Recovery of Linearly Mixed Sparse Sources From Multiple Measurement Vectors Using L1-Minimization

1st Seyyed Hamed Fouladi
Department of Electronic Systems
Norwegian Institute of Science and Technology
Trondheim, Norway
hamed.fouladi@ntnu.no

2nd Ilangko Balasingham
Department of Electronic Systems
Norwegian Institute of Science and Technology
Trondheim, Norway
ilangkob@iet.ntnu.no

***Abstract*—Multiple measurement vector (MMV) enables joint sparse recovery which can be applied in wide range of applications. Traditional MMV algorithms assume that the solution has independent columns or correlation among the columns. This assumption is not accurate for applications like signal estimation in photoplethysmography (PPG). In this paper, we consider a structure for the solution matrix decomposed into a sparse matrix with independent columns and a square mixing matrix. Based on this structure, we find the uniqueness condition for ℓ_1 minimization. Moreover, an algorithm is proposed that provides a new cost function based on the new structure. It is shown that the new structure increases the recovery performance especially in low number of measurements.**

I. Introduction

There are many signals of interest that can be well approximated in a few number of nonzero elements in a certain domain. Compressed sensing (CS) deals with sparse recovery from single measure vector (SMV). Multiple measurement vector (MMV) is the generalization of CS where multiple measurement vectors are available. This problem has attracted many applications such as image processing [1], [2], biomedical signal processing [3]–[5], speech processing [6] and radar signal processing [7]. The MMV problem can be formulated as

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X}, \tag{1}$$

where $\boldsymbol{Y} \in \mathbb{R}^{M\times L}$, $\boldsymbol{X} \in \mathbb{R}^{N\times L}$ and $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ are an observation matrix, a row sparse matrix and a known dictionary where $M \ll N$. In order to find the solution with the lowest number of nonzero rows, the following optimization problem is considered,

$$\min_{\boldsymbol{X}} \mathcal{R}(\boldsymbol{X}) \quad \text{s.t. } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X}, \tag{2}$$

where $\mathcal{R}(\boldsymbol{X})$ shows the number of nonzero rows of $\boldsymbol{X}$.

Many algorithms have been proposed to solve MMV and SMV problems. For example, one of the approaches was sparse representation via a greedy search like orthogonal matching pursuits (OMP) [8], [9]. In [10], [11], it was proved that OMP can result in the sparsest solution under a specific condition. There were some works to extend the greedy algorithms for MMV problem like [12]. Beside the OMP, there are algorithms such as FOCUSS [13] and T-MSBL [14] whose cost functions are not convex but practically perform very well. The shortcoming of these algorithms was that they had local minima [15]. Another well-known approach was minimization of ℓ_1 norm to obtain the sparse solution like basis pursuit and LASSO [16], [17]. The ℓ_1 minimization algorithms attracted attentions due to their convex property and easy implementation. In [18], a condition was presented to guarantee the solution to the ℓ_1 minimization to be the sparsest solution.

In addition to the importance of the cost function, a suitable assumption on the structure of the solution matrix could improve the performance of the sparse recovery [14]. In [14], the linear correlation among the columns of the solution was modeled using an unknown covariance matrix. However, this model cannot well describe the solution matrix for the applications like PPG signal extraction. In this application, matrix $\boldsymbol{X}$ includes the mixtures of the PPG signal and undesired signals which are independent. Therefore, the dependence among the columns of $\boldsymbol{X}$ cannot be well captured by covariance matrix.

Motivated to find the solution for this problem, in this paper, we propose to decompose the solution matrix $\boldsymbol{X}$ into a mixing matrix $\boldsymbol{A}$ and a matrix $\boldsymbol{S}$ with independent and sparse columns; i.e., $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$. The matrix $\boldsymbol{A}$ captures the dependence among the columns of $\boldsymbol{X}$. Using this model, the covariance of each row of $\boldsymbol{X}$ is obtained by $\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{\Gamma}\boldsymbol{A}$, where $\boldsymbol{\Gamma}$ is a diagonal matrix whose elements are the variance of the columns of $\boldsymbol{S}$. The covariance matrix $\boldsymbol{B}$ can be obtained by second order statistics of $\boldsymbol{X}$. The successful algorithm, T-MSBL, exploited matrix $\boldsymbol{B}$ to model the structure of $\boldsymbol{X}$. However, the independent components in matrix $\boldsymbol{S}$ may not be obtained using the covariance matrix $\boldsymbol{B}$. Considering the independence assumption on sources gives the capability to estimate the independent components $\boldsymbol{S}$ and the mixture

This work was funded by the Research Council of Norway project the Medical Sensing, Localization, and Communication using Ultra Wideband Technology (MELODY) (Phase II) under the contract no. 225885.

