

# Improving Noise Robustness in Subspace-based Joint Sparse Recovery

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

**Abstract**—In a multiple measurement vector problem (MMV), where multiple signals share a common sparse support and are sampled by a common sensing matrix, we can expect joint sparsity to enable a further reduction in the number of required measurements. While a diversity gain from joint sparsity had been demonstrated earlier in the case of a convex relaxation method using an  $l_1/l_2$  mixed norm penalty, only recently was it shown that similar diversity gain can be achieved by greedy algorithms if we combine greedy steps with a MUSIC-like subspace criterion. However, the main limitation of these hybrid algorithms is that they often require a large number of snapshots or a high signal-to-noise ratio (SNR) for an accurate subspace as well as partial support estimation. One of the main contributions of this work is to show that the noise robustness of these algorithms can be significantly improved by allowing sequential subspace estimation and support filtering, even when the number of snapshots is insufficient. Numerical simulations show that a novel sequential compressive MUSIC (sequential CS-MUSIC) that combines the sequential subspace estimation and support filtering steps significantly outperforms the existing greedy algorithms and is quite comparable with computationally expensive state-of-art algorithms.

**Index Terms**—Compressed sensing, greedy algorithm, multiple measurement vector problems, subspace estimation

## I. INTRODUCTION

We study a multiple measurement vector (MMV) problem, where multiple signals share a same common sparse support set and each signal is measured by multiplying it by a measurement matrix. An MMV problem is one way in which multiple correlated signals can appear in a signal ensemble, and MMV problems also have many important applications [1], [2], [3], [4]. A central theme in these studies has been that joint sparsity within signal ensembles enables a further reduction in the number of required measurements [5], [6], where the number of measurements required per sensor must

account for the minimal features unique to that sensor [7], [8], [9], [10], [11], [12]. Indeed, for the case of an  $l_1/l_2$  mixed norm approach, Obozinski *et al.* [13] showed that a near optimal diversity gain can be achieved.

Recently, Kim *et al.* [7] and Lee *et al.* [12] independently showed that such a diversity gain can be also achieved in a new class of greedy algorithms by exploiting the so-called generalized (or extended) multiple signal classification (MUSIC) criterion [7], [12]. More specifically, these algorithms obtain a partial support estimate using a conventional MMV greedy algorithm, and then the atoms corresponding to the partial supports are augmented into a data matrix to obtain an augmented signal subspace estimate. Finally, a MUSIC-like [14] criterion is derived for the augmented subspace to find the remaining support. The hybridization makes these hybrid greedy algorithms fully utilize a diversity gain so that the algorithms outperform all the existing greedy methods.

The performance improvement of these greedy algorithms is substantial and nearly achieves the  $l_0$  bound when a signal subspace and partial support estimation are accurate due to a sufficient number of snapshots or high signal to noise ratio (SNR) [7], [12]. However, if either of these estimation is erroneous owing to an insufficient number of snapshots or low SNR, performance degrades. Similar observations have been made in the literature on classical array signal processing [15], [16]. In [15], prior knowledge of the direction-of-arrival (DOA) has been incorporated to improve the performance of MUSIC by filtering out the known sources via orthogonal projections. However, as shown in [16], such orthogonal projection is suboptimal from a statistical standpoint.

While increasing the number of snapshots is relatively easier in classical sensor array signal processing problems, in some MMV problems such as parallel MR imaging [2], an additional snapshot requires a hardware change by adding a new receiver coil. Hence, in these problems, exploiting other dimensions would be beneficial. We are aware that joint sparse recovery methods such as Bayesian approaches [17], [18], or convex optimization techniques [19], are shown to be statistically robust in the direction of arrival estimation problems, as first demonstrated by Malitov *et al.* [20] and further developed by Stoica *et al.* [18]. In particular, Stoica *et al.* derived so-called LIKES (LIKelihood-based Estimation of Sparse parameters) method obtained in a hyperparameter-free manner from the maximum-likelihood principle similar to M-SBL (Multiple Sparse Bayesian Learning) by Wipf *et al.* [17], but improves

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

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, Daejeon 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, Daejeon 305-701, Korea (phone:82-42-350-4320; fax:82-42-350-4310; e-mail:jong.ye@kaist.ac.kr)

This research was supported by the National Research Foundation of Korea (NRF) grant funded in part by the Korean government (MEST) (No. 2011-0030933).

its performance while reducing the complexity. However, these approaches are still computationally expensive for MMV problems with a large number of sensors, so we need a new greedy algorithm that achieves a similar optimal performance with a significantly reduced computational complexity.

Therefore, one of the main goals of this paper is to address how these hybrid greedy methods can be made robust without increasing the number of snapshots. One important contribution is a new theory explaining that the generalized MUSIC criterion is a special case of a new subspace criterion that can be used to derive two sequential strategies to improve the accuracy of an augmented signal subspace estimation. More specifically, a forward greedy subspace estimation step improves the robustness of an augmented signal subspace estimation by adding newly discovered atoms in the MUSIC step, whereas the backward support filtering provides additional robustness by eliminating the inaccurate portion of support estimates. By combining the two steps, we develop a novel sequential CS-MUSIC algorithm that is robust, even with a limited number of snapshots. Using theoretical noise analysis as well as numerical simulation, we show that the sequential CS-MUSIC is superior to the existing subspace-based greedy algorithms and exhibits similar performance behavior to the state-of-art algorithms with a significantly reduced computational complexity.

## II. GENERALIZED SUBSPACE CRITERION

### A. Notations and Mathematical Preliminaries

Throughout the paper,  $\mathbf{x}^i$  and  $\mathbf{x}_j$  correspond to the  $i$ -th row and  $j$ -th column of matrix  $X$ . When  $I$  is an index set,  $X^I$ ,  $A_I$  corresponds to a submatrix collecting corresponding rows of  $X$  and columns of  $A$ , respectively. The rows (or columns) in  $\mathbb{R}^n$  are in *general position* if any  $n$  collection of rows (or columns) are linearly independent.

*Definition 1 (Canonical form noiseless MMV [7]):*

Suppose 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  $\|X_*\|_0 = |\text{supp}X| = k$ , where  $m$ ,  $n$ , and  $r$  are positive integers ( $r \leq m < n$ ) that represent the number of sensor elements, an ambient space dimension, and the rank of an observation matrix, respectively. A canonical form noiseless multiple measurement vector (MMV) problem is given as an estimation problem of  $k$ -sparse vectors  $X \in \mathbb{R}^{n \times r}$  using the following formula:

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

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

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 [7]. Hence, in

this section, we assume that an MMV problem assumes the canonical form. However, this assumption will be relaxed later in noise analysis.

### B. Generalized Subspace Criterion

Note that the generalized MUSIC criterion in [7] requires  $0 \leq \delta_{2k-r+1}^L(A) < 1$ . This implies that, if a sensing matrix is obtained from a random Gaussian and if measurement is noiseless, then we have the following minimal sampling condition [7]:

$$m \geq (1 + \delta)(2k - r + 1) \text{ for some } \delta > 0.$$

If we have a redundant sampling  $m \gg 2k - r + 1$ , the following theorem can be used instead as the extension of the generalized MUSIC criterion in [7].

