IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 10, 2018, accepted December 17, 2018, date of publication December 28, 2018, date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889098

On Improving Recovery Performance in Multiple Measurement Vector Having Dependency

SEYYED HAMED FOULADI[1] AND ILANGKO BALASINGHAM[1,2]

[1]Department of Electronics and Telecommunications, Norwegian University of Science and Technology, 7491 Trondheim, Norway
[2]Intervention Center, Oslo University Hospital, 0027 Oslo, Norway

Corresponding author: Seyyed Hamed Fouladi (hamed.fouladi@ntnu.no)

This work was supported by the Research Council of Norway Project "Medical Sensing, Localization, and Communication Using Ultra Wideband Technology (MELODY) (Phase II)" under Grant 225885.

ABSTRACT The multiple measurement vector (MMV) problem is applicable in a wide range of applications such as photoplethysmography (PPG), remote PPG measurement, heart rate estimation, and directional arrival estimation of multiple sources. Measurements in the aforementioned applications exhibit a dependency structure, which is not considered in the general MMV algorithms. Modeling the dependency or the correlation structure of the solution matrix to MMV problems can increase the recovery performance. The solution matrix $\boldsymbol{X}$ can be decomposed into a mixing matrix $\boldsymbol{A}$ and a sparse matrix with independent columns $\boldsymbol{S}$. The key idea of this model is that the matrix S can be sparser than the mixing matrix $\boldsymbol{A}$. Previous MMV algorithms did not consider such a structure for $\boldsymbol{X}$. This paper proposes two algorithms, which are based on orthogonal matching pursuit and basis pursuit, and derives the exact recovery guarantee conditions for both approaches. We compare the simulation results of the proposed algorithms with the conventional algorithms and show that the proposed algorithms outperform previous algorithms especially in the case of the low number of measurements.

INDEX TERMS Multiple measurement vectors, independent component analysis, orthogonal matching pursuit, basis pursuit.

I. INTRODUCTION

Sparse representation has attracted a wide range of applications among others such as imaging [1], [2], biomedical signal processing [3]–[5], radar signal processing [6]–[8], and remote sensing [9]. Sparse representation has been effective for handling these problems which are related to compressed sensing (CS) - a famous research topic in recent years [10]–[12]. Compressed sensing or single measurement vector (SMV) problem suggests the recovery of a sparse vector with a few number of measurements. The SMV problem can be written as

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^{N\times 1}$ and $\boldsymbol{y} \in \mathbb{R}^{M\times 1}$ are the solution vector and the observation vector, respectively. The matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ is the dictionary matrix, where $M \ll N$. Since $\boldsymbol{\Phi}$ is a fat matrix, there will be infinite solutions to (1). Moreover, the sparsest solution to (1) is the desired solution. The solution minimizes the following optimization problem

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0, \quad \text{s.t.}\ \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}, \tag{2}$$

where $\|\boldsymbol{x}\|_0$ denoted the ℓ_0-norm of $\boldsymbol{x}$, i.e. the number of nonzero elements of $\boldsymbol{x}$. One of the important issues is that to make sure the solution to the optimization problem (2) is unique. In order to have a unique solution, $\boldsymbol{\Phi}$ needs to satisfy the following inequality

$$|\text{Supp}(\boldsymbol{x})| < \frac{\text{Spark}(\boldsymbol{\Phi})}{2}, \tag{3}$$

where $\text{Supp}(\boldsymbol{x})$ shows the support of $\boldsymbol{x}$ or the set including the indexes of the nonzero elements of $\boldsymbol{x}$ and the operator $|.|$ gives the cardinality of a set. The operator Spark is defined as

Definition 1: The Spark *of a matrix* $\boldsymbol{\Phi}$ *is defined as the smallest number of linearly dependent columns of* $\boldsymbol{\Phi}$.

There are many applications such as electroencephalogram (EEG) [13] and photoplethysmogram (PPG) [3] signal estimation where multiple measurement vectors are available. Multiple measurement vector (MMV) problem is the generalization of the SMV problem in (1) which can be written as

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X}, \tag{4}$$

where $\boldsymbol{Y} \in \mathbb{R}^{M\times L}$ is the matrix including observation vectors and $\boldsymbol{X} \in \mathbb{R}^{N\times L}$ is the solution matrix ($L \ll M$). Here,

the target is to find the solution with the lowest number of nonzero rows, i.e.,

$$\min_{\boldsymbol{X}} \mathcal{R}(\boldsymbol{X}), \quad \text{subject to } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X}, \tag{5}$$

where $\mathcal{R}(\boldsymbol{X})$ denotes the number of nonzero rows of $\boldsymbol{X}$. The matrix $\boldsymbol{X}$ with the lowest number of rows can be uniquely recovered if the following condition is satisfied

$$\mathcal{R}(\boldsymbol{X}) < \frac{\text{Spark}(\boldsymbol{\Phi}) + \text{Rank}(\boldsymbol{Y}) - 1}{2}. \tag{6}$$

As it can be seen, the condition in (6) is less restrictive than the condition in (3) if $\text{Rank}(\boldsymbol{Y}) > 1$. This motivates to use multiple measurements.

Most existing MMV algorithms assume that each row of $\boldsymbol{X}$ is independent and identically distributed (i.i.d). This is not suitable for many real-world scenarios since, in practice, rows of $\boldsymbol{X}$ will have certain structures, like temporal structure [14]. In [14], it was shown that the recovery performance of exciting algorithms was affected by the temporal structure. In some applications, signals are block sparse [15]. This means that the desired signal can be grouped into blocks. Some of these block are nonzero. Some algorithms have been proposed to recover sparse signals with the block structure [16], [17]. Considering block structure in the signals have been found effective [18], [19].

In [20] and [21], it was shown that if $\boldsymbol{X}$ can be decomposed into matrix sparse $\boldsymbol{S}$ and mixing matrix $\boldsymbol{A}$, i.e. $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$, one can estimate $\boldsymbol{X}$ through estimating $\boldsymbol{S}$ and $\boldsymbol{A}$ using

$$\min_{\boldsymbol{S},\boldsymbol{A}} \sum_{i=1}^{L} \|\boldsymbol{s}_i\|_0, \quad \text{s.t. } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}, \quad \left\|\boldsymbol{a}^j\right\|_2 = 1, \tag{7}$$

where $\boldsymbol{a}^j$ is the j'th row of $\boldsymbol{A}$ and $\boldsymbol{s}_i$ is the i'th column of $\boldsymbol{S}$. The ℓ_2-norm of each row $\boldsymbol{A}$ is set to be 1 in order to avoid scale ambiguity.

In [20], it is proved that the recovered $\boldsymbol{X}$ using (7) is unique if the following inequity holds

$$\max_i \{r_i\} + \sum_{j=1}^{L} r_j < \text{Spark}(\boldsymbol{\Phi}), \tag{8}$$

where r_i is the number of nonzero elements of $\boldsymbol{s}_i$.

The problems (2), (5) and (7) are generally difficult to solve because combinatorial search for solving NP-hard problem often consumes intractable time [22], [23]. To overcome this problem, one way is to convexifying the ℓ_0-norm using ℓ_1-norm. This approach is called basis pursuit (BP) which converges to the sparsest solution under some specific conditions [24]. Another approach is to find the best columns in the dictionary $\boldsymbol{\Phi}$ using greedy search. This approach also leads to the sparsest representation under some specific conditions [24].

A. BASIS PURSUIT APPROACH

The principal of BP is to obtain the solution to (1) whose ℓ_1-norm is minimal [25],

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1, \quad \text{s.t. } \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}. \tag{9}$$

The above problem (9) can be seen as convexified problem (2). In [24] and [26]–[29], it has been proved that under certain condition the solutions to (2) and (9) can be equivalent. The exact recovery condition for (9) can be described using null-space property [28]: A r-sparse solution $\boldsymbol{x}$ of the linear system $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ is exactly recovered by solving the ℓ_1-optimization in (9) if and only if $\boldsymbol{\Phi}$ satisfies the null-space property of r, i.e. NSP(r), where the null-space property is defined as [29],

Definition 2: An $M \times N$ matrix, $\boldsymbol{\Phi}$, satisfies the null-space property of order r, NSP(r), if for any subset $B \subset \{1, \cdots, N\}$ and $|B| = r$; and for any nonzero vector $\boldsymbol{v}$ in the null-space of $\boldsymbol{\Phi}$, $\boldsymbol{v} \in Ker(\boldsymbol{\Phi})$, the following inequality holds

$$\left\|\boldsymbol{v}\big|_B\right\|_1 < \left\|\boldsymbol{v}\big|_{B^c}\right\|_1, \tag{10}$$

where B^c is the complement of the set B and $\boldsymbol{v}\big|_B$ is the vector $\boldsymbol{v}$ restricted on the set B.

Several sparse representation algorithms with the ℓ_1-norm minimization have been proposed to solve (9) [30]–[34]. One way is to use constrained optimization strategy to find the solution with the least ℓ_1-norm. The algorithms that address the non-differentiable unconstrained problem are given by reformulating it as a smooth differentiable constrained optimization problem [29]. In [31], an algorithm was proposed to obtain the sparse representation solution along with the gradient descent direction. In [30], an algorithm called the truncated Newton based interior-point method (TNIPM) was proposed to solve the ℓ_1-regularized problem.

