improved by introducing a new subspace-based criterion. More specifically, we show that the existing generalized MUSIC criterion is a special case of the new subspace criterion, and the new criterion can efficiently take advantage of the sampling redundancy to improve the noise robustness. More specifically, we extend the results in [1] to more general situations with various restricted
isometry property (RIP) conditions than those required in the original derivation in [1]. The new criterion, what we call the generalized subspace fitting criterion, can be reduced to the original form of the generalized MUSIC criterion under the RIP condition $0 \leq \delta_{k-r+1}^{L}(A) < 1$, and can encompass other general situations with more stringent RIP conditions. As a byproduct of the new criterion, we develop two alternative forms of compressive subspace fitting (CSF) algorithms with forward and backward support selection (forward and backward CSF), respectively, which are more robust to measurement noise, and hence, significantly improve the recovery performance. In particular, the forward CSF improves the noise robustness of MUSIC step by applying additional augmentation in the MUSIC step, whereas the backward CSF provides additional performance improvement by finding the optimal $k - r$ support as well. Using large system analysis, as well as numerical simulation, we show that the forward and backward CSFs are superior than existing CS-MUSIC or SA-MUSIC in all situations, and nearly achieve a $l_0$ bound for a sufficiently large $r$.

II. SUBSPACE FITTING CRITERION

A. Mathematical Preliminaries

Throughout the paper, $x_i$ and $x_j$ 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

Fig. 1. Recovery rates of compressive MUSIC when correct $k - r$ partial support are given by an “oracle” estimator. (a) $r = 3$, and (b) $r = 12$ where $m = 20$, $n = 100$. 
rows of $X$ and columns of $A$, respectively. The following definitions are also used throughout the paper.

**Definition 1**: The rows (or columns) in $\mathbb{R}^n$ are in general position if any $n$ collection of rows (or columns) are linearly independent.

**Definition 2**: $\text{Spark}(A)$ denotes the smallest number of linearly dependent columns of a matrix $A$.

**Definition 3 (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 $0 < \delta_k^L, \delta_k^R < 1$ such that
\[
(1 - \delta_k^L)\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta_k^R)\|x\|^2
\]
for all $x \in \mathbb{R}^n$ such that $\|x\|_0 \leq k$. A single RIP constant $\delta_k = \max\{\delta_k^L, \delta_k^R\}$ is often referred to as the RIP constant.

**Theorem 1**: (Generalized MUSIC criterion) Assume that $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, and $B \in \mathbb{R}^{m \times r}$ satisfy $AX = B$. Furthermore, we assume that $\|X\|_0 = k$ and $A$ satisfies the RIP condition with the left RIP constant $0 \leq \delta_{2k-r+1}^L(A) < 1$. Also, we assume that the nonzero rows of $X$ are in general position. If we are given $I_{k-r} \subset \text{supp}X$ with $|I_{k-r}| = k - r$ and $A_{I_{k-r}} \in \mathbb{R}^{m \times (k-r)}$, which consists of columns whose indices are in $I_{k-r}$, then for any $j \in \{1, \cdots, n\} \setminus I_{k-r}$,
\[
a_j^* \left[ P_{R(Q)} - P_{R(P_{n(Q)}A_{I_{k-r}})} \right] a_j = a_j^* P_{R(A_{I_{k-r}}B)} a_j = 0
\]
if and only if $j \in \text{supp}X$.

**B. Generalized Subspace Fitting Criterion**

Note that the generalized MUSIC criterion requires $0 \leq \delta_{2k-r+1}^L(A) < 1$. If the sensing matrix is random Gaussian and the measurement is noiseless, the RIP condition can be interpreted as the following sampling condition
\[
m \geq (1 + \delta)(2k - r + 1)
\]
for some $\delta > 0$.

Fig 2 illustrates the phase transition boundary of CS-MUSIC, when $k - r$ support are known a priori by an “oracle”. We can observe that if the number of samples $m$ is closed to the minimum sampling rate, CS-MUSIC nearly achieves the $l_0$-bound. However, as $m$ increases, we can observe that the resulting performance improvement grows slowly with different slopes. This phenomenon
is more severe when SNR becomes smaller. The results indicate that the noise sensitivity of MUSIC step is originated from that redundant sampling is not fully exploited. Indeed, such behavior is due to the use of the most relaxed form of RIP condition to derive the generalized MUSIC criterion

\[ 0 \leq \delta_{2k-r+1}^L(A) < 1. \]

With redundant sampling \( m \gg 2k - r + 1 \), RIP can be more stringent, and we need a new type of subspace criterion to deal with this situation. In the following theorem, we will show that the generalized MUSIC criterion in Theorem 1 that has been derived for the RIP condition \( 0 \leq \delta_{2k-r+1}^L(A) < 1 \), can be extended for the more general RIP conditions so that we can take advantage of redundant sampling.

---

**Fig. 2.** Phase transition map of compressive MUSIC (recovery rate \( \geq 99\% \)) with \( r = 12 \) and \( n = 200 \), when correct \( k - r \) partial support are given by an “oracle” estimator.

**Theorem 2 (Generalized Subspace Fitting Criterion):** Assume that \( 1 \leq l \leq r \) and we have a canonical MMV model \( AX = B \) where \( A \in \mathbb{R}^{m \times n} \), \( X \in \mathbb{R}^{n \times r} \), \( \|X\|_0 = k \), \( r < k < m < n \) and \( A \) satisfies the RIP condition with \( 0 \leq \delta_{2k-r+1}^L(A) < 1 \). Also, we assume that the nonzero rows of \( X \) are in general position. Then, suppose we are given an index set \( I \subset \{1, \ldots, n\} \) such that \( |I| \leq \min(2(k - r) + l, k) \) and \( |I \setminus \text{supp}X| \leq k - r + l \), then the following statements are
equivalent.