*Theorem 1:* Suppose  $1 \leq l \leq r$  and we have a canonical MMV model  $AX = B$  with a sensing matrix  $A$  that satisfies an RIP condition with  $0 \leq \delta_{2k-r+l}^L(A) < 1$ . Furthermore, suppose the nonzero rows of  $X$  are in general position. Then, for a given index set  $I \subset \{1, \dots, n\}$  such that  $|I| \leq \min(2(k-r) + l, k)$  and  $|I \setminus \text{supp}X| \leq k - r + l$ , the following statements are equivalent:

- (i)  $|I \cap \text{supp}X| \geq k - r + 1$ ;
- (ii)  $\text{rank}[A_I \ B] < |I| + r$ . (2)

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

$$\dim(R[A_{\tilde{I}} \ B]) \leq \dim(A_S) = k.$$

However,  $|\tilde{I}| + r = k + 1$  so that  $[A_{\tilde{I}} \ B]$  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 (2), there are  $\mathbf{p} \in \mathbb{R}^{|I|}$  and  $\mathbf{q} \in \mathbb{R}^r$  such that  $A_I \mathbf{p} + AX \mathbf{q} = \mathbf{0}$  and  $[\mathbf{p}, \mathbf{q}]^T \neq \mathbf{0}$ . If we let  $\tilde{\mathbf{p}} \in \mathbb{R}^n$  by  $\tilde{\mathbf{p}}^I = \mathbf{p}$  and  $\tilde{\mathbf{p}}^{I^c} = \mathbf{0}$ , we have  $A[\tilde{\mathbf{p}} + X\mathbf{q}] = \mathbf{0}$ . Since  $\|\tilde{\mathbf{p}} + X\mathbf{q}\|_0 \leq |I \setminus \text{supp}X| + |\text{supp}X| \leq 2k - r + l$ , by the RIP condition, we have  $\tilde{\mathbf{p}} + X\mathbf{q} = \mathbf{0}$  so that  $\text{supp}(\tilde{\mathbf{p}}) = \text{supp}(X\mathbf{q}) \subset \text{supp}X$ . If we assume that  $|I \cap \text{supp}X| \leq k - r$ , then  $\|\tilde{\mathbf{p}}\|_0 \leq k - r$  but  $\|X\mathbf{q}\|_0 \geq k - r + 1$  since the nonzero rows of  $X$  are in general position. This is impossible so that  $|I \cap \text{supp}X| \geq k - r + 1$ .  $\blacksquare$

Note that the conditions  $|I| \leq \min(2(k-r) + l, k)$  and  $|I \setminus \text{supp}X| \leq k - r + l$  in Theorem 1 do not imply that there is a unique index set  $I$ ; rather, Theorem 1 says that multiple index sets  $I$  can exist for a given  $l$ . For example, if  $l = 1$ , any index set  $I$  such that  $|I| = k - r + 1, \dots, \min(2(k-r) + 1, k)$  that satisfies the condition  $|I \setminus \text{supp}X| \leq k - r + l$ , can be used to test conditions (i)-(ii) in Theorem 1. Furthermore, if we choose  $|I| = k - r + 1$ , Theorem 1 is reduced to the following generalized MUSIC criterion in [7].

*Corollary 1 (Generalized MUSIC Criterion [7]):* Suppose we have a canonical MMV model  $AX = B$  with a

sensing matrix  $A$  that satisfies an RIP condition with  $0 \leq \delta_{2k-r+1}^L(A) < 1$ . Furthermore, suppose the nonzero rows of  $X$  are in general position. Then, for  $I_{k-r} \subset \text{supp}X$  with  $|I_{k-r}| = k-r$  and any  $j \in \{1, \dots, n\} \setminus I_{k-r}$ , we have  $j \in \text{supp}X$  if and only if

$$\text{rank}[A_{I_{k-r} \cup \{j\}} B] < k+1 \quad (3)$$

or equivalently

$$\mathbf{a}_j^* P_{R([A_{I_{k-r}} B])}^\perp \mathbf{a}_j = 0.$$

*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. (2) is equivalent to Eq. (3), which is equivalent to say  $\mathbf{a}_j \in R([A_{I_{k-r}} B])$  or  $\mathbf{a}_j^* P_{R([A_{I_{k-r}} B])}^\perp \mathbf{a}_j = 0$ . This concludes the proof. ■

*Remark 1:* If  $r = k$ , the conventional MUSIC criterion can be trivially derived.

*Remark 2:* The subspace  $R([A_{I_{k-r}} B])$  is called *augmented signal subspace*. This name was first coined in [12].

So far, we have shown that Theorem 1 can reproduce the existing results. However, one of the important byproducts of the theorem is the following form, which will be used extensively in the following sections.

*Corollary 2:* Suppose  $1 \leq l \leq r$  and we have a canonical MMV model  $AX = B$  with a sensing matrix  $A$  that satisfies an RIP condition with  $0 \leq \delta_{2k-r+l}^L(A) < 1$ . Furthermore, suppose the nonzero rows of  $X$  are in general position. Then, for an index set  $I \subset \{1, \dots, n\}$  such that  $|I| \leq \min(2(k-r)+l, k)$  and  $|I \setminus \text{supp}X| \leq k-r+l$ , if we have  $|I \cap \text{supp}X| = k-r+q$  for some  $q \geq 0$ , then

$$\text{rank}[A_I B] = |I| + r - q. \quad (4)$$

*Proof:* Take an  $I_{k-r} \subset (I \cap \text{supp}X)$  with  $|I_{k-r}| = k-r$  and let  $J_{k-r} = (I \cap \text{supp}X) \setminus I_{k-r}$ , where  $|J_{k-r}| = q$ . Then by Theorem 1, we have

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

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

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

since  $R([A_I B]) = R([A_{I \setminus J_{k-r}} B])$ . ■

### III. SEQUENTIAL COMPRESSIVE MUSIC ALGORITHM

By employing the results in the previous section, this section first develops forward or backward greedy steps. Then, by combining the two approaches, we can derive a novel *sequential CS-MUSIC* algorithm.

#### A. Forward Greedy: Sequential Subspace Estimation

In [7], the CS-MUSIC first determines  $k-r$  indices of  $\text{supp}X$  with CS-based algorithms such as 2-thresholding or S-OMP, and then it recovers the remaining  $r$  indices of  $\text{supp}X$  using the generalized MUSIC criterion. For this, a projection operator onto the noise subspace is calculated as the orthogonal complement of the augmented signal subspace  $R([A_{I_{k-r}} B])$ . However, the following result can further extend the existing generalized MUSIC criterion [7].

*Theorem 2:* Suppose  $1 \leq l \leq r$  and we have a canonical MMV model  $AX = B$  with a sensing matrix  $A$  that satisfies an RIP condition with  $0 \leq \delta_{2k-r+l}^L(A) < 1$ . Furthermore, suppose the nonzero rows of  $X$  are in general position. Then, if we have an index set  $I \subset \{1, \dots, n\}$  such that  $|I| \leq \min(2(k-r)+l-1, k-1)$ ,  $|I \setminus \text{supp}X| \leq k-r+l-1$  and  $|I \cap \text{supp}X| \geq k-r$ , we have for  $j \notin I$ ,  $j \in \text{supp}X$  if and only if

$$\text{rank}[A_I B] = \text{rank}[A_{I \cup \{j\}} B] \quad (5)$$

or equivalently

$$\mathbf{a}_j^* P_{R([A_I B])}^\perp \mathbf{a}_j = 0. \quad (6)$$

*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 Corollary 2 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, (5) is equivalent to  $\mathbf{a}_j \in R([A_I B])$ , which is also equivalent to (6). This completes the proof. ■

*Remark 3:* 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 implies that we first need to find  $I_{k-r}$  support using a compressive sensing algorithm, then we augment newly added supports into the initial estimate  $I_{k-r}$ . As will be shown later in noise analysis, such a greedy procedure improves the accuracy of the augmented signal subspace estimation.

*Remark 4:* The greedy procedure can even be performed in a critically sampled case, *i.e.*  $l = 1$ . In this case, we can augment atoms up to  $\min(2(k-r), k-1)$ , which is always bigger than adding only  $k-r$  atoms. However, the number of possible augmentation increases with a redundant sampling, which makes the algorithm more robust.