B. GREEDY APPROACH

The key idea of the greedy approach is to determine the support of $\boldsymbol{x}$ based on the relationship between the columns of $\boldsymbol{\Phi}$ and probe sample. The amplitude value is obtained by using the support of $\boldsymbol{x}$ and solving a least square problem [35], [36]. The matching pursuit (MP) algorithm is the earliest in the greedy approach for the sparse approximation [37]. The orthogonal matching pursuit (OMP) algorithm is an improved version of the MP algorithm [38], [39]. In each iteration, the OMP exploits the process of orthogonalization to project in the orthogonal direction. Greedy algorithms have been proposed based on the MP and OMP algorithms like the efficient OMP algorithm [40]. A regularized version of OMP (ROMP) algorithm recovered all r-sparse signals based on the restricted isometry property (RIP) of random frequency measurements [41].

A sufficient condition for exact sparse recovery for OMP was presented in [24]

$$\max_j \|\boldsymbol{\Phi}_B^+ \boldsymbol{\phi}_j\|_1 < 1, \quad j \notin \text{Supp}(\boldsymbol{x}), \tag{11}$$

where B is equal to $\text{Supp}(\boldsymbol{x})$, $\boldsymbol{\Phi}_B^+$ is a matrix with the columns whose indexes are in the support of $\boldsymbol{x}$. $\boldsymbol{\phi}_j$ denotes the j'th column of $\boldsymbol{\Phi}$.

C. CONTRIBUTION

In this paper, the solution matrix $\boldsymbol{X}$ is decomposed into a sparse matrix $\boldsymbol{S}$ and a mixing matrix $\boldsymbol{A}$, i.e., $\boldsymbol{X} = \boldsymbol{SA}$. The benefit of the mixing model is that $\boldsymbol{S}$ is sparser than $\boldsymbol{X}$ in some applications. This led us to develop two algorithms to estimate $\boldsymbol{S}$ and $\boldsymbol{A}$ unlike traditional MMV algorithms which estimate $\boldsymbol{X}$. One of the proposed algorithm is based on minimizing the sum of the ℓ_1-norm of the columns of $\boldsymbol{S}$. We call the algorithm independent component analysis basis pursuit (ICABP). We will derive the exact recovery guarantee conditions for both known and unknown $\boldsymbol{A}$ cases. We show that estimating $\boldsymbol{S}$ and $\boldsymbol{A}$ instead of $\boldsymbol{X}$ can improve the recovery performance. This structure fits to the applications like rPPG, PPG signal extraction, image separation or directional arrival estimation of multiple sources [42]–[46]. We also propose an algorithm based on OMP, called independent component analysis orthogonal matching pursuit (ICAOMP). Moreover, we derive the condition which guarantees the exact recovery of the ICAOMP algorithm.

This paper is organized as follows. In section II, the problem formulation and the solution structure are discussed. The proposed algorithms are presented in Section III. The conditions of recovery guarantee for the proposed algorithms are derived in Section IV. Experimental results are provided for evaluating the proposed algorithms in Section V.

Notations:

- $\|\boldsymbol{s}\|_p$, $\|\boldsymbol{A}\|_{\mathcal{F}}$ denote the ℓ_p-norm of the vector $\boldsymbol{s}$, and the Frobenius norm of the matrix $\boldsymbol{A}$, respectively.
- $\mathcal{R}(\boldsymbol{X})$ denotes the number of nonzero rows in the matrix $\boldsymbol{X}$.
- diag$\{a_1, \cdots, a_L\}$ denotes a diagonal matrix with diagonal elements being $a_1, \cdots, a_L$.
- For a matrix $\boldsymbol{A}$ and a vector $\boldsymbol{s}$, $\boldsymbol{A}_{(i,j)}$, and $\boldsymbol{s}_{(i)}$ denote the element that lies in the ith row and the jth column of $\boldsymbol{A}$ and ith element of $\boldsymbol{s}$, respectively.
- $\boldsymbol{A} \otimes \boldsymbol{B}$ represents the Kronecker product of the two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Tr$(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$. $\boldsymbol{A}^T$ denotes the transpose of $\boldsymbol{A}$.
- vec$(\boldsymbol{A})$ denotes the vectorization of the matrix $\boldsymbol{A}$ formed by stacking its columns into a single column vector.
- $[N]$ denotes the set $\{1, \cdots, N\}$.
- For a vector $\boldsymbol{s} \in \mathbb{R}^N$, Supp$(\boldsymbol{s})$ denotes $\{i \in [N] : \boldsymbol{s}_{(i)} \neq 0\}$. the support of $\boldsymbol{s}$. For a matrix $\boldsymbol{A} \in \mathbb{R}^{N\times L}$, Supp$(\boldsymbol{A})$ is $\{i \in [N] : \boldsymbol{A}_{(\vec{i})} \neq \boldsymbol{0}\}$.
- $|K|$ is the cardinality of the set K.
- $\boldsymbol{A}^{(k)}$ and $\boldsymbol{a}^{(k)}$ show the matrix $\boldsymbol{A}$ and the vector $\boldsymbol{a}$ updated in the k'th step of the proposed algorithm, respectively.

II. PROBLEM FORMULATION

The MMV problem can be expressed as

$$\boldsymbol{Y} = \boldsymbol{\Phi X}, \tag{12}$$

where $\boldsymbol{Y} \in \mathbb{R}^{M\times L}$, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ and $\boldsymbol{X} \in \mathbb{R}^{N\times L}$ are the observation matrix, the dictionary and the solution matrix, respectively ($M \ll N$ and $L \ll N$). The independent components are mixed and observed through some channels or sensors. Mathematically speaking, each column of $\boldsymbol{X}$ represents a mixture of these signals, i.e.,

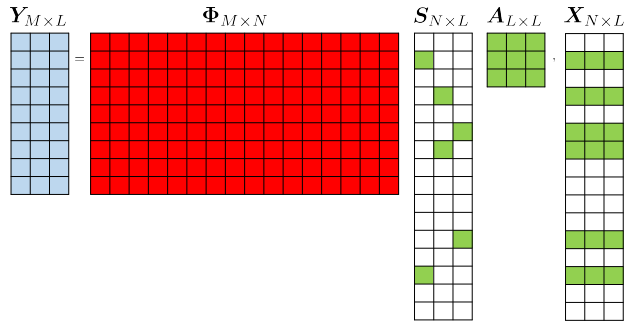

FIGURE 1. Visual representation of (12) and (15).

$$\boldsymbol{X} = \boldsymbol{SA}, \tag{13}$$

where $\boldsymbol{A} \in \mathbb{R}^{L\times L}$ is an unknown full-rank mixing matrix, and $\boldsymbol{S} \in \mathbb{R}^{N\times L}$ is an unknown source matrix including L source vectors where each column shows a sparse source. A source $\boldsymbol{s}_i$, i.e., the i'th column of $\boldsymbol{S}$, is an unknown sparse vector. It is assumed that the sources are independent. The multiplication of the source matrix $\boldsymbol{S}$ and the mixing matrix $\boldsymbol{A}$ results in matrix $\boldsymbol{X}$, which consists of L linear mixtures of the sources. The matrix $\boldsymbol{A}$ captures the dependences among columns of $\boldsymbol{X}$. The motivation of this modeling is that the matrix $\boldsymbol{X}$ can have a higher number of nonzero elements than each column $\boldsymbol{S}$ as shown Fig. 1. The following trivial lemma describes the minimum and maximum value of nonzero rows of $\boldsymbol{X}$ in terms of the columns of $\boldsymbol{S}$.

Lemma 1: Let the columns of a matrix $\boldsymbol{S} \in \mathbb{R}^{N\times L}$ be r_i-sparse and independent, and a matrix $\boldsymbol{A} \in \mathbb{R}^{L\times L}$ be a full-rank matrix. If $\boldsymbol{X} = \boldsymbol{SA}$, the number of nonzero rows of the matrix $\boldsymbol{X}$ meets the following inequality

$$\max_i \{|\text{Supp}(\boldsymbol{s}_i)|\} \leq \mathcal{R}(\boldsymbol{X}) \leq \sum_{i=1}^{L} |\text{Supp}\,(\boldsymbol{s}_i)|\,. \tag{14}$$

Proof: See Appendix A. □

Intuitively, since $\boldsymbol{X}$ has a higher number of nonzero rows, if we could find $\boldsymbol{A}$, it would be easier to solve the problem. This will be discussed in detail in Section IV. Fig 1 shows the problem while $\boldsymbol{X}$ is composed of the linear mixtures of independent sources.

Using (12) and (13), one can write the observation matrix in terms of $\boldsymbol{S}$ and $\boldsymbol{A}$ as

$$\boldsymbol{Y} = \boldsymbol{\Phi SA}. \tag{15}$$

Since we want to find the sparsest $\boldsymbol{S}$ satisfying (15), we would ideally like to solve the optimization problem in (7).

Remark 1: $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ is considered as solution where the i'th column of $\tilde{\boldsymbol{S}}$ is r_i sparse and $\tilde{\boldsymbol{S}}$ is the sparsest solution.

Remark 2: In (7), the solution $\tilde{\boldsymbol{S}}$ has the least number of nonzero elements. However, the aim of the MMV problem is to find the matrix $\tilde{\boldsymbol{X}}$, therefore $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ is also called the solution in this work.

Before tackling the problem in (7), we show in the following theorem that solving (7) leads to the matrix $\boldsymbol{X}$ with the least number of nonzero rows. This is important because MMV algorithms' object is to find $\boldsymbol{X}$ with the least nonzero rows.