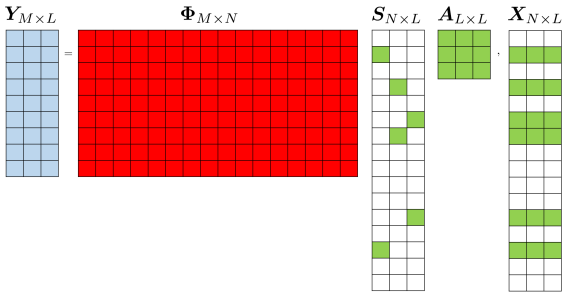

Fig. 1: Visual representation of (1) and (3).

matrix $\boldsymbol{A}$ using matrix $\boldsymbol{X}$ [19]. Using this model, the problem based on minimizing the ℓ_1 norm of the columns of $\boldsymbol{S}$ is formulated. We derive a condition that the solution to the optimization problem is equal to $\tilde{\boldsymbol{X}}$ with the least number of nonzero rows. In addition, an algorithm is proposed which solves the problem based on coordinate descent. Finally, some simulation results are provided to evaluate the efficiency of the proposed method compared to traditional algorithms.

The following notations are for a vector $\boldsymbol{s}$, a matrix $\boldsymbol{A}$ and a set K. $\|\boldsymbol{s}\|_0, \|\boldsymbol{s}\|_1, \|\boldsymbol{s}\|_2, \|\boldsymbol{A}\|_{\mathcal{F}}$ denote the ℓ_0 norm of the vector $\boldsymbol{s}$, the ℓ_1 norm of $\boldsymbol{s}$, the ℓ_2 norm of $\boldsymbol{s}$ and the Frobenius norm of the matrix $\boldsymbol{A}$, respectively. $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{R}(\boldsymbol{s})$ denote the number of nonzero rows in the matrix $\boldsymbol{A}$ and the number of nonzero elements in the vector $\boldsymbol{s}$ which equals to $\|\boldsymbol{s}\|_0$, respectively. $\boldsymbol{A}_{(i,j)}$, and $\boldsymbol{s}_{(i)}$ denote the element that lies in the i-th row and the j-th column of $\boldsymbol{A}$ and i-th element of $\boldsymbol{s}$ respectively. $\boldsymbol{A}^T$ denotes the transpose of $\boldsymbol{A}$. $\boldsymbol{A}_{(\vec{i})}$ and $\boldsymbol{A}_{(\downarrow j)}$ denote the i'th row of $\boldsymbol{A}$ and the j'th column of $\boldsymbol{A}$. If $\boldsymbol{s} \in \mathbb{R}^N$, the support of $\boldsymbol{s}$, Supp$(\boldsymbol{s})$, denotes $\{i \in \{1, \cdots, N\} : \boldsymbol{s}_{(i)} \neq 0\}$. If $\boldsymbol{A} \in \mathbb{R}^{N\times L}$, Supp$(\boldsymbol{A})$ is $\{i \in \{1, \cdots, N\} : \boldsymbol{A}_{(\vec{i})} \neq \boldsymbol{0}\}$. $|K|$ is the cardinality of the set K. The vector $\boldsymbol{s}$ equals to difference between two nonnegative variables $\boldsymbol{s}_+, \boldsymbol{s}_- \succeq \boldsymbol{0}$; i.e., $\boldsymbol{s}_+ - \boldsymbol{s}_-$.