(i) \(|I \cap \operatorname{supp} X| \geq k - r + 1;\)

(ii) \(\operatorname{rank}[A_I B] < |I| + r;\)

(iii) There exist nonzero \(c \in \mathbb{R}^{|I|}\) and \(d \in \mathbb{R}^r\) such that \(\|Bd - A_I c\| = 0.\)

**Proof:** (i)\(\iff\)(ii): Assume that \(|I \cap \operatorname{supp} X| \geq k - r + 1\). Then \(|I| + r \leq 2k - r + l \leq m\) since \(\delta^L_{2k-r+l}(A) < 1\). If we take \(\tilde{I} \subset (I \cap \operatorname{supp} X)\) such that \(|\tilde{I}| = k - r + 1\), then

\[
\dim(R[A_I B]) \leq \dim(A_S) = k.
\]

However, \(|\tilde{I}| + r = k + 1\) so that \([A_I B]\) is not of full column rank. Hence \([A_I B]\) is not also of full column rank since \(\tilde{I} \subset I\).

Conversely, if we assume (6), there are \(p \in \mathbb{R}^{|I|}\) and \(q \in \mathbb{R}^r\) such that \(A_I p + AXq = 0\) and \([p, q]^T \neq 0\). If we let \(\hat{p} \in \mathbb{R}^n\) by \(\hat{p}^I = p\) and \(\hat{p}^c = 0\), we have \(A[I\hat{p} + Xq] = 0\). Since \(\|\hat{p} + Xq\|_0 \leq |I| - \operatorname{supp} X| + |\operatorname{supp} X| \leq 2k - r + l\), by the RIP condition, we have \(\hat{p} + Xq = 0\) so that \(\operatorname{supp}(\hat{p}) = \operatorname{supp}(Xq) \subset \operatorname{supp} X\). If we assume that \(|I \cap \operatorname{supp} X| \leq k - r\), then \(\|\hat{p}\|_0 \leq k - r\) but \(\|Xq\|_0 \geq k - r + 1\) since the nonzero rows of \(X\) are in general position. This is impossible so that \(|I \cap \operatorname{supp} X| \geq k - r + 1\).

(ii)\(\iff\)(iii): If \(\operatorname{rank}[A_I B] < |I| + r\), it is not full column rank. Hence there exist \([c^T, d^T]^T \neq 0\) such that \(A_I c + Bd = 0\). If we assume that \(c = 0\), then we have \(Bd = 0\). However, by the assumption of the canonical MMV model, the columns of \(B\) are linearly independent so that we have \(d = 0\), which contradicts to \([c^T, d^T]^T \neq 0\). Furthermore, since \(\operatorname{spark}(A) > 2k - r + l\), we have \(|I| < \operatorname{spark}(A)\). Hence, by a similar reasoning, we also have \(d \neq 0\). Finally, we can see that the converse is trivial.

Note that the criterion (iii) in Theorem 2 is similar to a subspace fitting criterion in classical sensor array processing [3], [20]. The main difference from the conventional subspace fitting in array signal processing is that the new criterion also deals with situation where some correct supports exist with the support estimate \(I\). Hence, we refer this new condition as generalized subspace fitting criterion.

It is important to note that the assumptions \(|I| \leq \min(2(k - r) + l, k)\) and \(|I \setminus \operatorname{supp} X| \leq k - r + l\) in Theorem 2 do not claim that there is a unique such index set \(I\); rather, it says there...
may be multiple of index set $I$ for a given $l$. For example, if $l = 1$, any index set $I$ such that $|I| = k - r + 1, \cdots, \min(2(k - r) + 1, k)$ that satisfies the condition $|I \setminus \text{supp} X| \leq k - r + l$, can be used to test the condition (i)-(iii) in Theorem 2. In particular, if we choose $|I| = k - r + 1$, Theorem 2 can be used to derive the following generalized MUSIC criterion.

\textbf{Corollary 1:} Assume that we have a canonical MMV model $AX = B$ where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, $\|X\|_0 = k$, $r < k < m < n$ and $A$ satisfies the RIP condition with $0 \leq \delta_{2k-r+1}^L(A) < 1$. Also we assume that the nonzero rows of $X$ are in general position. Then if we have $I_{k-r} \subset \text{supp} X$ with $|I_{k-r}| = k - r$, for any $j \in \{1, \cdots, n\} \setminus \text{supp} X$, $j \in \text{supp} X$ if and only if

$$\text{rank}[A_{I_{k-r} \cup \{j\}} B] < k + 1$$

or equivalently

$$a_j^* P_R([A_{I_{k-r}} B]) a_j = 0.$$  \hfill (7)

\textbf{Proof:} For an index set $I$ such that $|I| = k - r + 1$, the condition $|I \setminus \text{supp} X| \leq k - r + 1$ always holds. Therefore, $|I \cap \text{supp} X| = k - r + 1$ is equivalent to $I \subset \text{supp} X$. Hence, for $I = I_{k-r} \cup \{j\}$ such that $I_{k-r} \subset \text{supp} X$ and $|I_{k-r}| = k - r$, Eq. (6) is equivalent to Eq. (7), which is equivalent to say $a_j \in R([A_{I_{k-r}} B])$ or $a_j^* P_R([A_{I_{k-r}} B]) a_j = 0$. This concludes the proof. \hfill \blacksquare

Furthermore, if we have additional condition $r = k$, we have the following conventional MUSIC criterion.

\textbf{Corollary 2:} Assume that we have a canonical MMV model $AX = B$ where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, $\|X\|_0 = k$, $r < k < m < n$ and $A$ satisfies the RIP condition with $0 \leq \delta_{k+1}^L(A) < 1$. Also we assume that the nonzero rows of $X$ are in general position. Then for any $j \in \{1, \cdots, n\}$, $j \in \text{supp} X$ if and only if

$$\text{rank}[a_j B] < k + 1$$

or equivalently

$$a_j^* P_{R(B)} a_j = 0.$$  \hfill (8)

The equivalence of (8) and the MUSIC criterion (3) is trivial since $P_{R(B)} = QQ^*$. Up to now, we have shown that the generalized subspace fitting criterion in Theorem 2 can reproduce the existing results. However, the main advantage of the generalized subspace fitting criterion is its flexibility to be extended for more general situation. For example, if we set $l = r$ in Theorem 2.
we have the following corollary, which will be used in backward form of compressive subspace fitting algorithm.