Theorem 2 leads us to the following sequential algorithm (*SeqSubspace*), as in Table I. Note that the algorithm can be combined with any joint sparse recovery algorithm that provides a  $k-r$  initial support estimate.

TABLE I

Algorithm: $I = \text{SeqSubspace}(A, B, I_{k-r})$
<ul style="list-style-type: none"> <li>- Set <math>q = 0</math> and <math>I = I_{k-r}</math>.</li> <li>- While <math>q &lt; r</math>, do the following procedure:               <ol style="list-style-type: none"> <li>1. Perform an SVD of <math>[A_I B] = [U_1, U_0] \text{diag}[\Sigma_1, \Sigma_0] [V_1, V_0]^*</math>, where <math>\Sigma_1 = \text{diag}[\sigma_1, \dots, \sigma_k]</math> and <math>\Sigma_0 = \text{diag}[\sigma_{k+1}, \dots, \sigma_{k+q}]</math> and <math>\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{k+q}</math>;</li> <li>2. Take <math>j_q = \arg \min_{j \notin I} \ P_{R(U_1)}^\perp \mathbf{a}_j\ ^2</math>.</li> <li>3. Set <math>I := I \cup \{j_q\}</math>, let <math>q := q + 1</math> and goto step 1.</li> </ol> </li> <li>- Return <math>I</math>.</li> </ul>

### B. Backward Greedy: Support Filtering

As discussed before, we can easily expect that the performance of the generalized MUSIC step is highly dependent on the selection of  $k-r$  correct indices of the support of  $X$ . Note that this is a very stringent condition. In practice, even though the first consecutive steps of, for example, S-OMP, may not provide all true partial supports, it is more likely that among a  $k$ -sparse support estimate of S-OMP, part of the supports (not in sequential order) can be correct. In fact, an information theoretical analysis of a partial support recovery condition in single measurement vector CS (SMV-CS) [21] showed that the required SNR condition of a partial support recovery is much more relaxed than that for a full support recovery. 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 correct indices from an estimate of  $\text{supp}X$ .

Indeed, our new algorithm requires that  $k-r+1$  supports (not in sequential order) out of a larger support estimate is correct. Then, the location of a correct  $k-r$  support can be readily estimated using the following backward support filtering. Compared to a forward greedy procedure that improves the accuracy of the signal subspace estimation, the backward support filtering criterion can improve the accuracy of a partial support recovery, and, hence, the corresponding accuracy of an augmented signal subspace.

*Theorem 3 (Backward support filtering criterion):*

Suppose  $1 \leq l \leq r$  and we have a canonical MMV model  $AX = B$  with a sensing matrix  $A$  that satisfies an RIP condition with  $0 \leq \delta_{2k-r+l}^L(A) < 1$ . Furthermore, suppose the nonzero rows of  $X$  are in general position. Then, if we have an index set  $I \subset \{1, \dots, n\}$  such that  $|I| \leq \min(2(k-r) + l, k)$ ,  $|I \setminus \text{supp}X| \leq k-r+l$  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

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

*Proof:* Assume that  $|I \cap \text{supp}X| = k-r+q$ , where  $q \geq 1$ . Then, by Corollary 2, we have  $\text{rank}[A_I B] = |I|+r-q$ . Noting that  $I \setminus \{j\}$  satisfies the assumptions of Corollary 2 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  $\mathbf{a}_j \in R([A_{I \setminus \{j\}} B])$  if and only if  $j \in \text{supp}X$ . Hence  $\mathbf{a}_j^* P_{R([A_{I \setminus \{j\}} B])}^\perp \mathbf{a}_j = 0$  if and only if  $j \in \text{supp}X$ . That completes the proof. ■

Theorem 3 informs us 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 the backward support filtering criterion as described in Table II.

TABLE II

Algorithm: $I_{k-r} = \text{SupportFiltering}(A, B, I)$
<ul style="list-style-type: none"> <li>- For all <math>j \in I</math>, calculate the quantities <math>\zeta(j) := \ P_{R([A_{I \setminus \{j\}} B])}^\perp \mathbf{a}_j\ ^2</math>.</li> <li>- Making an ascending ordering of <math>\zeta(j)</math> for <math>j \in I</math>, choose indices that correspond to the first <math>k-r</math> indices and put these indices into <math>I_{k-r}</math>.</li> <li>- Return <math>I_{k-r}</math>.</li> </ul>

*Remark 5:* Due to the condition  $|I| \leq \min(2(k-r) + l, k)$  in Theorem 3, we can include  $k$ -sparse support estimate  $I$  in a support filtering step if  $r < (k+l)/2$ . Note that this is always true regardless of  $r$  if  $l = k$  or  $\delta_{2k}^L < 1$ . However, if  $r \geq (k+l)/2$ , we can use the following heuristics. First, just include the first  $2(k-r)+l$  support estimate of  $I$  for a support filtering. Since, in most greedy algorithms, the earlier greedy steps are more likely to succeed, correct  $k-r+1$  supports are more likely to be included. Hence, we can filter out the remaining indices  $j \in I$  such that  $j \notin \text{supp}X$ .

### C. Sequential CS-MUSIC

By combining the forward and the backward greedy steps, this paper develops the following sequential CS-MUSIC algorithm described in Table III. Note that this algorithm assumes that the sparsity level  $k$  is given as *a priori* knowledge. (The estimation problem of an unknown  $k$  will be discussed later.)

## IV. NOISY PERFORMANCE ANALYSIS OF SEQUENTIAL CS-MUSIC

### A. Improving Noise Robustness Using Sequential Subspace Estimation

In practice, measurements are noisy, so the theory we have derived for noiseless measurements should be modified. Suppose a noisy MMV model is given by:

$$Y = AX + W,$$

TABLE III

Algorithm: $I_k = \text{SeqCSMUSIC}(A, Y, k, r)$
Input: $k, r, A \in \mathbb{R}^{m \times n}, Y \in \mathbb{R}^{m \times N}$
Output: $k$ support estimate $I_k$
- Estimate the $k$ support estimate $I_k$ of $\text{supp}X$ using any MMV algorithm.
- $U := \text{Rank-}r$ signal subspace estimate of $R(Y)$ .
- $I_{k-r} := \text{SupportFiltering}(A, U, I_k)$ .
- $I_k := \text{SeqSubspace}(A, U, I_{k-r})$ .
- Return $I_k$ .

where  $Y \in \mathbb{R}^{m \times N}$  are noisy measurements corrupted by an additive noise  $W \in \mathbb{R}^{m \times N}$ , and  $N$  denotes the number of snapshots. Then, using singular value decomposition, we can find the following canonical MMV problem:

$$\tilde{S} = A\tilde{X} + \tilde{W},$$

where  $\tilde{S} \in \mathbb{R}^{m \times r}$  is the rank- $r$  signal subspace estimate of  $Y$  and  $r \leq N$  denotes the numerical rank of  $Y$ . Due to the noise,  $\tilde{S}$  is perturbed from the noiseless signal subspace  $S$  such that  $R(\tilde{S}) = R(A\tilde{X})$ , which leads to errors in the augmented signal subspace. The following theorem characterizes how much perturbation in an augmented signal subspace can be endured by a generalized MUSIC step.