II. Problem Formulation

In applications like PPG or remote PPG signal extraction the desired signal is sparse and independent from noise and artifacts. The independent components are mixed and observed through some channels or sensors. Each column of $\boldsymbol{S}$ is considered as a source. Then, these sources are mixed and result in matrix $\boldsymbol{X}$. Mathematically speaking, each column of $\boldsymbol{X}$ represents a mixture of these signals; i.e.,

$$\boldsymbol{X}_{N\times L} = \boldsymbol{S}_{N\times L}\boldsymbol{A}_{L\times L}, \tag{3}$$

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. The matrix structure (1) and the problem (3) is visually shown in Fig 1. The figure shows the case where the columns of $\boldsymbol{S}$ have no common support. In such a case, $\boldsymbol{S}$ is much sparser than $\boldsymbol{X}$. The following proposition describes the number of nonzero rows of $\boldsymbol{X}$ in terms of the columns of the sources $\boldsymbol{S}_{(\downarrow i)}, i \in \{1, \cdots, L\}$.

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

$$\max_i \left\{\left|\text{Supp}(\boldsymbol{S}_{(\downarrow i)})\right|\right\} \leq \mathcal{R}(\boldsymbol{X}) \leq \sum_{i=1}^{L} \left|\text{Supp}\left(\boldsymbol{S}_{(\downarrow i)}\right)\right|. \tag{4}$$

Proof. See Appendix A. □

The lower bound in Proposition 1 shows that the number of nonzero rows of $\boldsymbol{X}$ is always larger than the number of nonzero elements of the columns of $\boldsymbol{S}$. Intuitively, this means that if the matrix $\boldsymbol{A}$ is estimated correctly, it is better to minimize the number of nonzero elements of $\boldsymbol{S}$ than the number of nonzero rows of $\boldsymbol{X}$. Therefore, a new cost function is defined to obtain the sparsest $\boldsymbol{S}$ unlike the previous MMV approaches which search for row sparse matrix $\boldsymbol{X}$. To find the sparsest $\boldsymbol{S}$, the number of nonzero elements of the columns of $\boldsymbol{S}$ need to be minimized which leads to,

$$(\boldsymbol{P}_0): \min_{\boldsymbol{S},\boldsymbol{A}} \sum_{i=1}^{L} \|\boldsymbol{S}_{(\downarrow i)}\|_0, \text{ subject to } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}, \boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}. \tag{5}$$

Due to non-convexity of ℓ_0 norm, we change the problem in (5) to the following ℓ_1 norm minimization problem,

$$(\boldsymbol{P}_1): \min_{\boldsymbol{S},\boldsymbol{A}} \sum_{i=1}^{L} \|\boldsymbol{S}_{(\downarrow i)}\|_1, \text{ subject to } \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}, \boldsymbol{X} = \boldsymbol{S}\boldsymbol{A} \text{ and } \left\|\boldsymbol{A}_{(\vec{i})}\right\|_2 = 1. \tag{6}$$

Since if $\{\boldsymbol{S}, \boldsymbol{A}\}$ is a solution, $\{\frac{1}{b}\boldsymbol{S}, b\boldsymbol{A}\}$ will also be a solution for a bounded constant b. The last constraint in (6) is applied on the rows $\boldsymbol{A}$ to avoid the scaling ambiguity. Note that any scale or permutation ambiguity is not important because they result in the same matrix $\boldsymbol{X}$.

Remark 1. *The solution of the minimization problem in (6) is $\{\boldsymbol{S}, \boldsymbol{A}\}$. However, we also call $\boldsymbol{X} = \boldsymbol{S}\boldsymbol{A}$ as the solution because the target of the MMV problem is to find $\boldsymbol{X}$.*

Remark 2. *This paper 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 more information when noise is not contaminated with the signals (See the model in (1).). 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 (6) can be solved with the assumptions in this paper.*

III. RECOVERY GUARANTEE

This section deals with the condition that results in the solution of (6) to be equal to $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ where $\tilde{\boldsymbol{S}}$ has the least number of nonzero elements. Conditions for the both known and unknown $\tilde{\boldsymbol{A}}$ cases are derived.