**Corollary 3:** Assume that we have a canonical MMV model $AX = B$ where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, $\|X\|_0 = k$, $r < k < m < n$ and $A$ satisfies the RIP condition with $0 \leq \delta_{2k}^L(A) < 1$. Also, we assume that the nonzero rows of $X$ are in general position. Then, if we have an index set $I \subset \{1, \ldots, n\}$ such that $|I| \leq k$, the following are equivalent:

(i) $|I \cap \text{supp}
\mathbf{X}| \geq k - r + 1$;

(ii) rank$[A_I \mathbf{B}] < |I| + r$;

(iii) There exist nonzero $c \in \mathbb{R}^{|I|}$ and $d \in \mathbb{R}^r$ such that $\|Bd - A_Ic\| = 0$.

**Proof:** For $l = r$, the RIP condition becomes $0 \leq \delta_{2k}^L(A) < 1$ and for $|I| \leq k$, the condition $|I \setminus \text{supp}
\mathbf{X}| \leq k \leq 2k - r$ is trivially satisfied. This concludes the proof.

III. **FORWARD AND BACKWARD COMPRESSIVE SUBSPACE FITTING**

In this section, using the results from the previous section, we will develop two novel variation of compressive MUSIC algorithm with forward or backward support selection step to improve the noise robustness. We first start with the following Proposition.

**Proposition 1:** Assume that $1 \leq l \leq r$ and we have a canonical MMV model $AX = B$ where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, $\|X\|_0 = k$, $r < k < m < n$ and $A$ satisfies the RIP condition with $0 \leq \delta_{2k-r+l}^L(A) < 1$. Also, we assume that the nonzero rows of $X$ are in general position. Then, if we have an index set $I \subset \{1, \ldots, n\}$ such that $|I| \leq \min(2(k-r)+l, k)$ and $|I\setminus\text{supp}
\mathbf{X}| \leq k-r+l$, and if we have $|I \cap \text{supp}
\mathbf{X}| = k - r + q$ for some $q \geq 0$, then

$$\text{rank}[A_I \mathbf{B}] = |I| + r - q. \tag{9}$$

**Proof:** Take an $I_{k-r} \subset (I \cap \text{supp}
\mathbf{X})$ with $|I_{k-r}| = k - r$ and let $J_{k-r} = (I \cap \text{supp}
\mathbf{X}) \setminus I_{k-r}$, where $|J_{k-r}| = q$. Then by Theorem 2 we have

$$\text{rank}[A_{I \setminus J_{k-r}} \mathbf{B}] = |I \setminus J_{k-r}| + r = |I| + r - q.$$ 

Then for any $j \in J_{k-r}$, $a_j \in R([A_{I_{k-r}} \mathbf{B}]) = R(A_S)$ so that

$$\text{rank}[A_I \mathbf{B}] = \text{rank}[A_{I \setminus J_{k-r}} \mathbf{B}] = |I| + r - q$$

since $R([A_I \mathbf{B}]) = R([A_{I \setminus J_{k-r}} \mathbf{B}])$. $\blacksquare$
A. Forward Compressive Subspace Fitting

In [1], the compressive MUSIC first determine \( k - r \) indices of \( \text{supp} X \) with CS-based algorithms such as 2-thresholding or S-OMP, and it recovers remaining \( r \) indices of \( \text{supp} X \) with a generalized MUSIC criterion, which uses the projection operator onto the orthogonal complement of \( R([A_{I_{k-r}} B]) \). However, the following theorem can modify the MUSIC step to improve the noise robustness.

**Theorem 3 (Forward subspace fitting criterion):** Assume that \( 1 \leq l \leq r \) and we have a canonical MMV model \( AX = B \) where \( A \in \mathbb{R}^{m \times n} \), \( X \in \mathbb{R}^{n \times r} \), \( \| X \|_0 = k \), \( r < k < m < n \) and \( A \) satisfies the RIP condition with \( 0 \leq \delta_{2k-r+l}(A) < 1 \). Also, we assume that the nonzero rows are in general position. Then, if we have an index set \( I \subset \{1, \cdots, n\} \) such that \( |I| \leq \min(2(k-r)+l-1, k-1) \), \( |I \cap \text{supp} X| \leq k - r + l - 1 \) and \( |I \cap \text{supp} X| \geq k - r \), we have for \( j \notin I \), \( j \in \text{supp} X \) if and only if

\[
\text{rank}[A_I B] = \text{rank}[A_{I \cup \{j\}} B] \tag{10}
\]

or equivalently

\[
a^*_j P_{R([A_I B])} a_j = 0. \tag{11}
\]

**Proof:** By the condition we have \( |I \cup \{j\}| \leq \min(2(k-r)+l, k) \) and \( |(I \cup \{j\}) \setminus \text{supp} X| \leq k - r + l \) so that we can apply Proposition [1] for \( I \) and \( I \cup \{j\} \) since \( |I \cap \text{supp} X| \geq k - r \). If \( |I \cap \text{supp} X| = k - r + q \) for some \( q \geq 0 \), then we have \( \text{rank}[A_I B] = |I| + r - q \). Then for any \( j \notin I \), if we have \( j \in \text{supp} X \), \( |(I \cup \{j\}) \cap \text{supp} X| = k - r + (q + 1) \) so that we have

\[
\text{rank}[A_{I \cup \{j\}} B] = |I| + 1 + r - (q + 1) = |I| + r - q = \text{rank}[A_I B].
\]

On the other hand, if we have \( j \notin \text{supp} X \), \( |(I \cup \{j\}) \cap \text{supp} X| = k - r + q \) so that we have

\[
\text{rank}[A_{I \cup \{j\}} B] = |I| + 1 + r - q > \text{rank}[A_I B].
\]

Finally, (10) is equivalent to \( a_j \in R([A_I B]) \), which is also equivalent to (11). This completes the proof.

Note that \( R([A_I B]) = R([A_{I_{k-r}} B]) \) and \( \dim R([A_{I_{k-r}} B]) = k \) for all \( I \subset \text{supp} X \) and \( k - r \leq |I| \leq \min(2(k-r)+l-1, k-1) \). This indicates that we first need to find \( I_{k-r} \) support using a compressive sensing algorithm, then we augment the newly added support into the initial estimate \( I_{k-r} \) to improve the noise robustness of the estimation of the \( k \)-dimensional signal subspace. In
summary, we propose the following forward CSF algorithm as in Table I. Note that the algorithm is very similar to CS-MUSIC; however, compared to the CS-MUSIC, support augmentation step is also added in MUSIC step, which is the main ingredient to improve the noise robustness.