Theorem 1: If $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ is the solution to (7) and the condition (8) is satisfied, the multiplication $\tilde{\boldsymbol{S}}$ and $\tilde{\boldsymbol{A}}$, i.e. $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ will have the least number of nonzero rows in $\boldsymbol{X}$s satisfying $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X}$.

Proof: See Appendix B. □

Due to the complexity of minimizing the ℓ_0-norm, two algorithms are proposed, one based on orthogonal matching pursuit and the other one based on basis pursuit approach which will be discussed in the following section.

III. ALGORITHMS

One should ideally solve the optimization (7), however, the problem in (7) is computationally complicated because of the ℓ_0-norm minimization.

A. INDEPENDENT COMPONENT ANALYSIS BASIS PURSUIT (ICABP)

The problem in (7) can be convexified by the following optimization,

$$\min_{\boldsymbol{S},\boldsymbol{A}} \sum_{i=1}^{L} \|\boldsymbol{s}_i\|_1, \quad \text{s.t. } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}, \quad \left\|\boldsymbol{a}^j\right\|_2 = 1. \tag{16}$$

In order to solve (16), an algorithm is proposed which minimizes $\sum_{i=1}^{N} \|\boldsymbol{s}_i\|_1$ and the corresponding mixing matrix $\boldsymbol{A}$. This can be written similar to LASSO as [47]

$$\min_{\boldsymbol{S},\boldsymbol{A}} \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\|_{\mathcal{F}}^2 + \lambda \sum_i \|\boldsymbol{s}_i\|_1. \tag{17}$$

If applying the operator vec on both sides of (15), it can be written as

$$\begin{aligned} \text{vec}_{L,M}\left(\boldsymbol{Y}^T\right) &= \text{vec}_{L,M}\left((\boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A})^T\right) \\ &= (\boldsymbol{\Phi} \otimes \boldsymbol{A}^T)\text{vec}_{L,N}(\boldsymbol{S}^T). \end{aligned} \tag{18}$$

For a simpler notation, let $\boldsymbol{y} = \text{vec}_{L,M}(\boldsymbol{Y}^T)$, $\boldsymbol{s} = \text{vec}_{L,N}(\boldsymbol{S}^T)$ and $\boldsymbol{\phi}_{\boldsymbol{A}} = \boldsymbol{\Phi} \otimes \boldsymbol{A}^T$. We can write (18) as

$$\boldsymbol{y} = \boldsymbol{\phi}_{\boldsymbol{A}}\boldsymbol{s}. \tag{19}$$

This leads to an SMV while the dictionary matrix $\boldsymbol{\phi}_{\boldsymbol{A}}$ is not known due to an unknown $\boldsymbol{A}$. Therefore (17) can be equivalently written as

$$\min_{\boldsymbol{s},\boldsymbol{A}} \mathcal{L}(\boldsymbol{s},\boldsymbol{A}) = \min_{\boldsymbol{s},\boldsymbol{A}} \frac{1}{2} \left\|\boldsymbol{y} - \boldsymbol{\phi}_{\boldsymbol{A}}\boldsymbol{s}\right\|_2^2 + \lambda \|\boldsymbol{s}\|_1. \tag{20}$$

The gradient projection sparse reconstruction (GPSR) algorithm can be employed to solve (20).

The sparse representation solution $\boldsymbol{s}$ can be formulated by its positive and negative parts as

$$\boldsymbol{s} = \boldsymbol{s}_+ + \boldsymbol{s}_-, \quad \boldsymbol{s}_+ \succeq \boldsymbol{0}, \ \boldsymbol{s}_- \succeq \boldsymbol{0}, \tag{21}$$

where the operators $(.)_+$ and $(.)_-$ are

$$\begin{aligned} \boldsymbol{s}_+ &= \begin{cases} (\boldsymbol{s}_+)_{(i)} = \boldsymbol{s}_{(i)} & \boldsymbol{s}_{(i)} > 0 \\ 0, & \text{otherwise}, \end{cases} \\ \boldsymbol{s}_- &= \begin{cases} (\boldsymbol{s}_-)_{(i)} = -\boldsymbol{s}_{(i)} & \boldsymbol{s}_{(i)} < 0 \\ 0, & \text{otherwise} \end{cases}. \end{aligned} \tag{22}$$

With this notation, the optimization problem (20) can be reformulated as

$$\min_{\boldsymbol{s},\boldsymbol{A}} \frac{1}{2} \left\|\boldsymbol{y} - \boldsymbol{\phi}_{\boldsymbol{A}}(\boldsymbol{s}_+ - \boldsymbol{s}_-)\right\|_2^2 + \lambda \left(\boldsymbol{1}_{LN}^T\boldsymbol{s}_+ + \boldsymbol{1}_{LN}^T\boldsymbol{s}_-\right), \tag{23}$$

where $\boldsymbol{1}_{LN} = [1, \cdots, 1]^T \in \mathbb{R}^{LN}$. The cost function in (23) can be rewritten as

$$\begin{aligned} \min_{\boldsymbol{s},\boldsymbol{A}} \mathcal{G}(\boldsymbol{s},\boldsymbol{A}) &= \min_{\boldsymbol{s},\boldsymbol{A}} \frac{1}{2}\boldsymbol{u}^T\boldsymbol{C}_{\boldsymbol{A}}\boldsymbol{u} + \boldsymbol{c}^T\boldsymbol{u}, \\ \boldsymbol{u} &= \begin{bmatrix} \boldsymbol{s}_+ \\ \boldsymbol{s}_- \end{bmatrix}, \end{aligned} \tag{24}$$

$$\begin{aligned} \boldsymbol{c} &= \lambda\boldsymbol{1}_{2LN} + \begin{bmatrix} -\boldsymbol{\phi}_{\boldsymbol{A}}^T\boldsymbol{y} \\ \boldsymbol{\phi}_{\boldsymbol{A}}^T\boldsymbol{y} \end{bmatrix}, \\ \boldsymbol{C}_{\boldsymbol{A}} &= \begin{bmatrix} \boldsymbol{\phi}_{\boldsymbol{A}}^T\boldsymbol{\phi}_{\boldsymbol{A}} & -\boldsymbol{\phi}_{\boldsymbol{A}}^T\boldsymbol{\phi}_{\boldsymbol{A}} \\ -\boldsymbol{\phi}_{\boldsymbol{A}}^T\boldsymbol{\phi}_{\boldsymbol{A}} & \boldsymbol{\phi}_{\boldsymbol{A}}^T\boldsymbol{\phi}_{\boldsymbol{A}} \end{bmatrix}. \end{aligned} \tag{25}$$

The GPSR algorithm exploits the gradient descent and standard line-search method [48] to solve (24). With this approach, the vector $\boldsymbol{u}$ is updated as

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} - \mu^{(k)} \frac{\partial \mathcal{G}}{\partial \boldsymbol{u}}\bigg|_{\boldsymbol{u}=\boldsymbol{u}^{(k)},\boldsymbol{A}=\boldsymbol{A}^{(k)}}, \tag{26}$$

where $\partial\mathcal{G}/\partial\boldsymbol{u} = \boldsymbol{c} + \boldsymbol{C}_{\boldsymbol{A}}\boldsymbol{u}$ and μ is the step size.

A closed-form for updating $\boldsymbol{A}$ is obtained by taking the derivative of $\mathcal{G}$ with respect to $\boldsymbol{A}$ and setting the result to zero.

$$\frac{\partial\mathcal{G}}{\partial\boldsymbol{A}} = \frac{\partial\mathcal{L}}{\partial\boldsymbol{A}} = \frac{\partial\left(\|\boldsymbol{y} - \boldsymbol{\phi}_{\boldsymbol{A}}\boldsymbol{s}\|_2^2\right)}{\partial\boldsymbol{A}} = \frac{\partial\left(\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\|_{\mathcal{F}}^2\right)}{\partial\boldsymbol{A}} = 0. \tag{27}$$

Solving (27) leads to the $\boldsymbol{A}$ with the least error given $\boldsymbol{S}$. The derivative in (27) can be expanded as

$$\begin{aligned} \frac{\partial\mathcal{G}}{\partial\boldsymbol{A}} &= \frac{\partial}{\partial\boldsymbol{A}}(\text{Tr}\left(\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\right)\left(\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\right)^T) \\ &= \frac{\partial}{\partial\boldsymbol{A}}(\text{Tr}(-2\boldsymbol{Y}\boldsymbol{A}^T\boldsymbol{S}^T\boldsymbol{\Phi}^T + \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{S}^T\boldsymbol{\Phi}^T)) \\ &= \frac{\partial}{\partial\boldsymbol{A}}(\text{Tr}(-2\boldsymbol{Y}^T\boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A} + \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{S}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{S})) \\ &= -2\boldsymbol{S}^T\boldsymbol{\Phi}^T\boldsymbol{Y} + 2\boldsymbol{S}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A} = 0. \end{aligned} \tag{28}$$

One can update $\boldsymbol{A}^{(k+1)}$ in terms of $\boldsymbol{S}^{(k)}$ using (28)

$$\boldsymbol{A}^{(k+1)} = \left(\left(\boldsymbol{S}^{(k)}\right)^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{S}^{(k)}\right)^{-1}\left(\boldsymbol{S}^{(k)}\right)^T\boldsymbol{\Phi}^T\boldsymbol{Y}, \tag{29}$$

where $\boldsymbol{S}^{(k)}$ can be simply obtained by

$$\boldsymbol{S}^{(k)^T} = \text{vec}_{L,N}^{-1}\left(\left(\boldsymbol{s}^{(k)}\right)\right), \tag{30}$$

TABLE 1. ICABP: The independent component analysis basis pursuit algorithm.