*Theorem 4:* For  $0 \leq l < r$ , if we have  $I_{k-r+l} \subset \text{supp}X$  such that  $|I_{k-r+l}| = k-r+l$  and singular value decomposition of  $[A_{I_{k-r}} \ S]$  and  $[A_{I_{k-r+l}} \ \tilde{S}]$  as

$$[A_{I_{k-r}} \ S] = U_1 \Sigma_1 V_1, \quad [A_{I_{k-r+l}} \ \tilde{S}] = \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^* + \tilde{U}_0 \tilde{\Sigma}_0 \tilde{V}_0^*,$$

where  $\tilde{\Sigma}_1 = \text{diag}[\tilde{\sigma}_1, \dots, \tilde{\sigma}_k]$ ,  $\tilde{\Sigma}_0 = \text{diag}[\tilde{\sigma}_{k+1}, \dots, \tilde{\sigma}_{k+l}]$  and  $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_{k+l}$ , then, for any  $j \notin \text{supp}X$  and  $q \in \text{supp}X$  we have

$$\|P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j\|^2 > \|P_{R(\tilde{U}_1)}^\perp \mathbf{a}_q\|^2$$

provided that

$$\|P_{R(\tilde{U}_1)} - P_{R(U_1)}\| < 1 - \gamma. \quad (7)$$

and  $m, n \rightarrow \infty$  and  $\gamma = \lim_{n \rightarrow \infty} k/m < 1$ . In other words, a generalized MUSIC step finds correct supports if Eq. (7) is satisfied.

*Proof:* Noting that  $\|P_{R(U_1)}^\perp \mathbf{a}_j\|^2 = 0$  for  $j \in \text{supp}X$  by the generalized MUSIC criterion, for any  $j \notin \text{supp}X$  and  $q \in \text{supp}X$ , we have

$$\begin{aligned} & \|P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j\|^2 - \|P_{R(\tilde{U}_1)}^\perp \mathbf{a}_q\|^2 \\ &= \mathbf{a}_j^* P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j - \mathbf{a}_q^* P_{R(\tilde{U}_1)}^\perp \mathbf{a}_q - \mathbf{a}_q^* [P_{R(\tilde{U}_1)} - P_{R(U_1)}] \mathbf{a}_q \\ &= \mathbf{a}_j^* P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j - \|P_{R(\tilde{U}_1)} - P_{R(U_1)}\| \|\mathbf{a}_q\|^2. \end{aligned} \quad (8)$$

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

[22],  $\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} \|\mathbf{a}_j\|^2/m = 1$  since  $\lim_{n \rightarrow \infty} (\log n)/m = 0$ . Since  $\lim_{n \rightarrow \infty} (\log(n-k))/(m-k) = 0$ , by Lemma 3 in [22], we have  $\lim_{n \rightarrow \infty} \min_{j \notin \text{supp}X} m \mathbf{a}_j^* P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j / (m-k) = 1$  so that  $\lim_{n \rightarrow \infty} \min_{j \notin \text{supp}X} \mathbf{a}_j^* P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j / \max_{1 \leq j \leq n} \|\mathbf{a}_j\|^2 = 1 - \gamma$ . Hence, Eq. (8) is positive provided that Eq. (7) holds in the large system limit. This completes the proof. ■

Therefore, by minimizing the perturbation in the augmented signal subspace  $\|P_{R(\tilde{U}_1)} - P_{R(U_1)}\|$ , we can make the generalized MUSIC step more robust. Unfortunately, the direct minimization of the perturbation of the subspace is not easy. Instead, we are interested in minimizing the following upper-bound of the perturbation, whose proof can be found in Appendix A:

$$\|P_{R(\tilde{U}_1)} - P_{R(U_1)}\| \leq \frac{\Delta}{\sigma_k([A_{I_{k-r+l}} \ S]) - \Delta}, \quad (9)$$

where  $\Delta = \|S - \tilde{S}\|$  and  $\sigma_k([A_{I_{k-r+l}} \ S])$  denotes the  $k$ -th largest singular value of  $[A_{I_{k-r+l}} \ S]$ . Then, we have the following theorem:

*Theorem 5:* Let  $l \geq 0$ . For  $I_{k-r+l} \subset \text{supp}X$  such that  $|I_{k-r+l}| = k-r+l$ , the generalized MUSIC steps find the remaining  $r-l$  support provided that

$$\frac{\sigma_k([A_{I_{k-r+l}} \ S])}{\Delta} > 1 + \frac{1}{1 - \gamma} \quad (10)$$

and  $m, n \rightarrow \infty$  and  $\gamma = \lim_{n \rightarrow \infty} k/m < 1$ .

*Proof:* This can be trivially proven by pugging Eq. (9) in the inequality (7). ■

Note that for  $I_{k-r+l} \subset \text{supp}X$ , we have  $\text{rank}([A_{I_{k-r+l}} \ S]) = k$  so that the set of columns of  $[A_{I_{k-r+l}} \ S]$  is a frame in  $R(A_{\text{supp}X})$  with lower frame bound  $\sigma_k^2([A_{I_{k-r+l}} \ S])$ . In this case, as shown in Fig. 1,  $\sigma_k([A_{I_{k-r+l}} \ B])$  is an increasing function of  $l$ , so as  $l$  increases, the frame becomes more redundant and the lower frame bound become larger. Hence, the left side of Eq. (10) becomes larger.

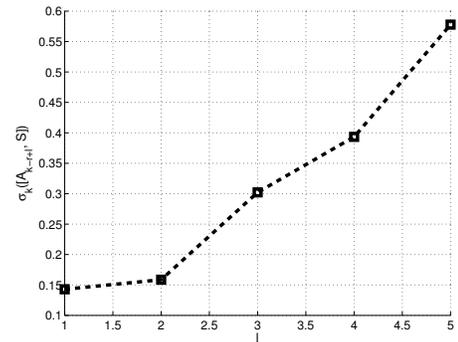


Fig. 1.  $\sigma_k([A_{k-r+l}, S])$  values with increasing  $l$ . Simulation parameters are  $n = 128, k = 8, r = 6$ .

This observation provides us an important error correction scheme. Note that the SNR condition Eq. (10) is still the same

even if we find a support index  $j_l$  in a greedy manner as follows:

$$j_l = \arg \min_{j \notin I_{k-r+l}} \|P_{R(\tilde{U}_1)}^\perp \mathbf{a}_j\|^2. \quad (11)$$

However, as SNR condition Eq. (10) is a sufficient condition, a non-zero probability of  $j_l$  being in the true support exists even though Eq. (10) is not satisfied. (This is especially true if we select only one index rather than choosing all  $r-l$  indices). If a *correctly* found index  $j_l$  is augmented for the next step of sequential subspace estimation, then it is more likely that the condition in Eq. (10) can be satisfied in the following greedy steps since the left side term of Eq. (10) is an increasing function of  $l$  thanks to the inclusion of a correct index  $j_l$ . As soon as an SNR condition is satisfied, the remaining greedy steps will succeed since the condition is sufficient. Therefore, even when a sufficient SNR condition is not satisfied initially, the proposed sequential subspace estimation technique exploits the possibility of finding a correct index to improve the noise robustness, which was not possible in an original MUSIC step.

### B. Improving Noise Robustness Using Support Filtering

Using similar techniques, we can derive the following sufficient condition for the success of support filtering.

*Theorem 6:* Let  $m \geq k+r$ . Suppose we have an index set  $I \subset \{1, \dots, n\}$  such that  $|I| = k$ ,  $|I_C| \geq k-r+1$ , where  $I_C := (\text{supp}X) \cap I$ . Then, we have

$$\min_{j \notin \text{supp}X} \|P_{R[A_{I \setminus \{j\}}]}^\perp \tilde{s}\| \|\mathbf{a}_j\|^2 > \max_{j \in \{q_1, \dots, q_{k-r}\}} \|P_{R[A_{I \setminus \{j\}}]}^\perp \tilde{s}\| \|\mathbf{a}_j\|^2 \quad (12)$$

provided that

$$\frac{\tilde{\sigma}_k(q_{k-r}, I)}{\Delta} > 1 + \frac{1}{1 - \gamma(1 + \alpha)}, \quad (13)$$

and  $m, n \rightarrow \infty$ , where  $\tilde{\sigma}_k(q, I) = \max\{\sigma_k([A_{I_T} \ S]) : T \subset I_C \setminus \{q\}, |T| = k-r\}$  for  $q \in I_C$  and  $I_C = \{q_1, \dots, q_{|I_C|}\}$ , which satisfies  $\tilde{\sigma}_k(q_1, I) \geq \tilde{\sigma}_k(q_2, I) \geq \dots \geq \tilde{\sigma}_k(q_{|I_C|}, I)$ , and  $\gamma := \lim_{n \rightarrow \infty} k/m$ ,  $\alpha := \lim_{n \rightarrow \infty} r/k$ .