**TABLE I**

<table>
<thead>
<tr>
<th>Algorithm : Forward CSF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input:</strong> $Y = AX + N$ where $N$ is the measurement noise.</td>
</tr>
<tr>
<td><strong>1) compressive sensing step</strong></td>
</tr>
<tr>
<td>- Find $k - r$ indices of $\text{supp}X$ by any MMV compressive sensing algorithms such as 2-thresholding or S-OMP.</td>
</tr>
<tr>
<td>- Let $I_{k-r}$ be the set of selected indices and let $S = I_{k-r}$.</td>
</tr>
<tr>
<td><strong>2) forward subspace fitting</strong></td>
</tr>
<tr>
<td>- Set $l = 0$.</td>
</tr>
<tr>
<td>- While $l &lt; r$, do the following procedure:</td>
</tr>
<tr>
<td>1. Perform an SVD of $[ASY] = [U_1, U_0]\text{diag}([\Sigma_1, \Sigma_0])[V_1, V_0]^*$, where $\Sigma_1 = \text{diag}([\sigma_1, \cdots, \sigma_k]$ and $\Sigma_0 = \text{diag}([\sigma_{k+1}, \cdots, \sigma_{k+l}])$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{k+l}$.</td>
</tr>
<tr>
<td>2. Take $j_l = \text{arg min}<em>{j \notin S} | P</em>{R(U_1)}a_j |^2$.</td>
</tr>
<tr>
<td>3. Set $S = S \cup {j_l}$, let $l = l + 1$ and goto step 1.</td>
</tr>
<tr>
<td>- Then $S$ is the support estimation of $X$.</td>
</tr>
</tbody>
</table>

**B. Backward Compressive Subspace Fitting**

As discussed before, we can easily expect that the performance of the compressive MUSIC is very dependent on the selection of $k - r$ correct indices of the support of $X$. Note that this is very stringent condition. In practice, even though the first consecutive steps of, for example, S-OMP may not provide all true partial supports, it is more likely that among the $k$-sparse solution of S-OMP, part of the supports (not in a sequential order) can be correct. Hence, if the estimate of the support of $X$ has at least $k - r$ indices of the support of $X$ and we can identify them, then we can expect that the performance of the compressive MUSIC will be improved. When $\binom{k}{k-r}$ is small, we may apply the exhaustive search, but if both $k - r$ and $r$ are not small, then the exhaustive search is hard to apply so that we have to find some alternative method to identify the correct indices from the estimate of $\text{supp}X$. Indeed, the following backward subspace fitting criterion can address the problem.

More specifically, we can relax the stringent requirement of $k - r$ correct CS steps. Instead, the new algorithm requires that $k - r + 1$ supports (not in sequential order) out of a larger size
support estimate is correct. Then, the location of the correct $k - r$ support can be readily estimated using the following backward subspace fitting criterion. Compared to the forward subspace fitting criterion that improve the noise robustness of MUSIC step, the following backward subspace fitting criterion can also improve the noise robustness of the compressive sensing step for partial support recovery. In particular, by utilizing both backward and forward subspace fitting, we can improve the noise robustness of both CS step as well as MUSIC step.

**Theorem 4 (Backward subspace fitting criterion):** Assume that $1 \leq l \leq r$ and we have a canonical MMV model $AX = B$ where $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times r}$, $\|X\|_0 = k$, $r < k < m < n$ and $A$ satisfies the RIP condition with $0 \leq \delta_{2k-r+l}^L(A) < 1$. Also, we assume that the nonzero rows are in general position. Then, if we have an index set $I \subset \{1, \cdots, n\}$ such that $|I| = \min(2(k-r)+l, k)$, $|I \setminus \text{supp}X| \leq k - r + l$ and $|I \cap \text{supp}X| \geq k - r + 1$, then we have for $j \in I$, $j \in \text{supp}X$ if and only if

$$
\text{rank}[A_{I \setminus \{j\}} B] = \text{rank}[A_I B]
$$

or equivalently

$$
a_j^* P_{R([A_{I \setminus \{j\}} B])} a_j = 0.
$$

**Proof:** Assume that $|I \cap \text{supp}X| = k - r + q$ where $q \geq 1$. Then by Proposition II we have $\text{rank}[A_I B] = |I| + r - q$. Noting that $I \setminus \{j\}$ satisfies the assumptions of Proposition II for any $j \in I$, if we have $j \in \text{supp}X$, $|I \cap \text{supp}X| = k - r + (q - 1)$ so that we have

$$
\text{rank}[A_{I \setminus \{j\}} B] = |I| - 1 + r - (q - 1) = |I| + r - q = \text{rank}[A_I B].
$$

On the other hand, if we have $j \notin \text{supp}X$, $|I \cap \text{supp}X| = k - r + q$ so that we have

$$
\text{rank}[A_{I \setminus \{j\}} B] = |I| - 1 + r - q = |I| + r - q - 1 < \text{rank}[A_I B].
$$

Finally, due to the rank condition, we know $a_j \in R([A_{I \setminus \{j\}} B])$ if and only if $j \in \text{supp}X$. Hence $a_j^* P_{R([A_{I \setminus \{j\}} B])} a_j = 0$ if and only if $j \in \text{supp}X$. That completes the proof.

The above theorem states that if we have a partial estimate of support of $X$ that has at least $k - r + 1$ correct indices of support of $X$, we can identify the correct part of the estimated partial support of $X$ by using backward subspace fitting criterion by testing each atom one by one. In particular, if we have an RIP condition $0 \leq \delta_{2k}^L(A) < 1$ and have an estimated $k$ supports (using any compressive sensing algorithm) that includes at least $k - r + 1$ correct indices, then we can identify the correct part of the estimated support as described as following corollary.
Corollary 4: Assume that we have a canonical MMV model \( AX = B \) where \( A \in \mathbb{R}^{m \times n}, \) 
\( X \in \mathbb{R}^{n \times r}, \) \( \| X \|_0 = k, \) \( r < k < m < n \) and \( A \) satisfies the RIP condition with \( 0 \leq \delta_{2k}^L(A) < 1. \) 
Also, we assume that the nonzero rows are in general position. Then, if we have an index set \( I \subset \{ 1, \cdots, n \} \) such that \( |I| = k \) and \( |I \cap \text{supp} X| \geq k - r + 1, \) for \( j \in I, \) we have \( j \in \text{supp} X \) if and only if \[
\text{rank}[A_{I \setminus \{j\}} B] = \text{rank}[A_I B]
\] or equivalently \[
\mathbf{a}_j^* P_{R([A_{I \setminus \{j\}} B])] \mathbf{a}_j = 0.
\]
Corollary 4 allows us the following backward CSF as described in Table II.