Theorem 1. *Let* $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ *be the solution to* $(\boldsymbol{P}_1)$ *where* $\text{Supp}(\tilde{\boldsymbol{S}}_{(\downarrow i)}) = B_i$ *where* $B_i \subset \{1, \cdots, L\}$ *and* $|B_i| = r_i$. *A sufficient condition for* $\tilde{\boldsymbol{X}}$ *to be the unique solution to* $(\boldsymbol{P}_1)$ *when matrix* $\tilde{\boldsymbol{A}}$ *is known, is that*

$$\begin{aligned} &\|\boldsymbol{\Phi}_{B_I}^{+}\boldsymbol{\Phi}_j\|_1 < 1, \forall j \notin B_I, \\ &I = \arg\max_i (r_i), \quad 1 \le i \le L \end{aligned} \tag{7}$$

where $\boldsymbol{\Phi}_{B_i}$ *includes the columns of* $\boldsymbol{\Phi}$ *whose indexes are in the set* B_i *and* $\boldsymbol{\Phi}_{B_i}^{+}$ *is the generalized inverse of* $\boldsymbol{\Phi}_{B_i}$, *i.e.,* $(\boldsymbol{\Phi}_{B_i}^T\boldsymbol{\Phi}_{B_i})^{-1}\boldsymbol{\Phi}_{B_i}^T$.

Proof. See Appendix B. □

Now the condition is presented for the unknown $\tilde{\boldsymbol{A}}$ case in the following theorem.

Theorem 2. *Let* $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ *be the solution to* $(\boldsymbol{P}_1)$ *where* $\text{Supp}(\tilde{\boldsymbol{S}}_{(\downarrow i)}) = B_i$ *and* $B_i \subset \{1, \cdots, N\}$ *where* $|B_i| = r_i$. *A sufficient condition for* $\tilde{\boldsymbol{X}}$ *to be the unique solution by solving the minimization problem in* $(\boldsymbol{P}_1)$ *when matrix* $\tilde{\boldsymbol{A}}$ *is full-rank, is that*

$$\|\boldsymbol{\Phi}_B^{+}\boldsymbol{\Phi}_j\|_1 < 1, \quad \forall j \notin B, \tag{8}$$

irrespective of the estimation of $\boldsymbol{A}$, *where* $B = \bigcup_{i=1}^{L} B_i$ *and* $\boldsymbol{\Phi}_B$ *includes the columns of* $\boldsymbol{\Phi}$ *whose indexes are in the set* B.

Proof. See Appendix C. □

The condition in Theorem 2 guarantees that the solution obtained by minimizing (6) is also the sparsest solution.

IV. METHOD

In this section, we briefly describe the proposed algorithm that estimates both $\boldsymbol{S}$ and $\boldsymbol{A}$. The following cost function is defined using the idea of LASSO problem,

$$\min_{\boldsymbol{S},\boldsymbol{A}} \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\|_2^2 + \lambda \sum_{i=1}^{L} \left\|\boldsymbol{S}_{(\downarrow i)}\right\|_1 \text{ s.t. } \left\|\boldsymbol{A}_{(\vec{i})}\right\|_2 = 1. \tag{9}$$

To solve (9), an iterative algorithm is developed that estimates $\boldsymbol{A}$ and $\boldsymbol{S}$. In order to update $\boldsymbol{A}$, the following update rule can be used due to the fact that only the term $\|\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{A}\|_2^2$ in (9) depends on $\boldsymbol{A}$,

$$\boldsymbol{A} = \left(\boldsymbol{S}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{S}\right)^{-1}\boldsymbol{S}^T\boldsymbol{\Phi}^T\boldsymbol{Y}. \tag{10}$$

The estimate in (10) may result in a matrix $\boldsymbol{A}$ with the rows not equal to 1. After the estimate, we simply normalize the rows of $\boldsymbol{A}$.

$$\boldsymbol{A}_{(\vec{i})} = \frac{1}{\left\|\boldsymbol{A}_{(\vec{i})}\right\|_2}\boldsymbol{A}_{(\vec{i})} \tag{11}$$