Input : $\boldsymbol{\Phi}$, $\boldsymbol{Y}$, λ, ε, β and $\boldsymbol{y} = \text{vec}_{L,M}(\boldsymbol{Y}^T)$.

1) Initialization: $k = 0$, $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{S} = \boldsymbol{\Phi}^+\boldsymbol{Y}$.
2) $k = k + 1$.
3) Obtain $\phi_{\boldsymbol{A}^{(k)}} = \boldsymbol{\Phi} \otimes \left(\boldsymbol{A}^{(k)}\right)^T$.
4) Update $\boldsymbol{u}$ by $\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} - \mu^{(k)}\left(\boldsymbol{c} + \boldsymbol{C}_{\boldsymbol{A}^{(k)}}\boldsymbol{u}^{(k)}\right)$, and accordingly update $\boldsymbol{S}^{(k+1)}$.
5) Update μ by $\mu^{(k)} = \frac{(\boldsymbol{h}^{(k)})^T\boldsymbol{h}^{(k)}}{(\boldsymbol{h}^{(k)})^T\boldsymbol{C}_{\boldsymbol{A}^{(k)}}\boldsymbol{h}^{(k)}}$.
6) Update $\boldsymbol{A}^{(k+1)}$ by (29)
7) Set each row of $\boldsymbol{A}^{(k+1)}$ to 1.
8) If (34) is not satisfied, go to step 2.
9) Return $\boldsymbol{S}$ and $\boldsymbol{A}$.

$\boldsymbol{s}$ is given using (24) and (21). In order to set the step size, one may use the approach employed in GPSR [31],

$$\mu^{(k)} = \arg\min_{\mu} \mathcal{G}(\boldsymbol{u}^{(k)} - \mu\boldsymbol{h}^{(k)}), \tag{31}$$

where $\boldsymbol{h}$ is defined as

$$\boldsymbol{h}_{(i)}^{(k)} = \begin{cases} \left(\dfrac{\partial\mathcal{G}(\boldsymbol{u}^{(k)}, \boldsymbol{A}^{(k)})}{\partial\boldsymbol{u}}\right)_{(i)}, & \\ \quad u_{(i)}^{(k)} > 0 \text{ or } \left(\frac{\partial\mathcal{G}(\boldsymbol{u}^{(k)},\boldsymbol{A}^{(k)})}{\partial\boldsymbol{u}}\right)_{(i)} < 0 \\ 0, \quad \text{otherwise.} \end{cases} \tag{32}$$

The closed-form solution for (31) is given in [31]

$$\mu^{(k)} = \frac{(\boldsymbol{h}^{(k)})^T\boldsymbol{h}^{(k)}}{(\boldsymbol{h}^{(k)})^T\boldsymbol{C}_{\boldsymbol{A}}\boldsymbol{h}^{(k)}}. \tag{33}$$

Moreover, the GPSR algorithm uses the backtracking linear search method [31] that ensures the step size of the gradient decreases in each iteration. In [31], it is proposed that the stop condition of the backtracking linear search should satisfy the following inequality

$$\begin{aligned} &\mathcal{G}((\boldsymbol{u}^{(k)} - \mu^{(k)}\partial\mathcal{G}(\boldsymbol{u}^{(k)}))_+) \\ &\quad > \mathcal{G}(\boldsymbol{u}^{(k)}) - \beta\left(\frac{\mathcal{G}(\boldsymbol{u}^{(k)})}{\partial\boldsymbol{u}}\right)^T \\ &\qquad \times \left(\boldsymbol{u}^T - \mathcal{G}((\boldsymbol{u}^{(k)} - \mu^{(k)}\partial\mathcal{G}(\boldsymbol{u}^{(k)}))_+)\right), \end{aligned} \tag{34}$$

where β is a small constant value. Table 1 lists the steps of the proposed algorithm.

Remark 3: This work deals with the case that the number of columns of $\boldsymbol{Y}$ is equal to the number of columns of $\boldsymbol{X}$, and matrices $\boldsymbol{X}$ and $\boldsymbol{A}$ are full-rank. The case that the number of columns of $\boldsymbol{Y}$ is larger than the number of columns of $\boldsymbol{X}$, means that the columns of $\boldsymbol{Y}$ are linearly dependent. However, in this case, the extra columns of $\boldsymbol{Y}$ do not provide additional information when noise is not contaminating the signals (See the model in (12)). Therefore, one can apply a dimension reduction algorithm on $\boldsymbol{Y}$ to find a matrix whose number of columns is equal to Rank($\boldsymbol{Y}$). Then, the problem in (17) can be solved with the assumptions in this work.

B. INDEPENDENT COMPONENT ANALYSIS ORTHOGONAL MATCHING PURSUIT (ICAOMP)

In this part, we describe how to apply the OMP algorithm to MMV problem when the solution matrix has the linear mixture structure. Suppose that the i'th column of $\tilde{\boldsymbol{S}}$ is r_i sparse and its support is B_i. The aim is to correctly find the support of $\tilde{\boldsymbol{S}}$. One way of looking at this problem is to apply the OMP algorithm on each column of $\boldsymbol{Y}$. Each column of $\boldsymbol{Y}$ is a linear combination of the columns of $\boldsymbol{\Phi}$ whose indexes are in $\bigcup_i B_i$. However, when the exact $\boldsymbol{A}$ is known, the multiplication of $\boldsymbol{Y}$ with the inverse of $\boldsymbol{A}$ results in $\boldsymbol{Y}'$ whose columns can only be represented with the linear mixture of the columns of $\boldsymbol{\Phi}$ with the indexes in B_i, i.e.

$$\boldsymbol{y}_i' = \sum_{j\in B_i} \boldsymbol{S}_{(j,i)}\boldsymbol{\phi}_j,$$

where $\boldsymbol{y}_i'$ is the i'th column of $\boldsymbol{Y}'$. Clearly, the cardinality of $\bigcup_i B_i$ is larger than the cardinality of B_i for all i. Then, in order to represent the columns of $\boldsymbol{Y}'$, one may need a smaller number of columns of $\boldsymbol{\Phi}$ compared with the sparse representation of $\boldsymbol{Y}$. First, the mixing matrix is estimated, and the columns of $\boldsymbol{Y}$ are demixed. Then, OMP is applied on the demixed observation.

In each iteration, the matrix $\boldsymbol{A}$ is updated by minimizing $\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\|_{\mathcal{F}}^2$ using the estimate of $\boldsymbol{S}$ in the last step. Since the same function is minimized similar to the ICABP in Section III-A, the same formulation will be obtained as in (29),

$$\boldsymbol{Y}^{(k_1)} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\left(\boldsymbol{A}^{(k_1)}\right)^{-1}. \tag{35}$$

Now we can apply the OMP algorithm on each column of $\boldsymbol{Y}^{(k_1)}$.

For each k_1 we apply the OMP algorithm. Let $B_i^{(k)}$ be the set including the indexes of the support of $\boldsymbol{s}_i^{(k)}$ in the k'th iteration. Table 2 lists the steps of the algorithm.

IV. RECOVERY GUARANTEE

In this section, the conditions, which guarantee the exact support can be obtained for the proposed algorithms, are studied.

A. UNIQUENESS CONDITION FOR ℓ_1-MINIMIZATION

The ICABP algorithm exploits the ℓ_1-norm to estimate $\boldsymbol{S}$ and $\boldsymbol{A}$, and accordingly $\boldsymbol{X}$. The aim is to find the condition in which the solution to (17) is the sparsest solution.

We start with the ideal case in which the mixing matrix is known or estimated exactly with zero error.

Theorem 2: Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ resulting in $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ to be a solution to the problem (P_1) where the columns of matrix $\tilde{\boldsymbol{S}}$ are supported on $B_i \subset [N]$ and $|B_i| = r_i$. If $\tilde{\boldsymbol{A}}$ is known and $\boldsymbol{\Phi}$

TABLE 2. ICAOMP: The independent component analysis orthogonal matching pursuit algorithm.

Input : $\boldsymbol{\Phi}, \boldsymbol{Y}, \lambda, \varepsilon, \boldsymbol{y} = \text{vec}_{L,M}(\boldsymbol{Y}^T)$

1) Initialization for matrix $\boldsymbol{A}$: $k_1 = 0$, $\boldsymbol{A}^{(0)} = \boldsymbol{I}$.
2) Initialization: $k = 0$, $\boldsymbol{S}^{(0)} = \boldsymbol{\Phi}^+\boldsymbol{Y}$ and $\boldsymbol{R}^{(0)} = \boldsymbol{Y}$.
3) $k = k + 1$.
4) Obtain $b_i^{(k+1)} = \arg\max_j(|\langle \boldsymbol{r}_i^{(k)}, \boldsymbol{\phi}_j\rangle|)$ for all i, where $\boldsymbol{r}_i$ is the i'th column of $\boldsymbol{R}$.
5) Obtain $B_i^{(k+1)} = B_i^{(k)} \cup \{b_i^{(k+1)}\}$ for all i.
6) Obtain $\boldsymbol{s}_i^{(k+1)} = \boldsymbol{\Phi}^+_{B_i^{(k)}}\boldsymbol{y}_i^{(k_1)}$ for all i, where $\boldsymbol{y}_i$ is the i'th column of $\boldsymbol{Y}$.
7) Update $\boldsymbol{R}^{(k+1)}$ by $\boldsymbol{R}^{(k+1)} = \boldsymbol{Y}^{(k_1)} - \boldsymbol{\Phi}\boldsymbol{S}^{(k)}$
8) Go to step 2 if $k \mod \|\boldsymbol{r}\|_1 \neq 0$.
9) Update $\boldsymbol{A}^{(k_1+1)}$ by (29).
10) Set each row of $\boldsymbol{A}^{(k_1+1)}$ to 1.
11) Go to step 1.
12) Return $\boldsymbol{S}$ and $\boldsymbol{A}$.

satisfies the null space property $R = \max\{r_i\}$, i.e., NSP(R), the solution with the least ℓ_1-norm is equal to the solution $\tilde{\boldsymbol{X}}$.