<table>
<thead>
<tr>
<th>Algorithm: Backward CSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input: ( Y = AX + N ) where ( N ) is the measurement noise.</td>
</tr>
<tr>
<td>1) compressive sensing step</td>
</tr>
<tr>
<td>- Let ( S = \emptyset. )</td>
</tr>
<tr>
<td>- If ( r &lt; k, ) estimate ( k ) indices of ( \text{supp} X ) with existing MMV algorithms such as mixed norm approach, thresholding estimate or S-OMP and so on. Then, let ( I ) be the support estimate of ( X. )</td>
</tr>
<tr>
<td>- If ( r = k, ) goto step 3.</td>
</tr>
<tr>
<td>2) Backward subspace fitting</td>
</tr>
<tr>
<td>- For ( j \in I, ) calculate the quantities ( \zeta(j) := | P_{R([A_{I \setminus {j}} Y])] \mathbf{a}_j |_2^2. )</td>
</tr>
<tr>
<td>- Making an ascending ordering of ( \zeta(j) ) for ( j \in I, ) choose indices that corresponds the first ( k - r ) indices and put these indices into ( S. )</td>
</tr>
<tr>
<td>3) Forward subspace fitting</td>
</tr>
<tr>
<td>- Set ( l = 0. )</td>
</tr>
<tr>
<td>- While ( l &lt; r, ) do the following procedure:</td>
</tr>
<tr>
<td>1. Perform an SVD of ( [A_S Y] = [U_1, U_0] \text{diag}[\Sigma_1, \Sigma_0][V_1, V_0]^*, ) where ( \Sigma_1 = \text{diag}[\sigma_1, \cdots, \sigma_k] ) and ( \Sigma_0 = \text{diag}[\sigma_{k+1}, \cdots, \sigma_{k+l}] ) where ( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{k+l}. )</td>
</tr>
<tr>
<td>2. Take ( j_l = \arg \min_{j \not\in S} | P_{R(U_j)} \mathbf{a}_j |_2^2. )</td>
</tr>
<tr>
<td>3. Set ( S = S \cup { j_l }, ) let ( l = l + 1 ) and goto step 1.</td>
</tr>
<tr>
<td>- Then ( S ) is the support estimation of ( X. )</td>
</tr>
</tbody>
</table>

The paradigm shift from early termination of CS algorithm after \( k - r \) step to selecting the correct \( k - r \) supports out of \( k \)-sparse solution by any CS algorithm is much more significant and fundamental than just an algorithmic improvement. In particular, by converting the problem
as a partial support recovery problem, we expect that we can adapt rich information theoretical analysis tools that have been developed for partial support recovery in single measurement vector CS (SMV-CS) [21], [22].

IV. PERFORMANCE ANALYSIS OF FORWARD- AND BACKWARD- CSF

In practice, the measurements are noisy, so the theory we derived for noiseless measurement should be modified. In our previous work [1], 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. Even though the derivation was based on a large system model with a Gaussian sensing matrix, it has provided very useful insight. Therefore, we again employed a large system model.

Definition 4: A large system noisy canonical MMV model, 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:

$$\text{minimize} \quad \|X\|_0$$
$$\text{subject to} \quad Y = AX + N,$$

where $A \in \mathbb{R}^{m \times n}$ is a random matrix with i.i.d. $N(0, 1/m)$ entries, $N = [n_1, \ldots, n_r] \in \mathbb{R}^{m \times r}$ is an additive noise matrix, $m = m(n) \to \infty$, $k = k(n) \to \infty$ as $n \to \infty$ and $\text{rank}(AX) = r \leq k = \|X\|_0$ 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, where $m$, $k$ and $r$ satisfy $k/m < 1 - \epsilon$ and $r/k < 1 - \epsilon$.

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 $n \to \infty$. The following theorem provides a theoretical justification why augmenting generalized MUSIC step improves the noise robustness.

Theorem 5: Assume that we have a noisy canonical MMV model $Y = AX$. For $0 \leq l < r$, if we have $I_{k-r+l} \subseteq \text{supp}X$ such that $|I_{k-r+l}| = k - r + l$ and singular value decomposition of $[A_{I_{k-r+l}} Y]$ as

$$[A_{I_{k-r+l}} Y] = \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^* + \tilde{U}_0 \tilde{\Sigma}_0 \tilde{V}_0^*,$$
where $\tilde{\Sigma}_1 = \text{diag}[\tilde{\sigma}_1, \cdots, \tilde{\sigma}_k]$, $\tilde{\Sigma}_0 = \text{diag}[\tilde{\sigma}_{k+1}, \cdots, \tilde{\sigma}_{k+l}]$ and $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \cdots \geq \tilde{\sigma}_{k+l}$, then in the large system limit,

$$j_l = \arg \min_{j \not\in I_{k-r+l}} \|P_{R(\tilde{U}_i)}^j a_j\|^2 \in \text{supp}X$$

(13)

provided that

$$\frac{\sigma_k([A_{I_{k-r+l}} B])}{\|N\|} > \frac{1}{1 - \gamma} + 1$$

(14)

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

Proof: See Appendix A.