This normalization is done in order to avoid scaling ambiguity which is well-known in BSS problem. By multiplying the inverse of $\boldsymbol{A}$ to both sides of (1), one obtain

$$\boldsymbol{Y}'_{(\downarrow i)} = \boldsymbol{\Phi}\boldsymbol{S}_{(\downarrow i)} \quad \text{for all } i. \tag{12}$$

(12) can be considered as L SMV problem. The cost function in (9) can be modified to

$$\frac{1}{2}\min_{\boldsymbol{S}} \left\|\boldsymbol{Y}'_{(\downarrow i)} - \boldsymbol{\Phi}\boldsymbol{S}_{(\downarrow i)}\right\|_2^2 + \lambda \left\|\boldsymbol{S}_{(\downarrow i)}\right\|_1 \text{ for all } i. \tag{13}$$

where $\boldsymbol{Y}' = \boldsymbol{Y}\boldsymbol{A}^{-1}$ and λ is a regularization parameter. The cost function in (13) can be optimized using traditional ℓ_1 minimization algorithms. Coordinate descent is used which is an efficient procedure in ℓ_1 minimization [20]. Table I lists the steps of the proposed algorithm based on coordinate descent.

TABLE I: The proposed algorithm

Input : $\boldsymbol{\Phi}$, $\boldsymbol{Y}$, λ and ε.
1) Initialization: $k = 0$, $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{S} = \boldsymbol{0}$, $\boldsymbol{R} = \boldsymbol{Y}\boldsymbol{A}^{-1}$.
2) $k = k + 1$, $j = k \bmod (N + 1)$.
3) Obtain the residual $\boldsymbol{R}$ by $\boldsymbol{R}_{(l,i)} = \boldsymbol{Y}'_{(l,i)} - \sum_{k \neq j} \boldsymbol{\Phi}_{(l,k)}\boldsymbol{S}_{(k,i)}$.
4) Obtain univariate OLS solution for all i $s_i = \left(\boldsymbol{\Phi}_{(\downarrow j)}\right)^T \boldsymbol{R}_{(\downarrow i)}$.
5) Update $\boldsymbol{S}_{(i,j)}$ (univariate LASSO solution), $\boldsymbol{S}_{(i,j)} = \text{sign}(s_i)(|s_i| - \lambda)_+$.
6) If $k \bmod (N) = 0$, update $\boldsymbol{A}$ using (10).
7) $\boldsymbol{A}_{(\vec{i})} = \frac{1}{\left\|\boldsymbol{A}_{(\vec{i})}\right\|_2}\boldsymbol{A}_{(\vec{i})}$ for all i.
8) Obtain $\boldsymbol{Y}' = \boldsymbol{Y}\boldsymbol{A}^{-1}$.
9) If the value $\left\|\boldsymbol{R} - \boldsymbol{\Phi}_{(\downarrow j)}\boldsymbol{S}_{(\vec{j})}\right\|_{\mathcal{F}} < \varepsilon$, return $\boldsymbol{S}$ and $\boldsymbol{A}$, otherwise go to step 2.
10) Return $\boldsymbol{S}$ and $\boldsymbol{A}$.

V. EXPERIMENTAL RESULTS

In this section, the performance of the proposed algorithm is evaluated and compared with traditional algorithms. Each experiment includes 500 trials. 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 $\tilde{\boldsymbol{S}}_{(\downarrow i)}$ is generated using a uniform random variable from 1 to N. The nonzero elements of $\tilde{\boldsymbol{S}}$ are from a Laplace distribution with unit variance. The elements of $\tilde{\boldsymbol{A}}$ are from Gaussian distribution with unit variance. The matrices $\boldsymbol{Y}$ and $\tilde{\boldsymbol{X}}$ are $\boldsymbol{\Phi}\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}$ respectively. Note that $\tilde{\boldsymbol{X}}$ shares common support among its column as the assumption for MMV approaches. In order to evaluate the performance, the number of missed elements in the support of $\hat{\boldsymbol{X}}$ is found by

$$\alpha = \frac{\left|\text{Supp}\left(\tilde{\boldsymbol{X}}\right) - \text{Supp}\left(\hat{\boldsymbol{X}}\right)\right|}{\left|\text{Supp}\left(\tilde{\boldsymbol{X}}\right)\right|}. \tag{14}$$

The following algorithms are used for comparison purpose.