Proof: See Appendix C. □

This is, however, not a practical case, since matrix $\boldsymbol{A}$ is unknown. In the following theorem the condition that guarantees exact recovery in the case when $\boldsymbol{A}$ is unknown.

Theorem 3: Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ be the sparsest solution giving $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$, where the columns of matrix $\tilde{\boldsymbol{S}}$ are independent and supported on $B_i \subset [N]$ and $|B_i| = r_i$. If matrix $\boldsymbol{\Phi}$ satisfies the null space property $\left|\bigcup_{i=1}^N B_i\right|$, i.e., NSP($\left|\bigcup_{i=1}^N B_i\right|$), the matrix $\tilde{\boldsymbol{S}}$ is the sparsest by solving the minimization problem (16).

Proof: See Appendix D. □

Theorem 2 describes the case where $\boldsymbol{A}$ is estimated exactly without any error. Since $B_i \subseteq \bigcup_{i=1}^N B_i$, the condition in Theorem 2 can be much less restrictive than the condition in Theorem 3. The condition in Theorem 3 is the same as the condition for general MMV in [49] with a different cost function. Theorem 3 explains that $\boldsymbol{X}$ can be uniquely recovered irrespective of the estimate of $\boldsymbol{A}$. The first condition is for the ideal case (unrealistic) and the second one is very loose. Although the estimation of $\boldsymbol{A}$ depends on the algorithm used, different algorithms can estimate $\boldsymbol{A}$ with some accuracy [50]. Therefore, we derive a practical condition based on the accuracy of the estimate of $\boldsymbol{A}$. Since the matrix $\boldsymbol{A}$ can be estimated with some accuracy, i.e. $\left\|\boldsymbol{t}^T\tilde{\boldsymbol{A}}\hat{\boldsymbol{A}}^{-1}\right\|_1 < (1+\alpha)\left\|\boldsymbol{t}^T\right\|_1$ for a small α and a vector $\boldsymbol{t}$, a practical condition can be obtained as presented in the following theorem. We need the following definition in [51] and lemma for the theorem.

Definition 3: For any set $B \subset [N]$ with $|B| < \kappa$, a matrix $\boldsymbol{\Phi}$ is said to satisfy the ℓ_1 stable null space property of order κ with constant τ, $0 < 1 - \tau < 1$, if

$$\left\|\boldsymbol{v}|_B\right\|_1 < (1-\tau)\left\|\boldsymbol{v}|_{B^c}\right\|_1, \tag{36}$$

for all $\boldsymbol{v} \in \text{Ker}(\boldsymbol{\Phi})$.

Lemma 2: If $\{\bar{\boldsymbol{S}}, \bar{\boldsymbol{A}}\}$ is the solution to the equation $\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}$ whose columns' ℓ_1-norms are the least, then the following inequality holds

$$\sum_{i=1}^N \|\bar{\boldsymbol{s}}_i\|_1 \leq \sum_{i=1}^N \left\|\boldsymbol{\Phi}^+\boldsymbol{y}_i\right\|_1. \tag{37}$$

Proof: Assume

$$\sum_{i=1}^N \left\|\boldsymbol{\Phi}^+\boldsymbol{y}_i\right\|_1 < \sum_{i=1}^N \|\bar{\boldsymbol{s}}_i\|_1. \tag{38}$$

Based on the assumption in the lemma, $\bar{\boldsymbol{S}}$ is the solution with the least ℓ_1-norm. But using (38), the solution $\{\boldsymbol{\Phi}^+\boldsymbol{Y}, \boldsymbol{I}\}$, where $\boldsymbol{I}$ is an $L \times L$ identity matrix, will have a smaller ℓ_1-norm than $\{\tilde{\boldsymbol{S}}, \bar{\boldsymbol{A}}\}$. This is a contradiction and thus completes the lemma. □

Theorem 4: Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ be the solution giving $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ where the columns of matrix $\tilde{\boldsymbol{S}}$ are independent and supported on $B_i \subset [N]$ and $|B_i| = r_i$. The solution to (16) is equal to $\tilde{\boldsymbol{X}}$ if the following conditions hold

1) $\left\|\boldsymbol{t}^T\tilde{\boldsymbol{A}}\hat{\boldsymbol{A}}^{-1}\right\|_1 < (1+\alpha)\left\|\boldsymbol{t}^T\right\|_1$ *for any vector $\boldsymbol{t} \in \mathbb{R}^N$,*
2) *Matrix $\boldsymbol{\Phi}$ satisfies null space property of the order $\|\boldsymbol{r}\|_\infty$ with constant τ, where*

$$\tau > \frac{\alpha}{\delta}\frac{1}{N}\frac{\sum_{i=1}^N \left\|\boldsymbol{\Phi}^+\boldsymbol{y}_i\right\|_1}{\|\boldsymbol{r}\|_1}, \tag{39}$$

3) $\left|\tilde{\boldsymbol{S}}_{(i,j)}\right| > \delta$ *for $\delta > 0$,*

where $\boldsymbol{r} = [r_1, r_2, \cdots, r_N]$.

Proof: See Appendix E. □

The condition in Theorem 4 depends on the accuracy of the estimate of $\boldsymbol{A}$; i.e., α. If α goes to 0, in other words, if $\boldsymbol{A}$ is estimated exactly, τ will go to 0. Then, the conditions in Theorem 2 and Theorem 4 will be the same. This means we require the less restrictive condition for a better estimate of $\boldsymbol{A}$.

B. RECOVERY GUARANTEE FOR ICAOMP

In this subsection, the conditions are found which if satisfied, ICAOMP will find the support of $\tilde{\boldsymbol{X}}$. Here we start with the ideal case that the matrix $\tilde{\boldsymbol{A}}$ is known.

Theorem 5: Let $\tilde{\boldsymbol{S}}$ be the sparsest solution to (12) where the columns of matrix $\tilde{\boldsymbol{S}}$ are supported on $B_i \subset [N]$ and $|B_i| = r_i$. If matrix $\tilde{\boldsymbol{A}}$ is known, the solution $\tilde{\boldsymbol{X}}$ is unique if the matrix $\boldsymbol{\Phi}$ satisfies the following inequality

$$\left\|\boldsymbol{\Phi}^+_{B_I}\boldsymbol{\phi}_j\right\|_1 < 1, \quad I = \arg\max_i (r_i), \quad 1 \leq i \leq L. \tag{40}$$

Proof: See Appendix F. □

Theorem 5 shows that the condition is restricted by the set with the highest cardinality value.

Now we describe the condition for the ICAOMP recovery guarantee when $\boldsymbol{A}$ is unknown.

Theorem 6: Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ giving the matrix $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ to be the sparsest solution to (12) where the columns of matrix $\tilde{\boldsymbol{S}}$ are supported on $B_i \subset [N]$ and $|B_i| = r_i$. The matrix $\tilde{\boldsymbol{X}}$ can be

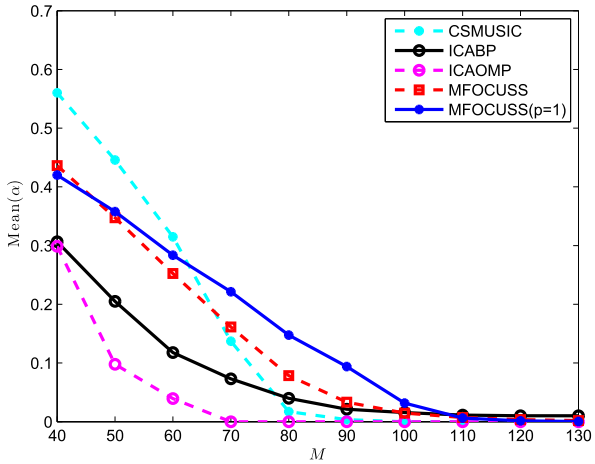

FIGURE 2. Mean of α in terms of number of measurements M.