Note that if $I_{k-r+l} \subset \text{supp}X$ for $1 \leq l \leq r$, we have rank$([A_{I_{k-r+l}} B]) = k$ so that the columns of $[A_{I_{k-r+l}} B]$ is a frame in $R(AS)$ with lower frame bound $\sigma_k^2([A_{k-r+l} B])$. Hence, in this case, $\sigma_k([A_{I_{k-r+l}} B])$ is an increasing function of $l$, so the implication of Theorem 5 is that as $l$ increases, the frame becomes more redundant and the lower frame bound become larger. Hence, the SNR requirement of success of MUSIC step can be improved by augmenting newly estimated support in calculating the signal subspace. As will be shown in numerical results, such advantages are significant especially when the measurement are noisy. Moreover, the following theorem shows that backward support selection is also feasible in spite of noise.

Theorem 6: Assume that we have a noisy canonical large system MMV model $Y = AX$ with $0 \leq \delta_{2k}(A) < 1$. Then if we have an index set $I \subset \{1, \cdots, n\}$ which satisfies $|I| = k$, $|I \cap \text{supp}X| \geq k - r + 1$ and

$$\frac{\sigma_r(B)}{\|N\|} > \frac{1}{1 - \gamma(1 + \alpha)} + 1,$$

(15)

where $\gamma := \lim_{n \to \infty} k/m$ and $\alpha := \lim_{n \to \infty} r/k$, we have

$$\min_{j \not\in \text{supp}X} a_j^* P_{R(A_{\setminus\{j\}})} y_j a_j > \max_{j \in \text{supp}X} a_j^* P_{R(A_{\setminus\{j\}})} y_1 a_j$$

(16)

in the large system limit.

Proof: See Appendix B.

V. NUMERICAL RESULTS

In this section, we demonstrate the performance of forward/backward CSF. These new algorithms are compared to S-OMP algorithm and original compressive MUSIC. We declared the algorithm as a success if the estimated support is the same as the true $\text{supp}X$, and the success rates were averaged over 500 experiments. The simulation parameters were as follows: $m \in \{1, 2, \ldots, 60\}$,
Elements of sensing matrix $A$ were generated from a Gaussian distribution having zero mean and variance of $1/m$, and then each column of $A$ is normalized to have an unit norm. The $\text{supp}X$ were chosen randomly, and the maximum iteration was set to $k$ for the S-OMP algorithm. Subspace S-OMP is used to find the partial support recovery for CS-MUSIC and forward CSF algorithms, while conventional S-OMP is used to find $k$ supports for backward CSF algorithm.

Fig. 3 illustrates the 3D phase transition maps with respect to $k$, $m$ and $r$ for various MMV algorithms, when $n = 200$. Each column of Fig. 3 (from left to right) indicates the result using CS-MUSIC, forward CSF, backward CSF, and S-OMP, respectively; and SNR = 40dB for the first row and SNR = 20dB for the second row, respectively. The overlayed curves denote the theoretical noiseless $l_0$ bound. Note that the performance of CS-MUSIC is better than S-OMP at SNR = 40dB but is degraded as the noise level increased; however, forward and backward CSFs are more robust than CS-MUSIC, and outperform both CS-MUSIC and S-OMP algorithms.
Fig. 4 illustrates the boundary for phase transition map by various MMV algorithms (recovery rate \( \geq 99\% \)) when \( n = 200, r \in \{3,12\} \), and \( \text{SNR} \in \{40,20dB\} \). The overlayed solid red curves denote the theoretical \( l_0 \) bound. As shown in Fig. 4, the boundary of CS-MUSIC is greatly degraded as the noise level is increased. Moreover, CS-MUSIC does not take advantage of redundant sampling to improve noise robustness compared to S-OMP and we can even observe performance crossing between the two algorithms. However, forward and backward CSFs solved these problems and nearly achieved theoretical \( l_0 \) bound when \( r = 12 \).

![Figure 4](image)

Fig. 4. Phase transition maps of various MMV algorithms (recovery rate \( \geq 99\% \)) when \( n = 200 \). (a) \( r = 3 \), \( \text{SNR} = 40dB \) (b) \( r = 3 \), \( \text{SNR} = 20dB \), (c) \( r = 12 \), \( \text{SNR} = 40dB \), and (d) \( r = 12 \), \( \text{SNR} = 20dB \). The overlayed solid red curves denote the theoretical \( l_0 \) bound and dotted red curves denote the bound for successful partial support recovery with subspace S-OMP.

In order to compare the proposed algorithm with other methods more clearly, the recovery rates of various algorithms are plotted in Fig. 5 when \( n = 100 \) and \( m \in \{20,40\} \). The first and second row of Fig. 5 are the recovery rates for \( \text{SNR} = 40dB \), and the last row is for \( \text{SNR} = 20dB \); whereas the first and second column are for \( r = 3 \) and \( r = 6 \), respectively. Under the same noise level of \( \text{SNR} = 40dB \), CS-MUSIC outperforms S-OMP when \( m = 20 \); however, the recovery rate of CS-MUSIC is less than that of S-OMP for a certain range of \( k \) when \( m = 40 \). This phenomenon again tells the original CS-MUSIC algorithm does not fully take advantage of the oversampled...
measurement when the measurement is noisy. Moreover, when \( m = 40 \), the recovery rate of CS-MUSIC rapidly decreases as the noise level increases. However, forward CSF overcomes these problems, and consistently outperformed S-OMP. Moreover, further enhancement can be achieved by backward CSF as shown in Fig. 5.

Fig. 5. Recovery rates by various MMV algorithms when \( n = 100 \), \( m \in \{20, 40\} \), and \( \text{SNR} \in \{40, 20\} \). The vertical lines correspond to the theoretical \( l_0 \) bound.
Next, we performed the simulation on the recovery results for three different types of the RIP condition using various MMV algorithms. More specifically, we assumed the sensing matrix follows $A \sim \mathcal{N}(a, 1/m)$ and then normalized each column of $A$ to have an unit norm. We set $a = 0, 1,$ and $2$, where larger $a$ represents worse RIP condition. In this simulation, we set SNR = 40dB, $n = 100$, $m = 40$, and $r = 9$. Fig. 6(a)-(c) show the results of CS-MUSIC, forward CSF, and backward CSF, respectively. Note that the backward CSF is significantly robust for RIP condition compared to other two methods due to the backward procedure of CSF improves the compressed sensing stage (S-OMP) which heavily depends on the RIP condition of the sensing matrix $A$, and nearly approaches the vertical lines that denote the theoretical $l_0$ bound.