- T-MSBL, proposed in [14] (http://sccn.ucsd.edu/~zhang/TMSBL_code.zip).

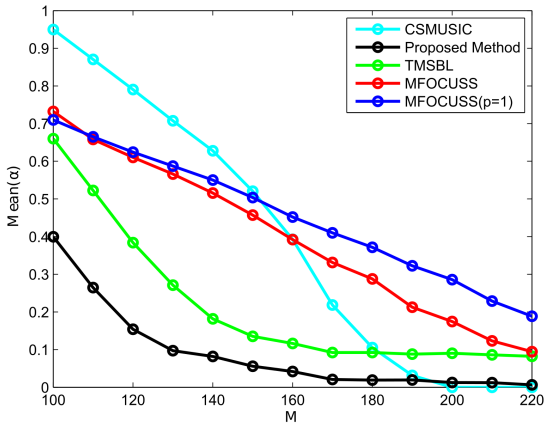

Fig. 2: Mean α in terms of number of measurements.

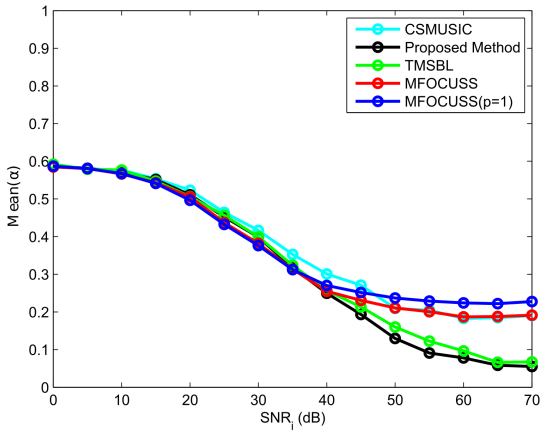

Fig. 3: Mean α in terms of input SNR.

- CS-MUSIC, proposed in [12] (http://bispl.weebly.com/compressive-music.html).
- MFOCUSS, the regularized M-FOCUSS proposed in [13] (http://dsp.ucsd.edu/~zhilin/MFOCUSS.m). We set its p norm $p = 0.8$, as suggested by the authors.
- MFOCUSS with $p = 1$.

In the first experiment, the performance of the algorithms are evaluated with different number of measurements. M is increased from 100 to 220. N and L are 500 and 5 respectively and $\left|\text{Supp}\left(\tilde{\boldsymbol{S}}_{(\downarrow i)}\right)\right|$ is 30 for all i.

Fig. 2 shows the mean of α averaged over all trials. As it can be seen, the performance of all algorithms improves as the number of measurements increases as expected. One interesting observation is that the proposed method outperforms the other algorithms especially in lower number of measurements.

In the next experiment, all algorithms are compared in noisy environment. The values for M, N, and L are 110, 250 and 3 respectively. The matrices $\boldsymbol{S}$, $\boldsymbol{A}$ and $\boldsymbol{\Phi}$ are generated with the same distribution as the previous experiment. The number of nonzero elements of each column of $\boldsymbol{S}$ is 30.

Fig. 3 shows the performance the algorithms improves as the input SNR increases. The input SNR is calculated as $\|\tilde{\boldsymbol{X}}\|_2^2/\|\boldsymbol{N}\|_2^2$. In noisy environment the performance of the proposed algorithm and TMSBL are very similar. Both outperformed other algorithms for the input SNRs were larger than 40 dB.

VI. Conclusion

We addressed the multiple measurement vector (MMV) problem, where the structure of the row-sparse solution matrix $\tilde{\boldsymbol{X}}$ was modeled with multiplication of a mixing matrix and a sparse matrix with independent columns. We derived a condition that guaranteed the ℓ_1 minimization solution to be the sparsest solution. Then, an algorithm was proposed based on coordinate descent where the structure of the solution was taken into account. The experimental results showed the proposed algorithm improved the performance of the sparse recovery.