*Proof:* See Appendix B. ■

In Theorem 6,  $\tilde{\sigma}_k(q_{k-r}, I)$  increases when an initial support estimation has more correct support because of the equation in which  $\tilde{\sigma}_k(q, I)$  is given by the maximum value out of  $\binom{|I_C|}{k-r}$  possibilities. Moreover, if we increase the ratio  $m/(k+r)$ , then the right-hand side of (13) decreases so that we can expect a greater possibility of accurate support filtering with an increased redundant number of samples than the critical sampling rate.

To confirm a support filtering useful for performance improvement, we examine the cases where the sufficient condition for an initial support estimation is less favorable than that of a support filtering. Characterization of such cases should be done with respect to a particular initial support estimation algorithm. For example, in the case of a subspace S-OMP

for an initial support estimation, a sufficient condition for the success of subspace S-OMP for  $\alpha > 0$  is given by [7]:

$$\|P_{R(\tilde{U}_1)} - P_{R(U_1)}\| < \frac{\alpha - \alpha\sqrt{\gamma}(2 - F(\alpha))}{2}, \quad (14)$$

where  $F(\alpha)$  is an increasing function such that  $F(1) = 1$  and  $\lim_{\alpha \rightarrow 0^+} F(\alpha) = 0$ , which is defined as  $F(\alpha) = \frac{1}{\alpha} \int_0^{4t_1(\alpha)} x d\lambda_1(x)$ ,  $d\lambda_1(x) = (\sqrt{(4-x)x})/(2\pi x) dx$  is the probability measure with support  $[0, 4]$ ,  $0 \leq t_1(\alpha) \leq 1$  satisfies  $\int_0^{2t_1(\alpha)} ds_1(x) = \alpha$ , and  $ds_1(x) = (1/\pi)\sqrt{4-x^2} dx$  is a probability measure with support  $[0, 2]$ . Now, the SNR condition for support filtering in Eq. (13) can be translated into a threshold of the allowable augmented subspace perturbation of  $1 - \gamma(1 + \alpha)$ , when  $m \geq r+k$ . Hence, the gap between the two bounds is given by

$$f(\gamma, \alpha) = \frac{\alpha - \alpha\sqrt{\gamma}(2 - F(\alpha))}{2} - (1 - \gamma(1 + \alpha)).$$

Fig. 2 characterizes the function  $f(\gamma, \alpha)$ . In region A,  $f(\gamma, \alpha) < 0$  and  $m > r+k$ , hence, the support filtering has more noise robustness and can correct errors from subspace S-OMP. As we can see from Fig. 2, support filtering is effective in most of the practical sampling rate, and especially when we have redundant samples or  $\text{rank}(Y)$  is relatively small. The larger size of region A that favors support filtering again confirms that support filtering is a quite useful technique to improve noise robustness.

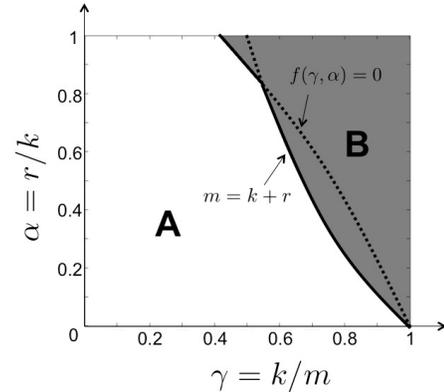


Fig. 2. Feasibility region where support filtering becomes more effective than subspace S-OMP.

## V. NUMERICAL RESULTS

In this section, we perform extensive numerical experiments to validate the proposed algorithm under various experimental conditions, and compare it with respect to existing joint sparse recovery algorithms.

### A. Dependency on Snapshot Number

First, we demonstrate that a sequential CS-MUSIC is less sensitive to the number of snapshots. The simulation parameters were as follows:  $m \in \{1, 2, \dots, 30\}$ ,  $n = 128$ ,  $k =$

$8, r = 4$ , and  $N \in \{6, 16, 256\}$ , respectively. The elements of a sensing matrix  $A$  were generated from a Gaussian distribution having zero mean and variance of  $1/m$ , and then each column of  $A$  was normalized to have an unit norm. An unknown signal  $X$  with  $\text{rank}(X) = r \leq k$  was generated using the same procedure as in [12]. Specifically, we randomly generated a support  $I$ , and then the corresponding nonzero signal components were obtained by

$$X^I = \Psi \Lambda \Phi, \quad (15)$$

where  $\Psi \in \mathbb{R}^{k \times r}$  and  $\Lambda$  were set to random orthonormal columns and the identity matrix, respectively, and  $\Phi \in \mathbb{R}^{r \times N}$  were made using Gaussian random distribution with zero mean and variance of  $1/N$ . After generating noiseless data, we added zero mean white Gaussian noise to have  $SNR = 30dB$  measurements. We declared success if an estimated support was the same as a true  $\text{supp}X$ , and success rates were averaged over 1000 experiments.

Fig. 3 shows success rates of a sequential CS-MUSIC compared to that of CS-MUSIC or SA-MUSIC. Since SA-MUSIC in [12] is equivalent to CS-MUSIC for a normalized  $A$  matrix, the original code of SA-MUSIC was used for fair comparison. As shown in Fig 3, sequential CS-MUSIC exhibits nearly similar recovery performance for various snapshot numbers, whereas the original form of CS-MUSIC/SA-MUSIC requires a large number of snapshots to achieve maximum performance.

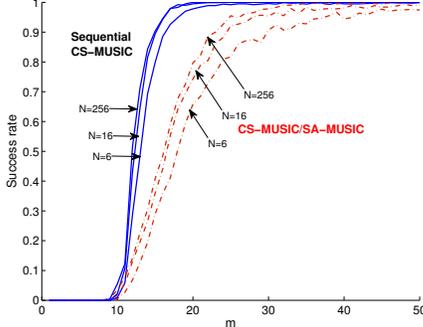


Fig. 3. Snapshot dependent performance behaviour of the sequential CS-MUSIC and the original CS-MUSIC/SA-MUSIC. The simulation parameters are  $n = 128, k = 8, r = 4$  and  $SNR = 30dB$ .

In order to identify the contribution of the forward and backward greedy steps in the performance improvement, we perform additional experiments using the same simulation setup. Fig. 4 illustrates the performances of sequential CS-MUSIC, a variation of sequential CS-MUSIC without backward support filtering, and the original CS-MUSIC/SA-MUSIC algorithm for  $N = 6, 16$ , respectively. Here, an initial  $k - r$  support for CS-MUSIC/SA-MUSIC and the sequential CS-MUSIC were estimated using an identical subspace S-OMP algorithm in [7], [12] so that performance differences came only from the sequential subspace estimation step. For a bigger  $N$  where

the signal subspace error is small, performance improvement due to the sequential subspace estimation was not remarkable. However, the advantages of sequential subspace estimation is especially noticeable for a small number of snapshots where a subspace estimation is prone to error. On the other hand, the backward support filtering is beneficial for all ranges of snapshots since it corrects the contribution of a partial support estimation error in a subspace S-OMP step.