![Fig. 6. Recovery rates by various RIP condition of $A$ (larger mean value provides worse RIP condition). The measurements are corrupted by additive Gaussian noise of SNR = 40dB and $n = 100$, $m = 40$, and $r = 9$. (a) CS-MUSIC, (b) forward CSF, and (c) backward CSF. The vertical lines correspond to the theoretical $l_0$ bound.](image)

To show the relationship between the recovery performance in the condition number of matrices $X$, we performed the simulation on the the recovery results for three different types of the source model $X$. More specifically, the singular values of $X$ are given by $\sigma_j = \tau^{j-1}$ for $j = 1, \cdots, \text{rank}(X)$. In this simulation, we also set SNR = 40dB, $n = 100$, $m = 40$, and $r = 9$. The results in Fig. 7 provide evidence of the significant impact of the condition number of $X$. Again, forward and backward CSF consistently outperform the original CS-MUSIC.

VI. CONCLUSION

The original version of compressive MUSIC or SA-MUSIC estimates $k-r$ entries of the support using conventional MMV algorithms, while the remaining $r$ support indices are estimated using a generalized MUSIC criterion so that it optimally combines the CS and MUSIC. Although the
Fig. 7. Recovery rates by various condition number of $X$ (smaller $\tau$ provides larger condition number). The measurements are corrupted by additive Gaussian noise of $\text{SNR} = 40\text{dB}$ and $n = 100$, $m = 40$, and $r = 9$. (a) CS-MUSIC, (b) forward CSF, and (c) backward CSF. The vertical lines correspond to the theoretical $l_0$ bound.

Compressive MUSIC or SA-MUSIC produces significantly better performances than any other conventional MMV algorithms in the noiseless case, the performance improvement is reduced when the multiple measurements are noisy. We showed that the performance degradation is due to the MUSIC step that does not fully exploit the redundant sampling to improve noise robustness. To address this problem, we developed a novel subspace fitting criterion that extends the generalized MUSIC criterion for a general RIP setting, and introduced a forward and backward compressive subspace fitting algorithm which significantly improve the robustness from the noise. Extensive numerical simulation demonstrated that the new algorithm consistently outperforms existing CS-MUSIC and S-OMP, and nearly achieves the $l_0$-bound robustly for sufficiently large snapshots.

**APPENDIX A**

To obtain the SNR bound for each iteration in generalized MUSIC step with augmentation, we use the following theorem for the perturbation of signal subspace. To state the results, we define the followings:

**Definition 5:** Let $P_L$ and $P_M$ denote the orthogonal projections onto subspaces $L$ and $M$. We define the angle between two subspace $L$ and $M$ as

$$\sin \theta(L, M) = \|(I - P_M)P_L\|$$

for any unitary invariant norm.
If we assume that \( \dim(L) = \dim(M) \), it is known that \[23\], \[24\]

\[
\sin \theta(L, M) = \sin \theta(M, L)
\]

and

\[
\|P_L - P_M\| = \sin \theta(L, M).
\]

**Theorem 7 (Wedin \[23\], \( \sin \theta \) theorem):** Assume that \( G \in \mathbb{R}^{m \times q} \) has the singular value decomposition

\[
G = U \Sigma V^* = U_1 \Sigma_1 V_1^* + U_0 \Sigma_0 V_0^* = G_1 + G_0
\]

where \( G_1 := U_1 \Sigma_1 V_1^* \) and \( G_0 := U_0 \Sigma_0 V_0^* \). Also, for a perturbed matrix \( \tilde{G} \in \mathbb{R}^{m \times q} \) of \( A \), assume that \( \tilde{G} \) has the singular value decomposition

\[
\tilde{G} = \tilde{U} \tilde{\Sigma} \tilde{V}^* = \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^* + \tilde{U}_0 \tilde{\Sigma}_0 \tilde{V}_0^* = \tilde{G}_1 + \tilde{G}_0
\]

where \( \tilde{G}_1 := \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1 \) and \( \tilde{G}_0 := \tilde{U}_0 \tilde{\Sigma}_0 \tilde{V}_0^* \), and \( U_1 \) and \( \tilde{U}_1 \) (or \( U_0 \) and \( \tilde{U}_0 \)) are the matrices of same size. If there exist \( \alpha \geq 0 \) and \( \delta > 0 \) such that

\[
\sigma_{\min}(\tilde{G}_1) \geq \alpha + \delta \text{ and } \sigma_{\max}(G_0) \leq \alpha,
\]

then for every unitary invariant norm,

\[
\sin \theta(R(\tilde{G}_1), R(G_1)) \leq \frac{\epsilon}{\delta},
\]

where

\[
\epsilon := \max(\|R_1\|, \|R_2\|), \quad R_1 := -NV_1, \quad R_2 := -N^*U_1.
\]

**Proof:** For the proof, see \[23\].

*Proof of Theorem 5* For a noiseless measurement \([A_{I_{k-r+i}}, B]\), we have

\[
\sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_{k+l} = 0
\]

so that we have \( \sigma_{\min}(\tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^*) \geq \sigma_k([A_{I_{k-r+i}}, B]) - \|N\| \) and \( \sigma_{\max}(U_0 \Sigma_0 V_0^*) = 0 \) so that if we have \( \|N\| < \sigma_k([A_{I_{k-r+i}}, B]) \), we can apply Theorem 7. By Theorem 7 we have

\[
\sin \theta(R(\tilde{U}_1), R(U_1)) \leq \frac{\max(\|NV_1\|, \|N^*U_1\|)}{\sigma_k([A_{I_{k-r+i}}, B]) - \|N\|} \leq \frac{\|N\|}{\sigma_k([A_{I_{k-r+i}}, B]) - \|N\|}
\]
if \( \|N\| < \sigma_k([A_{I_k-r+l}] B) \). Noting that \( \|P_{R(U_1)}^\perp a_j\|^2 = 0 \) for \( j \notin \text{supp} X \) by the generalized MUSIC criterion, for any \( j \notin \text{supp} X \) and \( q \in \text{supp} X \) we have

\[
\|P_{R(U_1)}^\perp a_j\|^2 - \|P_{R(U_1)}^\perp a_q\|^2 = a_j^* P_{R(U_1)}^\perp a_j - a_q^* [P_{R(U_1)} - P_{R(U_1)}] a_q
\]

\[
= a_j^* P_{R(U_1)}^\perp a_j - \|P_{R(U_1)} - P_{R(U_1)}\|\|a_q\|^2
\]

\[
\geq a_j^* P_{R(U_1)}^\perp a_j - \frac{\|N\|}{\sigma_k([A_{I_k-r+l}] B)} - \|N\|\|a_q\|^2.
\]