Appendix

A. Proof of Proposition 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

$$\boldsymbol{X}_{(\vec{i})} = \boldsymbol{S}_{(\vec{i})}\boldsymbol{A}. \tag{15}$$

Based on (15), one can write

$$\text{Supp}(\boldsymbol{X}) \subseteq \bigcup_{i=1}^{L} \text{Supp}(\boldsymbol{S}_{(\downarrow i)}). \tag{16}$$

The cardinality of the right hand side of (16) is upper bounded by

$$|\text{Supp}(\boldsymbol{X})| \leq \left|\bigcup_{i=1}^{L} \text{Supp}(\boldsymbol{S}_{(\downarrow i)})\right| \leq \sum_{i=1}^{L} \left|\text{Supp}(\boldsymbol{S}_{(\downarrow i)})\right|. \tag{17}$$

The upper bound is given by (17).

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

$$\exists i; i \in \bigcup_{k=1}^{L} \text{Supp}(\boldsymbol{S}_{(\downarrow k)}) \text{ s.t. } \boldsymbol{X}_{(\vec{i})} = \sum_{j=1}^{L} \boldsymbol{S}_{(\vec{i})}\boldsymbol{A} = 0. \tag{18}$$

This means that $\text{Ker}(\boldsymbol{A}^T) \neq \{0\}$. Since $\boldsymbol{A}$ is square and full-rank, it contradicts. This completes the Theorem.

B. Proof of Theorem 1

Let $\tilde{\boldsymbol{S}}$ and $\check{\boldsymbol{S}}$ be the solutions of optimization problem $(\boldsymbol{P}_1)$, $\text{Supp}\left(\tilde{\boldsymbol{S}}_{(\downarrow i)}\right) = B_i$ and $\text{Supp}\left(\check{\boldsymbol{S}}_{(\downarrow i)}\right) = B'$. By multiplying $\boldsymbol{A}^{-1}$ to both sides of (1), the following equation is obtained using (3)

$$\boldsymbol{Y}\boldsymbol{A}^{-1} = \boldsymbol{\Phi}\tilde{\boldsymbol{S}} = \boldsymbol{\Phi}\check{\boldsymbol{S}}. \tag{19}$$

For each column, one can write

$$\boldsymbol{\Phi}\tilde{\boldsymbol{S}}_{(\downarrow i)} = \boldsymbol{\Phi}_{B'}\check{\boldsymbol{S}}_{(\downarrow i)}. \tag{20}$$

where $\boldsymbol{A}_3 = \boldsymbol{A}_2\boldsymbol{A}_1^{-1}$. Each column of both sides can be considered as a SMV problem. Then, using the uniqueness condition in [10] for ℓ_1 minimization for SMV problem, we can write

$$\|\boldsymbol{\Phi}_{B_i}^{+}\boldsymbol{\Phi}_j\|_1 < 1, \forall j \notin B_i, \quad 1 \leq i \leq L. \tag{21}$$

The most restrictive condition is for the set B_i with the largest cardinality.

C. Proof of Theorem 2

Let $\{\tilde{\boldsymbol{S}}, \tilde{\boldsymbol{A}}\}$ and $\{\check{\boldsymbol{S}}, \check{\boldsymbol{A}}\}$ be the solutions to optimization problem $(\boldsymbol{P}_1)$, $\text{Supp}\left(\tilde{\boldsymbol{S}}_{(\downarrow i)}\right) = B_i$ and $\text{Supp}\left(\check{\boldsymbol{S}}_{(\downarrow i)}\right) = \check{B}$. By multiplying $\check{\boldsymbol{A}}^{-1}$ to both sides of (1), the following equation is obtained using (3)

$$\boldsymbol{Y}\check{\boldsymbol{A}}^{-1} = \boldsymbol{\Phi}\boldsymbol{X}' = \boldsymbol{\Phi}\check{\boldsymbol{S}}, \tag{22}$$

where $\check{\boldsymbol{X}} = \tilde{\boldsymbol{S}}\tilde{\boldsymbol{A}}\check{\boldsymbol{A}}^{-1}$. Using (22), one can write

$$\boldsymbol{\Phi}_B\check{\boldsymbol{X}}'_{(\downarrow i)} = \boldsymbol{\Phi}_{\check{B}}\check{\boldsymbol{S}}'_{(\downarrow i)} \tag{23}$$