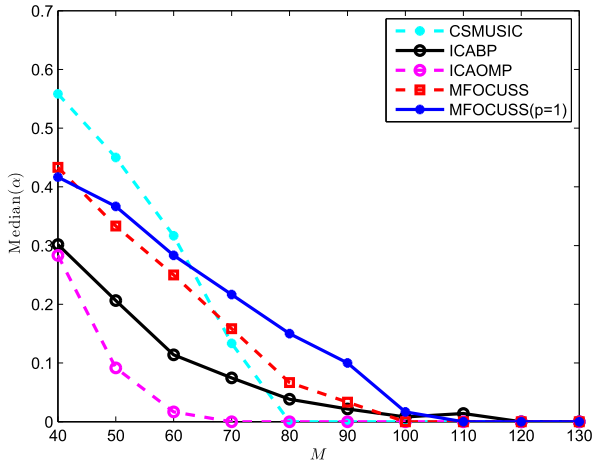

FIGURE 3. Mean of α in terms of number of measurements M.

uniquely recovered irrespective of the estimate of matrix $\boldsymbol{A}$, if matrix $\tilde{\boldsymbol{A}}$ is unknown and matrix $\boldsymbol{\Phi}$ satisfies the following inequality

$$\left\|\boldsymbol{\Phi}_B^{+}\boldsymbol{\phi}_j\right\|_1 < 1, \quad B = \bigcup_{i=1}^{L} B_i. \tag{41}$$

Proof: See Appendix G. □

Even though the condition in Theorem 6 is the same as the condition in [49], the algorithms are different. We show in Section V that the proposed algorithm is more effective when a certain structure is used on the solution matrix like for example PPG and rPPG applications.

V. RESULTS

The performance of the proposed algorithms is evaluated and compared with traditional algorithms. Different scenarios are considered and 500 trials are generated in each scenario. In the experiments, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ is generated by a Gaussian distribution with unit variance. The support of the vector $\tilde{\boldsymbol{s}}_i$ is generated by a uniform random variable from 1 to N. The nonzero elements of $\tilde{\boldsymbol{S}}$ are found from a unit variance Laplace distribution. The elements of $\tilde{\boldsymbol{A}}$ are Gaussian distributed with unit variance. The observation matrix $\boldsymbol{Y}$ and the matrix $\tilde{\boldsymbol{X}}$ are $\boldsymbol{\Phi}\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$, respectively. In noise-free cases, the number of missed elements in the support of $\hat{\boldsymbol{X}}$ is found by

$$\alpha = \frac{\left|\mathrm{Supp}\left(\tilde{\boldsymbol{X}}\right) - \mathrm{Supp}\left(\hat{\boldsymbol{X}}\right)\right|}{\left|\mathrm{Supp}\left(\tilde{\boldsymbol{X}}\right)\right|}. \tag{42}$$

The following algorithms are used for the comparison purpose.

- CS-MUSIC, proposed in [52].[1]
- MFOCUSS, the regularized M-FOCUSS proposed in [53].[2] We set its p-norm $p = 0.8$, as suggested by the authors.
- MFOCUSS with $p = 1$.

[1] http://bispl.weebly.com/compressive-music.html
[2] http://dsp.ucsd.edu/ zhilin/MFOCUSS.m

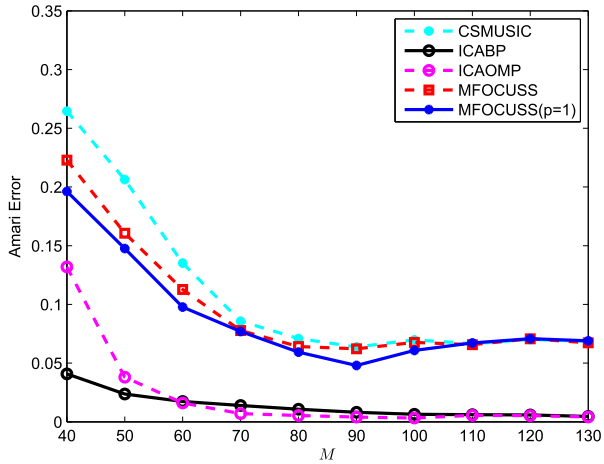

FIGURE 4. Mean of α in terms of number of measurements M.

In order to evaluate the estimation accuracy of $\boldsymbol{A}$, Amari error is used which is defined as

$$\text{Amari error} = \sum_{i=1}^{L}\left(\sum_{j=1}^{L}\frac{|\boldsymbol{H}_{(i,j)}|}{\max_k |\boldsymbol{H}_{(i,k)}|} - 1\right) + \sum_{j=1}^{L}\left(\sum_{i=1}^{L}\frac{|\boldsymbol{H}_{(i,j)}|}{\max_k |\boldsymbol{H}_{(k,j)}|} - 1\right). \tag{43}$$

where the matrix $\boldsymbol{H}$ equals $\tilde{\boldsymbol{A}}\hat{\boldsymbol{A}}^{-1}$. Since the traditional MMV algorithms do not estimate $\boldsymbol{A}$, one of the successful BSS algorithms called EBM in [54] is used to obtain $\hat{\boldsymbol{A}}$ from $\hat{\boldsymbol{X}}$.

A. RECOVERY PERFORMANCE IN TERMS OF NUMBER OF MEASUREMENTS

The motivation of this experiment is to evaluate the performance of the algorithms in terms of number of measurements. In this experiment, we generate 6 independent sources $\tilde{\boldsymbol{S}} \in \mathbb{R}^{200\times 6}$ which are mixed by a 6×6 mixing matrix $\tilde{\boldsymbol{A}}$. The mixture $\tilde{\boldsymbol{X}}$ is measured by the measurement matrix $\boldsymbol{\Phi}$ results in the observation vector $\boldsymbol{Y}$. Each column of $\tilde{\boldsymbol{S}}$ is 10 sparse.

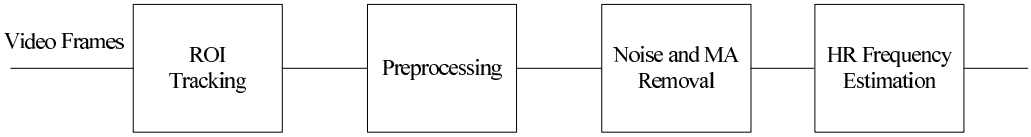

FIGURE 5. General steps for HR estimation.

The locations of the nonzero elements of each column are picked from a discrete uniform random variable from 1 to 200. The values of the nonzero elements are from a Laplace distribution with unit variance.

Fig. 2 and Fig. 3 show the mean and median of α. It can be seen that the proposed algorithms outperform especially for a low number of samples. Fig. 4 shows the Amari errors of the algorithms. It can be seen that the proposed algorithms have better estimation of the mixing matrix in comparison with traditional methods.

B. REAL DATA

In this experiment, the algorithms are evaluated by real data, i.e., heart rate (HR) estimation from face video recording. For this purpose, we briefly describe rPPG signal estimation. RPPG enables contact-less measuring of human cardiac activities by color variations on human skin using an RGB camera [55]. The main steps of rPPG and HR estimation are shown in Fig. 5. In ROI tracking step, an ROI including skin, e.g. face, is chosen and tracked over the time. In preprocessing step, three one-dimension signals in time are extracted from RGB channels where each element is the average of ROI's pixels in the corresponding frame. These three signal include rPPG signal, noise and motion artifacts. Then noise and motion artifacts are suppressed from the rPPG signal. In the final stage, the HR is estimated using spectrum estimation methods. The proposed approaches are applied after the preprocessing stage.

The reason that rPPG can be measured by a recording is that the blood's hemoglobin absorbs light differently than other tissues over time. When arterial blood volume changes during the cardiac cycles, light absorption of the human skin fluctuates [56]. RPPG captures the color variations in time during the recording. HR can be estimated by recording tiny color variations along with minor light intensity variation of the skin.

Since the target of this paper is not ROI tracking, we used face recording of a person who sat still in front of the camera. The frame rate was 30 frames per second (fps). The person's forehead was selected as ROI. After tracking ROI, in the preprocessing step, three one-dimensional signals from RGB channels were estimated. In order to suppress noise and motion artifacts, the signals were filtered using a bandpass filter which kept the coefficients in the range of 40 beats per minute (bpm) to 180 bpm. In this experiment, we used partially known support assumption by removing the coefficients in the range of [0 - 0.6] Hz and [3 - 15] Hz. Matrix $\boldsymbol{\Phi}$ is the DCT transformation matrix including the columns

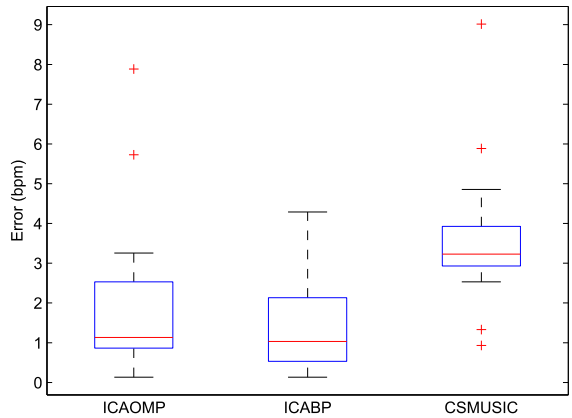

FIGURE 6. The error of heart rate estimation.

corresponds to [0.6 - 3] Hz. A pulse oximeter measuring heart rate frequency was used as a reference.

The error is simply obtained by the absolute value of the difference between the estimated heart rate by the algorithms and that measured by the pulse oximeter. Fig. 6 shows the error of heart rate estimation by the algorithms. The red line in the middle of each box shows the median of the error. As it can be seen in Fig. 6, the proposed algorithms outperformed CSMUSIC and TMSBL. Since MFOCUSS algorithm failed in the experiment, the corresponding results are removed for a clearer presentation.