Since \( a_{i,j} \)'s are i.i.d. normal distribution with zero mean and variance \( 1/m \) and \( a_j \) is independent of \( P_{R(U_1)}^\perp \) for any \( j \notin \text{supp} X \), \( a_j^* P_{R(U_1)}^\perp a_j \) is a chi-squared random variable of degree of freedom \( m - k \) since \( \text{rank}(\tilde{U}_1) = k \). Also, for each \( 1 \leq j \leq n \), \( m\|a_j\|^2 \) is a chi-squared random variable with degree of freedom \( m \) so that we have by Lemma 3 in [25],

\[
\lim_{n \to \infty} \frac{\max_{1 \leq j \leq n} \|a_j\|^2}{m} = 1,
\]

since \( \lim_{n \to \infty} (\log n)/m = 0 \). Since \( \lim_{n \to \infty} (\log (n - k))/(m - k) = 0 \), by Lemma 3 in [25], we have

\[
\lim_{n \to \infty} \min_{j \notin \text{supp} X} \frac{\max_j \|a_j\|^2}{m - k} = 1
\]

so that

\[
\lim_{n \to \infty} \frac{\min_{j \notin \text{supp} X} \|a_j\|^2}{\max_{1 \leq j \leq n} \|a_j\|^2} = \lim_{n \to \infty} \left( \frac{\min_{j \notin \text{supp} X} \|a_j\|^2}{m - k} \right) \left( \frac{1}{\max_{1 \leq j \leq n} \|a_j\|^2} \right) = 1 - \gamma.
\]

Hence, we have (13) if we have

\[
1 - \gamma - \left( \frac{\sigma_k([A_{I_k-r+l}] B)}{\|N\|} - 1 \right)^{-1} > 0
\]

in the large system limit. This completes the proof.
APPENDIX B

By the assumption $|I \cap \text{supp} X| \geq k - r + 1$, and the generalized MUSIC criterion, we have $\|P^\perp_{[A_{I \setminus \{j\}} B]} a_j\|^2 = 0$ for any $j \in \text{supp} X$. Then, for any $j \in \text{supp} X$ and $q \in \text{supp} X$, we have

$$\|P^\perp_{[A_{I \setminus \{j\}} Y]} a_j\|^2 - \|P^\perp_{[A_{I \setminus \{q\}} Y]} a_q\|^2 \geq \|P^\perp_{[A_{I_k \setminus \{q\}} Y]} a_q\|^2 - \|P^\perp_{[A_{I_k \setminus \{q\}} Y]} a_q\|^2 = \|P^\perp_{[A_{I_k \setminus \{q\}} Y]} a_q\|^2 + \|P^\perp_{[A_{I_k \setminus \{q\}} B]} a_q\|^2 \geq \|P^\perp_{[A_{I_k \setminus \{j\}} Y]} a_j\|^2 - \|P^\perp_{[A_{I_k \setminus \{q\}} Y]} a_q\|^2,$$

since $P^2 = P$ for any orthogonal projection operator $P$, where $I_{k-r,q} \subset (I \setminus \{q\}) \cap \text{supp} X$. For $j \notin \text{supp} X$, $a_j$ is statistically independent from $R([A_{I \setminus \{j\}} Y])$ so that $m\|P_R([A_{I \setminus \{j\}} Y]) a_j\|^2$ is chi-squared random variable with degree of freedom $m - k - r + 1$ so that

$$\lim_{n \to \infty} \frac{\min_{j \notin \text{supp} X} \|p^\perp_{[A_{I \setminus \{j\}}]} a_j\|^2}{\max_{1 \leq j \leq n} \|a_j\|^2} = \lim_{n \to \infty} \left( \frac{\min_{j \notin \text{supp} X} \|p^\perp_{[A_{I \setminus \{j\}}]} a_j\|^2}{m - k - r + 1} \frac{1}{\max_{1 \leq j \leq n} \|a_j\|^2} \right) = 1 - \gamma(1 + \alpha)$$

where $\gamma := \lim_{n \to \infty} k/m$ and $\alpha := \lim_{n \to \infty} r/k$.

On the other hand, for any $q \in \text{supp} X$, $m\|a_q\|^2$ is a chi-squared random variable of degree of freedom $m$, so that by Lemma 3 in [25] we have

$$\lim_{n \to \infty} \max_{q \in \text{supp} X} \|a_q\|^2 = 1.$$

Since $[A_{I_k \setminus \{q\}} B]$ has a full column rank, by the projection update rule, we have

$$P_R([A_{I_k \setminus \{q\}} B]) = P_R(A_{I_k \setminus \{q\}}) + P_R(P^\perp_R(A_{I_k \setminus \{q\}} B))$$

and

$$P_R([A_{I_k \setminus \{q\}} Y]) = P_R(A_{I_k \setminus \{q\}}) + P_R(P^\perp_R(A_{I_k \setminus \{q\}} Y)).$$

Then by applying Theorem 5 (or Theorem 7), we have

$$\|P_R([A_{I_k \setminus \{q\}} B]) - P_R([A_{I_k \setminus \{q\}} Y])\| \leq \|P_R(P^\perp_R(A_{I_k \setminus \{q\}} B)) - P_R(P^\perp_R(A_{I_k \setminus \{q\}} Y))\| \leq \|P_R(B) - P_R(Y)\| \leq \frac{\|N\|}{\sigma_{\min}(B) - \|N\|}.$$
Hence (16) holds if we have
\[ 1 - \gamma (1 + \alpha) - \left( \frac{\sigma_{\text{min}}(B)}{\|N\|} - 1 \right)^{-1} > 0 \]
in the large system limit. This completes the proof.

ACKNOWLEDGMENT

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No.2011-0000353).

REFERENCES