where $\check{\boldsymbol{X}}'$ and $\check{\boldsymbol{S}}'$ include the nonzero rows of $\check{\boldsymbol{X}}$ and $\check{\boldsymbol{S}}$ respectively. Clearly, $\left|\text{Supp}(\boldsymbol{X}'_{(\downarrow i)})\right| \leq |B|$. Therefore, $\boldsymbol{\Phi}_B$ is considered as the columns corresponding to nonzero rows of $\check{\boldsymbol{X}}'$. By multiplying $\boldsymbol{\Phi}_B^{+}$ to (23), we obtain,

$$\check{\boldsymbol{X}}'_{(\downarrow i)} = \boldsymbol{\Phi}_B^{+}\boldsymbol{\Phi}_{\check{B}}\check{\boldsymbol{S}}'_{(\downarrow i)}. \tag{24}$$

Using (24), one can write

$$\begin{aligned}\sum_{i=1}^{L}\left\|\tilde{\boldsymbol{S}}_{(\downarrow i)}\right\|_1 \leq& \sum_{i=1}^{L}\left\|\check{\boldsymbol{X}}'_{(\downarrow i)}\right\|_1 = \sum_{i=1}^{L}\left\|\boldsymbol{\Phi}_B^{+}\boldsymbol{\Phi}_{\check{B}}\check{\boldsymbol{S}}'_{(\downarrow i)}\right\|_1 \\ \leq& \sum_{i=1}^{L}\sum_{k=1}^{|\check{B}|}\left\|\boldsymbol{\Phi}_B^{+}\boldsymbol{\Phi}_{\check{B}(\downarrow k)}\right\|_1\left|\check{\boldsymbol{S}}'_{(k,i)}\right| \\ <& \sum_{i=1}^{L}\left\|\check{\boldsymbol{S}}_{(\downarrow i)}\right\|_1. \end{aligned} \tag{25}$$

The latter inequality holds in (25) if the condition in (8) is satisfied. The first inequality holds because $\tilde{\boldsymbol{S}}$ is the minimum of $(\boldsymbol{P}_1)$. This completes the theorem.

REFERENCES

[1] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Process. Magazine*, vol. 25, no. 2, pp. 14–20, 2008.

[2] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal Process. magazine*, vol. 25, no. 2, pp. 83–91, 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, 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, 2015.

[5] S. Aviyente, "Compressed sensing framework for EEG compression," in *Statistical Signal Process., 2007. SSP'07. IEEE/SP 14th Workshop on*. IEEE, 2007, pp. 181–184.

[6] S. Arberet, P. Vandergheynst, R. E. Carrillo, J.-P. Thiran, and Y. Wiaux, "Sparse reverberant audio source separation via reweighted analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 21, no. 7, pp. 1391–1402, 2013.

[7] T. Strohmer and B. Friedlander, "Compressed sensing for MIMO radar-algorithms and performance," in *Signals, Syst. and Computers, 2009 Conf. Record of the Forty-Third Asilomar Conf. on*. IEEE, 2009, pp. 464–468.

[8] 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, 2007.

[9] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. of Sel. topics in signal Process.*, vol. 4, no. 2, pp. 310–316, 2010.

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

[11] J. A. Tropp, A. C. Gilbert and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Acoust., Speech, and Signal Process., 2005. Proc.(ICASSP'05). IEEE Int. Conf. on*, vol. 5. IEEE, 2005, pp. v–721.

[12] 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, 2012.

[13] 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, 2005.

[14] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE J. of Sel. Topics in Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[15] Z. Zhang and B. D. Rao, "Recovery of block sparse signals using the framework of block sparse bayesian learning," in *Acoust., Speech and Signal Process. (ICASSP), 2012 IEEE Int. Conf. on*. IEEE, 2012, pp. 3345–3348.

[16] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[17] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. of the Royal Statistical Soc. Series B (Methodological)*, pp. 267–288, 1996.

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

[19] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2004, vol. 46.

[20] I. Rish and G. Grabarnik, *Sparse modeling: theory, algorithms, and applications*. CRC press, 2014.