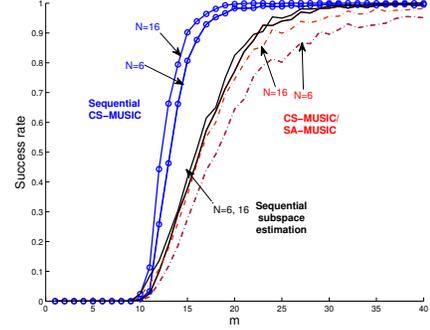


Fig. 4. Sequential CS-MUSIC, Sequential CS-MUSIC without greedy subspace estimation, and CS-MUSIC/SA-MUSIC performance for various snapshot numbers. Simulation parameters are  $n = 128, k = 8, r = 4$  and  $SNR = 30dB$ .

## B. Performance Comparison with State-of-Art Joint Sparse Recovery Algorithms

To compare the proposed algorithm with various state-of-art joint sparse recovery methods, the recovery rates of various joint sparse recovery algorithms such as CS-MUSIC/SA-MUSIC, S-OMP, MUSIC, and LIKES [18] are plotted in Fig. 5 along with those of a sequential CS-MUSIC. LIKES is at least as accurate as M-SBL[17] but it is faster than M-SBL and therefore it is more competitive [18], so we use LIKES for comparison instead of M-SBL. We do not include the results of mixed norm approach since the mixed norm approach was not sufficiently competitive to LIKES. For LIKES implementation, we used the original implementation provided by the original authors in [18]. Since the LIKES do not provide a exact  $k$ -sparse solution, we used the support for the largest  $k$  coefficients as a support estimate in calculating the perfect recovery ratio. Figs. 5(a) and (b) show the recovery rates for  $N = 8$  and  $256$ , respectively. Sequential CS-MUSIC outperforms S-OMP and the original CS-MUSIC/SA-MUSIC consistently, and its performance nearly achieves that of LIKES. Indeed, the additional sampling cost for a sequential CS-MUSIC compared to LIKES is very small. Considering that any subspace method needs additional redundancy (i.e.  $m \geq k + 1$ ) to avoid ambiguity in the signal subspace estimation, we believe that sequential CS-MUSIC nearly achieves the optimum performance. Furthermore, this high performance can be achieved at smaller computational complexity. Note that the complexity of the sequential CS-MUSIC is smaller than

LIKES and the complexity advantage is more noticeable for large snapshot numbers, as shown in Figs. 5(c)(d) for  $N = 8$  and 256, respectively. For the stopping criterion of LIKES, we employed the same method developed in the original implementation [18]. Note that the increased computational complexity of LIKES for small  $m$  is originated from increased iteration number to satisfy the stopping criterion for the case of insufficient number of measurements.

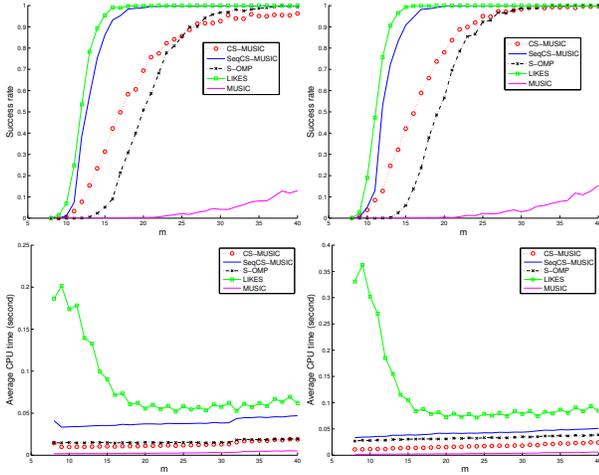


Fig. 5. Performance of various joint sparse recovery algorithm at (a)  $N = 8$  and (b)  $N = 256$ , when  $n = 128, k = 8, r = 4$ , and  $SNR = 30dB$ . (c)(d) Average CPU time for  $N = 8$  and  $N = 256$ , respectively.

To show the dependency of recovery performance on the condition number of  $X$ , we conducted simulations for two different types of  $X$ . More specifically, the  $j$ -th diagonal term of  $\Lambda$  in Eq. (15) is given by  $\sigma_j = \tau^{j-1}$  for  $j = 1, \dots, r$ . The results in Fig. 6(a) provide evidence that sequential CS-MUSIC is not greatly affected by the condition number of  $X$ , and appears less sensitive than LIKES. Next, we performed simulation studies for the different types of RIP conditions using various MMV algorithms. More specifically, we assumed that each component of a sensing matrix follows  $\mathcal{N}(a, 1/m)$  and then normalized each column of  $A$  to have a unit norm. The mean values are set to  $a = 0$  and 1, where a larger  $a$  represents a worse RIP condition. In this simulation,  $N = 64$  and the other parameters are the same as before. Fig. 6(b) shows that sequential CS-MUSIC is more robust than the original CS-MUSIC/SA-MUSIC for unfavorable RIP conditions. However, compared to LIKES, the sequential CS-MUSIC appears less robust to unfavorable RIP conditions, which is commonly observed in most of the greedy approaches.

### C. Fourier Sensing Matrix

Finally, we conducted similar numerical experiments using the Fourier sensing matrix. In this case, the source model in Eq. (15) is set to be complex valued. Fig. 7 illustrates the recovery performance of various MMV algorithms for the

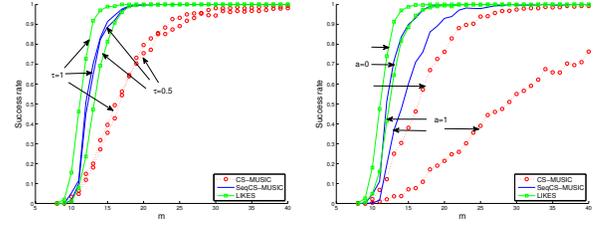


Fig. 6. Recovery rates of LIKES, sequential CS-MUSIC and CS-MUSIC/SA-MUSIC for (a) different condition numbers and (b) various RIP condition of  $A$  (larger mean value provides worse RIP condition). The simulation parameters were  $n = 128, k = 8, r = 4, N = 64$  and  $SNR = 30dB$ .

Fourier sensing matrix when  $n = 128, k = 8$ , and  $r = 4$  at the SNR of 30dB for  $N = 5$ . We again observed similar performance improvement as in the Gaussian matrix. Note that computation time gain of the proposed method over LIKES is more noticeable in Fourier sensing matrix cases.

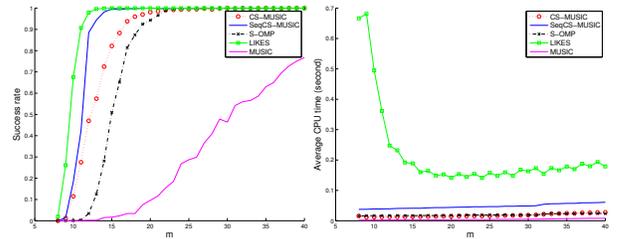


Fig. 7. (a) Performance of various joint sparse recovery algorithm from Fourier sensing matrix and (b) its computational time, when  $n = 128, k = 8, r = 4, N = 5$  and  $SNR = 30dB$ .

## VI. DISCUSSION

### A. Relation to Prior Constrained MUSIC and Sequential MUSIC

In a prior constrained MUSIC algorithm [15], the prior knowledge of a direction-of-arrival (DOA) is incorporated into the MUSIC criterion to improve estimation performance by filtering out the known source directions via orthogonal projections. If the prior knowledge of a partial support is exact, the prior constrained MUSIC is closely related to the generalized MUSIC step in CS-MUSIC/SA-MUSIC with an exact partial support estimate. However, as shown in [16] as well as in this paper, such an algorithm is affected by the resulting perturbation of the augmented subspace if the number of snapshots or SNR is not sufficiently high or the partial support knowledge is erroneous. There have been several approaches to improve the noise robustness of a prior constrained MUSIC (see [16] and references therein); however, to the best of our knowledge, we are not aware of any existing method that improves the robustness of an augmented signal subspace estimate using sequential subspace estimation and support filtering.