VI. CONCLUSION

In this paper, the MMV problem is addressed in such a case that the solution matrix has linear mixture structure. The solution model structure is applicable in practical scenarios like PPG or rPPG signal estimation. The solution is decomposed into mixing matrix $\boldsymbol{A}$ and matrix $\boldsymbol{S}$ with independent columns. The key idea of this model is that the matrix $\boldsymbol{S}$ can be much sparser than the mixture $\boldsymbol{X}$. Based on this model, two algorithms are proposed, one based on OMP and the other one based on basis pursuit. The conditions which guarantee the exact recovery of the solution are derived. It is shown that the estimation accuracy of matrix $\boldsymbol{A}$ is significant in the sparse recovery performance. Experimental results show that the proposed methods can improve the performance when a few measurements are available.

APPENDIX A
PROOF OF LEMMA 1

Since $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$, each row of matrix $\boldsymbol{X}$ can be written in terms of $\boldsymbol{S}$ and the full-rank matrix $\boldsymbol{A}$ as

$$\boldsymbol{x}^i = \boldsymbol{s}^i \boldsymbol{A}, \tag{44}$$

where $\boldsymbol{x}^i$ and $\boldsymbol{s}^i$ are the i'th rows of $\boldsymbol{X}$ and $\boldsymbol{S}$ respectively. Based on (44), one can write

$$\text{Supp}(\boldsymbol{X}) \subseteq \bigcup_{i=1}^{L} \text{Supp}(\boldsymbol{s}_i). \tag{45}$$

The cardinality of the right hand side of (45) is upper bounded by

$$|\text{Supp}(\boldsymbol{X})| \leq \left| \bigcup_{i=1}^{L} \text{Supp}(\boldsymbol{s}_i) \right| \leq \sum_{i=1}^{L} |\text{Supp}(\boldsymbol{s}_i)|. \tag{46}$$

The upper bound is given by (46).

For the lower bound we start with assuming that there is $\boldsymbol{X}$ such that

$$\mathcal{R}(\boldsymbol{X}) < \max_i \{r_i : i \in \{1, \cdots L\}\}.$$

This assumption means that

$$\exists i;\, i \in \bigcup_{k=1}^{L} \text{Supp}(\boldsymbol{s}_k) \quad \text{s.t.}\, \boldsymbol{x}^i = \sum_{j=1}^{L} \boldsymbol{s}^i \boldsymbol{A} = 0. \tag{47}$$

This means that $\text{Ker}(\boldsymbol{A}^T) \neq \emptyset$. Since $\boldsymbol{A}$ is square and full-rank, it contradicts with $\text{Ker}(\boldsymbol{A}^T) \neq \emptyset$. This completes the Lemma.

APPENDIX B
PROOF OF THEOREM 1

Assume that there exists an $\check{\boldsymbol{X}}$ having a smaller number of nonzero rows than $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$, i.e. $\mathcal{R}(\check{\boldsymbol{X}}) < \mathcal{R}(\tilde{\boldsymbol{X}})$. One can write

$$\boldsymbol{\Phi}\tilde{\boldsymbol{X}} - \boldsymbol{\Phi}\check{\boldsymbol{X}} = \boldsymbol{0},$$

or,

$$\boldsymbol{\Phi}\tilde{\boldsymbol{S}} - \boldsymbol{\Phi}\check{\boldsymbol{X}}\tilde{\boldsymbol{A}}^{-1} = \boldsymbol{0}.$$

This means that $\tilde{\boldsymbol{S}} - \check{\boldsymbol{X}}\tilde{\boldsymbol{A}}^{-1} \in \ker(\boldsymbol{\Phi})$. Let $\boldsymbol{V}$ be $\tilde{\boldsymbol{S}} - \check{\boldsymbol{X}}\tilde{\boldsymbol{A}}^{-1}$. Therefore, based on the condition (8),

$$|\text{Supp}(\boldsymbol{v}_i)| \geq \max_i |\text{Supp}\,(\tilde{\boldsymbol{s}}_i)| + \sum_{i=1}^{L} |\text{Supp}\,(\tilde{\boldsymbol{s}}_i)|, \tag{48}$$

where $\boldsymbol{v}_i$ is the i'th column of $\boldsymbol{V}$. Therefore, each column of $\check{\boldsymbol{X}}$ must have more than $\sum_{i=1}^{L} |\text{Supp}\,(\tilde{\boldsymbol{s}}_i)|$ nonzero elements. On the other hand, based on the assumption, $\mathcal{R}(\check{\boldsymbol{X}})$ is less than $\mathcal{R}(\tilde{\boldsymbol{X}})$. Therefore, according to Lemma 1, the inequality $\mathcal{R}(\check{\boldsymbol{X}}) < \sum_{i=1}^{L} |\text{Supp}\,(\tilde{\boldsymbol{s}}_i)|$ holds. This is a contradiction which completes the theorem.

APPENDIX C
PROOF OF THEOREM 2

Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ and $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ denote the solution of $(\boldsymbol{P}_1)$.

$$\boldsymbol{Y} = \boldsymbol{\Phi}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}} = \boldsymbol{\Phi}\check{\boldsymbol{S}}\tilde{\boldsymbol{A}}. \tag{49}$$

We multiply both sides of (49) with $\tilde{\boldsymbol{A}}^{-1}$,

$$\boldsymbol{\Phi}\tilde{\boldsymbol{S}} = \boldsymbol{\Phi}\check{\boldsymbol{S}}.$$

For each column we can write,

$$\boldsymbol{\Phi}\tilde{\boldsymbol{s}}_i = \boldsymbol{\Phi}\check{\boldsymbol{s}}_i. \tag{50}$$

Let $\boldsymbol{V} \in \text{Ker}\,(\boldsymbol{\Phi})$ and $\tilde{\boldsymbol{X}} + \boldsymbol{V} = \check{\boldsymbol{X}}$. Since $\boldsymbol{V} \in \text{Ker}(\boldsymbol{\Phi})$, then, $\boldsymbol{V}' = \boldsymbol{V}\tilde{\boldsymbol{A}}^{-1} \in \text{Ker}(\boldsymbol{\Phi})$. We can write

$$\bar{\boldsymbol{s}}_i + \boldsymbol{v}'_i = \check{\boldsymbol{s}}_i, \tag{51}$$

where $\boldsymbol{v}'_i$ is the i'th column of $\boldsymbol{V}'$. The summation of both sides of (51) over $1 \leq i \leq N$.

$$\begin{aligned}
\sum_{i=1}^{N} \|\tilde{\boldsymbol{s}}_i\|_1 &\leq \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{s}}_i + \boldsymbol{v}'_i\big|_{B_i} \right\|_1 + \left\| \boldsymbol{v}'_i\big|_{B_i} \right\|_1 \\
&\leq \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{s}}_i + \boldsymbol{v}'_i\big|_{B_i} \right\|_1 + \left\| \boldsymbol{v}'_i\big|_{B_i^c} \right\|_1 \\
&< \sum_{i=1}^{N} \left\| \check{\boldsymbol{s}}_i\big|_{B_i} \right\|_1 + \left\| \boldsymbol{v}'_i\big|_{B_i^c} \right\|_1 \\
&< \sum_{i=1}^{N} \left\| \check{\boldsymbol{s}}_i \right\|_1,
\end{aligned} \tag{52}$$

where $\boldsymbol{v}'_i\big|_{B_i}$ and $\boldsymbol{v}'_i\big|_{B_i^c}$ are the restrictions of $\boldsymbol{v}'_i$ on B_i and its complement B_i^c, respectively. The theorem is completed by (52).

APPENDIX D
PROOF OF THEOREM 3

Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ and $\check{\boldsymbol{X}} = \{\check{\boldsymbol{S}}, \check{\boldsymbol{A}}\}$, and denote the solutions of $(\boldsymbol{P}_1)$ where $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ and $\check{\boldsymbol{X}} = \check{\boldsymbol{S}}\check{\boldsymbol{A}}$,

$$\boldsymbol{Y} = \boldsymbol{\Phi}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}} = \boldsymbol{\Phi}\check{\boldsymbol{S}}\check{\boldsymbol{A}}. \tag{53}$$

By multiplying $\check{\boldsymbol{A}}^{-1}$ to both sides of (53), we obtain

$$\boldsymbol{Y} = \boldsymbol{\Phi}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\check{\boldsymbol{A}}^{-1} = \boldsymbol{\Phi}\check{\boldsymbol{S}}. \tag{54}$$

Since $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ is a solution to the problem, one can write using the fact that $\text{Supp}\left(\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\check{\boldsymbol{A}}^{-1}\right) \subset \bigcup_{i=1}^{N} B_i$,

$$\begin{aligned}
\left\|\tilde{\boldsymbol{S}}\right\|_1 &\leq \left\|\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\check{\boldsymbol{A}}^{-1} + \boldsymbol{V}'\big|_B\right\|_1 + \left\|\boldsymbol{V}'\big|_B\right\|_1 \\
&= \left\|\check{\boldsymbol{S}}\big|_B\right\|_1 + \left\|\boldsymbol{V}'\big|_B\right\|_1 \\
&< \left\|\check{\boldsymbol{S}}\big|_B\right\|_1 + \left\|\boldsymbol{V}'\big|_{B^c}\right\|_1 = \left\|\check{\boldsymbol{S}}\right\|_1,
\end{aligned} \tag{55}$$

where $B = \bigcup_{i=1}^{N} B_i$ and $\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\check{\boldsymbol{A}}^{-1} + \boldsymbol{V}' = \check{\boldsymbol{S}}$.