Sequential MUSIC and its variations [23], [24], [25], [26] may appear closely related to the proposed method. Indeed, Davies *et al.* [27] showed that the Recursively Applied and Projected (RAP)-MUSIC [26] is equivalent to a subspace S-OMP (SS-OMP) step - a partial support recovery estimation part of CS-MUSIC. However, as shown in [7], the SNR requirement of a subspace S-OMP is much tighter than that of a generalized MUSIC step. Therefore, switching from sequential MUSIC to the generalized MUSIC step would be beneficial. Moreover, our analysis in this paper showed that the sequential subspace estimation further relaxes the SNR condition sequentially. This implies that even if the SNR condition of the generalized MUSIC is not satisfied initially, during the sequential subspace estimation the SNR condition can be met and the overall recovery performance can be improved. To the best of our knowledge, this type of techniques have not been reported for any existing sequential MUSIC algorithms [23], [24], [25], [26].

### B. Sparsity Estimation

So far, our derivation assumes the prior knowledge of support size. In our previous work [7], we derived a sparsity estimation algorithm. Note that this algorithm can be incorporated at each greedy step to make the algorithm work, even without knowing the sparsity level *a priori*. In addition, there are various heuristics that could be used to estimate the sparsity in MUSIC type parametric methods. However, they have a nonzero probability of being wrong if the measurement is noisy. Typically, these sparsity estimation algorithms tend to overestimate, so various model order selection criteria have been often incorporated to avoid this overestimation [28].

Note that the RIP condition  $\delta_{2k-r+l}^L < 1$  implies that the maximal sparsity level that our algorithm can recover does not exceed  $m - 1$ , and the corresponding submatrix  $A_{\hat{I}}$  for the support estimate  $\hat{I}$  is always full column ranked. This implies that, as long as the condition number of  $A_{\hat{I}}$  is not bad and the noise levels are sufficiently small, we can implement thresholding techniques in a reconstruction domain to find a sparse signal, similar to other non-parametric sparse recovery approaches like iterative thresholding, M-SBL, etc. When such thresholding scheme may not be sufficiently accurate, the current technique has limitations and we need a new way to estimate the sparsity level. Though the sparsity estimation is very important topic, this is beyond scope of the current work, and will be reported elsewhere.

### C. Limitation of Noisy Analysis

Even though our noisy analysis provides useful insight on the origin of the noise robustness of a sequential CS-MUSIC, current analysis has two limitations. First, the analysis is based on the Gaussian sensing matrix using an asymptotic argument. Hence, the analysis should be modified for a general

sensing matrix such as Fourier. An RIP based analysis in SA-MUSIC [12] would work toward this goal. Second, the noisy performance analysis is based on comparing sufficient conditions. Since a sufficient condition is often more restricted than necessary, the analysis in this paper should be understood as a more conservative comparison. Finally, note that Theorem 4 is not practical to verify because its assumption in Eq. (7) is difficult to check. The same difficulty exists in applying Theorem 5 where  $\Delta$  in Eq. (10) is difficult to calculate in practise. However, Theorem 4 and Theorem 5 still provide us useful mathematical insight why sequential subspace estimation can provide noise robustness in the proposed method.

## VII. CONCLUSION

In this paper, we derived two greedy strategies to improve the noise robustness of recent hybrid joint sparse recovery algorithms such as CS-MUSIC and SA-MUSIC. Although these hybrid algorithms significantly outperform any other conventional greedy MMV algorithms, the performance improvement is reduced for a limited number of snapshots. We showed that the performance degradation is due to a perturbation in an augmented signal subspace estimation originating from an inaccurate subspace or partial support estimation. Furthermore, we demonstrated that even with limited number of snapshots, there are two different ways to improve the noise robustness of augmented signal subspace estimation: one by sequential subspace estimation and the other by filtering out incorrect support. We further explained that the two greedy steps are byproducts of a novel generalized subspace criterion. Theoretical analysis in noisy situations revealed the origins of the noise robustness of the proposed algorithm and led to the identification of sampling conditions where each greedy step becomes beneficial. Extensive numerical simulation demonstrated that the new algorithm consistently outperforms the existing greedy algorithms and nearly achieves optimal performance with minimal computational complexity.

### APPENDIX A

To obtain the perturbation bound Eq. (9) in sequential subspace estimation, we use the following theorem.

*Theorem 7:* [29] Assume that  $G \in \mathbb{R}^{m \times q}$  has the singular value decomposition

$$G = U\Sigma V^* = U_1\Sigma_1V_1^* + U_0\Sigma_0V_0^* = G_1 + G_0$$

where  $G_1 := U_1\Sigma_1V_1^*$  and  $G_0 := U_0\Sigma_0V_0^*$ . 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,

$$\sin \theta(R(\tilde{G}_1), R(G_1)) = \|P_{R(\tilde{G}_1)} - P_{R(G_1)}\| \leq \frac{\epsilon}{\delta},$$

where

$$\begin{aligned} \epsilon &:= \max(\|R_1\|, \|R_2\|) \\ R_1 &:= -(\tilde{G} - G)V_1, R_2 := -(\tilde{G} - G)^*U_1. \end{aligned}$$

(Proof of Eq. (9)) For a noiseless measurement  $[A_{I_{k-r+l}} S]$ , we have

$$\sigma_1 \geq \dots \geq \sigma_k > \sigma_{k+1} = \dots = \sigma_{k+l} = 0$$

so that we have  $\sigma_{\min}(\tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^*) \geq \sigma_k([A_{I_{k-r+l}} S]) - \|S - \tilde{S}\|$  and  $\sigma_{\max}(U_0 \Sigma_0 V_0^*) = 0$  so that if we have  $\Delta := \|P_{R(S)} - P_{R(\tilde{S})}\| < \sigma_k([A_{I_{k-r+l}} S])$ , we can apply Theorem 7. Then, we have

$$\begin{aligned} \frac{\epsilon}{\delta} &= \frac{\max(\|(S - \tilde{S})V_1\|, \|(S - \tilde{S})^*U_1\|)}{\sigma_k([A_{I_{k-r+l}} B]) - \|S - \tilde{S}\|} \\ &\leq \frac{\Delta}{\sigma_k([A_{I_{k-r+l}} S]) - \Delta} \end{aligned}$$

if  $\Delta < \sigma_k([A_{I_{k-r+l}} S])$ . This concludes the proof.

## APPENDIX B

By the assumption  $|I \cap \text{supp}X| \geq k - r + 1$  and the generalized MUSIC criterion, we have  $\|P_{R([A_{I \setminus \{j\}} S])}^\perp \mathbf{a}_j\|^2 = 0$  for any  $j \in \text{supp}X$ . Then, for any  $j \notin \text{supp}X$  and  $q \in \text{supp}X$ , we have

$$\begin{aligned} &\|P_{R([A_{I \setminus \{j\}} \tilde{S}])}^\perp \mathbf{a}_j\|^2 - \|P_{R([A_{I \setminus \{q\}} \tilde{S}])}^\perp \mathbf{a}_q\|^2 \\ &\geq \|P_{R([A_{I \setminus \{j\}} \tilde{S}])}^\perp \mathbf{a}_j\|^2 - \|P_{R([A_{I_{k-r,q}} \tilde{S}])}^\perp \mathbf{a}_q\|^2 \\ &= \|P_{R([A_{I \setminus \{j\}} \tilde{S}])}^\perp \mathbf{a}_j\|^2 - \|P_{R([A_{I_{k-r,q}} \tilde{S}])}^\perp \mathbf{a}_q\|^2 \\ &\quad + \|P_{R([A_{I_{k-r,q}} S])}^\perp \mathbf{a}_q\|^2 \\ &\geq \|P_{R([A_{I \setminus \{j\}} \tilde{S}])}^\perp \mathbf{a}_j\|^2 \\ &\quad - \|\mathbf{a}_q\|^2 \|P_{R([A_{I_{k-r,q}} \tilde{S}])} - P_{R([A_{I_{k-r,q}} S])}\|, \end{aligned} \tag{16}$$

since  $P^2 = P$  for any orthogonal projection operator  $P$ , where  $I_{k-r,q} \subset (I \setminus \{q\}) \cap \text{supp}X$ . For  $j \notin \text{supp}X$ ,  $\mathbf{a}_j$  is statistically independent from  $R([A_{I \setminus \{j\}} \tilde{S}])$  so that  $m \|P_{R([A_{I \setminus \{j\}} \tilde{S}])}^\perp \mathbf{a}_j\|^2$  is chi-squared random variable with at least  $m - k - r + 1$  degrees of freedom so that

$$\liminf_{n \rightarrow \infty} \min_{j \notin \text{supp}X} \mathbf{a}_j^* P_{R([A_{I \setminus \{j\}} \tilde{S}])}^\perp \mathbf{a}_j \geq 1 - \gamma(1 + \alpha), \tag{17}$$

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

On the other hand, for any  $q \in \text{supp}X$ ,  $m \|\mathbf{a}_q\|^2$  is a chi-squared random variable of  $m$  degrees of freedom, so that by Lemma 3 in [22] we have  $\lim_{n \rightarrow \infty} \max_{q \in \text{supp}X} \|\mathbf{a}_q\|^2 = 1$ . Since

$[A_{I_{k-r,q}} S]$  has a full column rank, using the bound in Eq. (9) for any  $I_{k-r,q} \subset (\text{supp}X \cap I) \setminus \{q\}$ , we have

$$\begin{aligned} &\|P_{R([A_{I_{k-r,q}} S])} - P_{R([A_{I_{k-r,q}} \tilde{S}])}\| \\ &\leq \frac{\Delta}{\max_{T \subset (\text{supp}X \cap I) \setminus \{q\}} \sigma_k([A_{I_T} S]) - \Delta}. \end{aligned}$$

If we let  $\tilde{\sigma}(q, I) = \max\{\sigma_k([A_{I_T} S]) : T \subset I_C \setminus \{q\}, |T| = k - r\}$ , where  $I_C = \{q_1, \dots, q_{|I_C|}\}$  and

$$\tilde{\sigma}(q_1, I) \geq \tilde{\sigma}(q_2, I) \geq \dots \geq \tilde{\sigma}(q_{|I_C|}, I),$$

we have

$$\|P_{R([A_{I_{k-r,q}} S])} - P_{R([A_{I_{k-r,q}} \tilde{S}])}\| \leq \frac{\Delta}{\tilde{\sigma}_k(q_{k-r}, I) - \Delta} \tag{18}$$

for  $q = q_1, \dots, q_{k-r}$ . Hence, by (16), (17) and (18), (12) holds if we have

$$1 - \gamma(1 + \alpha) - \frac{\Delta}{\tilde{\sigma}_k(q_{k-r}, I) - \Delta} > 0$$

in the large system limit. This completes the proof.

## ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MEST) (No.2011-0030933). The authors would like to thank Kiryung Lee and Yoram Bresler at UIUC for helpful discussions and providing an SA-MUSIC code, and thank P. Babu at Uppsala University for providing a LIKES code.

## REFERENCES

- [1] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Proc. Magazine*, pp. 67–94, July 1996.
- [2] K. P. Pruessmann, M. Weigher, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.
- [3] A. Joshi, W. Bangerth, and E. M. Sevick-Muraca, "Non-contact fluorescence optical tomography with scanning patterned illumination," *Optics Express*, vol. 19, no. 14, pp. 6516–6534, July 2006.
- [4] O. K. Lee, J. M. Kim, Y. Bresler, and J. C. Ye, "Compressive diffuse optical tomography: non-iterative exact reconstruction using joint sparsity," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1129–1142, 2011.
- [5] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham, and R.G. Baraniuk, "Distributed compressed sensing," *preprint*, 2005.
- [6] M.F. Duarte, M.B. Wakin, D. Baron, and R.G. Baraniuk, "Universal distributed sensing via random projections," in *Proceedings of the International Conference on Information Processing in Sensor Networks*, Nashville, TN, 2006, pp. 177–185.
- [7] J.M. Kim, O.K. Lee, and J.C. Ye, "Compressive MUSIC: revisiting the link between compressive sensing and array signal processing," *IEEE Trans. on Information Theory*, vol. 58, no. 1, pp. 278–301, 2012.
- [8] 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.
- [9] 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.
- [10] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. on Signal Processing*, vol. 56, pp. 4692–4702, 2009.

- [11] E. Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2516–2527, 2010.
- [12] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. on Information Theory*, vol. 58, no. 6, pp. 3613–3641, 2012.
- [13] G. Obozinski, M.J. Wainwright, and M.I. Jordan, "Support union recovery in high-dimensional multivariate regression," *The Annals of Statistics*, vol. 39, no. 1, pp. 1–47, 2011.
- [14] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [15] D.A. Linebarger, R.D. DeGroat, E.M. Dowling, P. Stoica, and G.L. Fudge, "Incorporating a priori information into MUSIC-algorithms and analysis," *Signal Processing*, vol. 46, no. 1, pp. 85–104, 1995.
- [16] P. Wirfalt, M. Jansson, G. Bouleux, and P. Stoica, "Prior knowledge-based direction of arrival estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 2540–2543.
- [17] 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.
- [18] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, pp. 1580–1590, 2012.
- [19] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. on Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.
- [20] 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.
- [21] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *Proceedings of the IEEE International Symposium of Information Theory (ISIT 2008)*, Toronto, Canada, 2008, pp. 2187–2191.
- [22] S. Rangan A.K. Fletcher and V.K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. on Information Theory*, vol. 55, no. 12, pp. 5758–5772, December 2009.
- [23] P. Stoica, P. Handel, and A. Nehoral, "Improved sequential MUSIC," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 31, no. 4, pp. 1230–1239, 1995.
- [24] S.K. Oh and C.K. Un, "A sequential estimation approach for performance improvement of eigenstructure-based methods in array processing," *IEEE Trans. on Signal Processing*, vol. 41, no. 1, pp. 457, 1993.
- [25] J.C. Mosher and R.M. Leahy, "Recursive MUSIC: a framework for EEG and MEG source localization," *IEEE Trans. on Biomedical Engineering*, vol. 45, no. 11, pp. 1342–1354, 1998.
- [26] J.C. Mosher and R.M. Leahy, "Source localization using recursively applied and projected (RAP) MUSIC," *IEEE Trans. on Signal Processing*, vol. 47, no. 2, pp. 332–340, 1999.
- [27] M.E. Davies and Y.C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. on Information Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.
- [28] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.
- [29] P.-Å. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT*, pp. 99–111, 1972.

PLACE  
PHOTO  
HERE

**Jong Min Kim** (M'12) 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, applied Fourier analysis, inverse problem and medical imaging and communication application.

PLACE  
PHOTO  
HERE

**Ok Kyun Lee** (S'09) received the B.S. degree in electrical, electronic and computer engineering in 2007 from the Hanyang University. He received the M.S. degree in electronic and electrical engineering in 2009 from the Korea Advanced Institute of Science and Technology (KAIST), where he is currently pursuing the Ph.D. degree. His research interests include compressive sensing, neuroimaging, and diffuse optical tomography.

PLACE  
PHOTO  
HERE

**Jong Chul Ye** (M'99) 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 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 Affiliated 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. He is a current member of IEEE EMBS Technical Committee on Biomedical Imaging and Image Processing (TC-BIIP).