The inequality in (54) shows that the solution is the sparsest, which completes the theorem.

APPENDIX E
PROOF OF THEOREM 4

Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ and $\{\check{\boldsymbol{S}}, \check{\boldsymbol{A}}\}$ denote the solutions.

$$\begin{aligned}
\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}} + \check{\boldsymbol{V}} &= \check{\boldsymbol{S}}\check{\boldsymbol{A}} \\
\tilde{\boldsymbol{S}} + \boldsymbol{V} &= \check{\boldsymbol{S}}\check{\boldsymbol{A}}\tilde{\boldsymbol{A}}^{-1} = \bar{\boldsymbol{S}}
\end{aligned} \tag{56}$$

Then, one can write

$$\sum_{i=1}^{N} \|\tilde{s}_i\|_1 \leq \sum_{i=1}^{N} \left\| \tilde{s}_i + v_i\big|_{B_i} \right\|_1 + \left\| v_i\big|_{B_i} \right\|_1 \leq \sum_{i=1}^{N} \left\| \bar{s}_i\big|_{B_i} \right\|_1 + \left\| v_i\big|_{B_i} \right\|_1, \tag{57}$$

using the definition of null space property, we can write

$$\begin{aligned}\sum_{i=1}^{N} \|\tilde{s}_i\|_1 &< \sum_{i=1}^{N} \left\| \bar{s}_i\big|_{B_i} \right\|_1 + (1-\tau) \left\| v_i\big|_{B_i^c} \right\|_1 \\ &\leq \sum_{i=1}^{N} \|\bar{s}_i\|_1 - \tau \left\| v_i\big|_{B_i^c} \right\|_1 \\ &\leq \sum_{i=1}^{N} (1+\alpha) \left\| \check{s}_i \right\|_1 - \tau \left\| v_i\big|_{B_i^c} \right\|_1 \\ &\leq \sum_{i=1}^{N} \left\| \check{s}_i \right\|_1 .\end{aligned} \tag{58}$$

The latter inequality is valid if

$$\sum_{i=1}^{N} \alpha \left\| \check{s}_i \right\|_1 < \sum_{i=1}^{N} \tau \left\| v_i\big|_{B_i^c} \right\|_1 \tag{59}$$

or

$$\frac{\alpha}{N} \sum_{i=1}^{N} \left\| \check{s}_i \right\|_1 < \tau \left\| v_i\big|_{B_i^c} \right\|_1, \text{ for all } i. \tag{60}$$

The minimum value for each element of $v_i\big|_{B_i^c}$ is δ, then, $\left\| v_i\big|_{B_i^c} \right\|_1$ is larger than $\delta\|\boldsymbol{r}\|_\infty$. The term $\sum_{i=1}^{N} \left\| \check{s}_i \right\|_1$ is less than $\sum_{i=1}^{N} \left\| \boldsymbol{\Phi}^+ \boldsymbol{y}_i \right\|_1$. The condition in (60) is satisfied if

$$\frac{\alpha}{N} \sum_{i=1}^{N} \left\| \boldsymbol{\Phi}^+ \boldsymbol{y}_i \right\|_1 < \tau\delta\|\boldsymbol{r}\|_\infty. \tag{61}$$

This completes the theorem.

APPENDIX F
PROOF OF THEOREM 5

By multiplying the inverse of $\tilde{\boldsymbol{A}}$ to both sides of (12), one can obtain

$$\boldsymbol{Y}' = \boldsymbol{\Phi}\boldsymbol{S}. \tag{62}$$

This leads to L independent SMV problem. Using the result from [24] for the exact recovery condition for OMP algorithm, we can write for each column that

$$\left\| \boldsymbol{\Phi}_{B_i}^{+} \boldsymbol{\phi}_j \right\|_1 < 1, \quad \text{for all } i,\ j \notin B_i. \tag{63}$$

This completes the proof of the theorem.

APPENDIX G
PROOF OF THEOREM 6

Let $\hat{\boldsymbol{A}}$ is the estimate of matrix $\boldsymbol{A}$. By multiplying the inverse of $\hat{\boldsymbol{A}}$ to both sides of (12), it can be written that

$$\boldsymbol{Y}'' = \boldsymbol{\Phi}\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\hat{\boldsymbol{A}}^{-1}. \tag{64}$$

The matrix each columns of $\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\hat{\boldsymbol{A}}^{-1}$ is supported on the union of the supports of the columns of $\tilde{\boldsymbol{S}}$. If we apply ICAOMP column by column, the exact support can be recovered if $\left\| \boldsymbol{\Phi}_B^{+} \boldsymbol{\phi}_j \right\|_1 < 1$.

REFERENCES

[1] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 14–20, Mar. 2008.
[2] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[3] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1902–1910, Aug. 2015.
[4] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.
[5] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process. (SSP)*, Aug. 2007, pp. 181–184.
[6] T. Strohmer and B. Friedlander, "Compressed sensing for MIMO radar—Algorithms and performance," in *Conf. Rec. 43rd Asilomar Conf. Signals, Syst. Comput.*, 2009, pp. 464–468.
[7] Y. Yu, A. P. Petropulu, and H. V. Poor, "MIMO radar using compressive sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 146–163, Feb. 2010.
[8] S. Goginneni and A. Nehorai, "Target estimation using sparse modeling for distributed MIMO radar," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5315–5325, Nov. 2011.
[9] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, 2006, pp. 134–142.
[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[11] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[12] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
[13] A. Shukla and A. Majumdar, "Row-sparse blind compressed sensing for reconstructing multi-channel EEG signals," *Biomed. Signal Process. Control*, vol. 18, pp. 174–178, Apr. 2015.
[14] Z. Zhang and B. D. Rao, "Sparse signal recovery in the presence of correlated multiple measurement vectors," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 3986–3989.
[15] S. Liu, F. Yang, J. Song, and Z. Han, "Block sparse Bayesian learning-based NB-IoT interference elimination in LTE-Advanced systems," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4559–4571, Oct. 2017.
[16] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.
[17] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
[18] S. Liu, F. Yang, W. Ding, X. Wang, and J. Song, "Two-dimensional structured-compressed-sensing-based NBI cancelation exploiting spatial and temporal correlations in MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9020–9028, Nov. 2016.
[19] S. Liu, F. Yang, X. Wang, J. Song, and Z. Han, "Structured-compressed-sensing-based impulsive noise cancelation for mimo systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6921–6931, Aug. 2017.

[20] S. H. Fouladi, S.-E. Chiu, B. D. Rao, and I. Balasingham, "Recovery of independent sparse sources from linear mixtures using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6332–6346, Dec. 2018.

[21] S. H. Fouladi and I. Balasingham, "Recovery of linearly mixed sparse sources from multiple measurement vectors using L_1-minimization," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Rome, Italy, Sep. 2018, pp. 563–567.

[22] D. Needell, J. Tropp, and R. Vershynin, "Greedy signal recovery review," in *Proc. 42nd Asilomar Conf. Signals, Syst. Comput.*, 2008, pp. 1048–1050.

[23] L. Li, X. Huang, and J. A. K. Suykens, "Signal recovery for jointly sparse vectors with different sensing matrices," *Signal Process.*, vol. 108, pp. 451–458, Mar. 2015.

[24] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[25] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[26] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1-minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[27] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[28] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[29] I. Rish and G. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*. Boca Raton, FL, USA: CRC Press, 2014.

[30] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.

[31] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[32] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1-problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.

[33] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York, NY, USA: Springer, 2011, pp. 185–212.

[34] M. A. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 2, Sep. 2005, p. II-782.

[35] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

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

[37] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[38] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[39] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput. Conf. Rec.*, 1993, pp. 40–44.

[40] S. N. Vitaladevuni, P. Natarajan, R. Prasad, and P. Natarajan, "Efficient orthogonal matching pursuit using sparse random projections for scene and video classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2312–2319.

[41] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.

[42] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.

[43] J. Bobin, Y. Moudden, J. L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 409–412, Jul. 2006.

[44] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2662–2674, Nov. 2007.

[45] M. J. Fadili, J.-L. Starck, J. Bobin, and Y. Moudden, "Image decomposition and separation using sparse representations: An overview," *Proc. IEEE*, vol. 98, no. 6, pp. 983–994, Jun. 2010.

[46] L. Zhao, J. Xu, J. Ding, A. Liu, and L. Li, "Direction-of-arrival estimation of multipath signals using independent component analysis and compressive sensing," *PLoS ONE*, vol. 12, no. 7, p. e0181838, 2017.

[47] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[49] J. Chen and X. Ho, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[50] K. E. Hild, D. Erdogmus, and J. C. Principe, "Experimental upper bound for the performance of convolutive source separation methods," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 627–635, Feb. 2006.

[51] Y. Gao, J. Peng, S. Yue, and Y. Zhao, "On the null space property of l_q-minimization for $0< q \leq 1$ in compressed sensing," *J. Function Spaces*, vol. 2015, Mar. 2015, Art. no. 579853.

[52] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 278–301, Jan. 2012.

[53] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[54] X.-L. Li and T. Adalı, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.

[55] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[56] T. Aoyagi, N. Kobayashi, and T. Sasaki, "Apparatus for determining the concentration of a light-absorbing material in blood," U.S. Patent 4 832 484, May 23, 1989.

Authors' photographs and biographies not available at the time of publication.