Fusion of sparse reconstruction algorithms for multiple measurement vectors

K G DEEPA¹, SOORAJ K AMBAT^{1,2,*} and K V S HARI¹

¹Statistical Signal Processing Laboratory, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India

²Naval Physical and Oceanographic Laboratory, Defence Research and Development Organisation, Kochi 682021, India

e-mail: deepa@ece.iisc.ernet.in; sooraj@ece.iisc.ernet.in; hari@ece.iisc.ernet.in

MS received 21 December 2015; accepted 2 May 2016

Abstract. We consider the recovery of sparse signals that share a common support from multiple measurement vectors. The performance of several algorithms developed for this task depends on parameters like dimension of the sparse signal, dimension of measurement vector, sparsity level, measurement noise. We propose a fusion framework, where several multiple measurement vector reconstruction algorithms participate and the final signal estimate is obtained by combining the signal estimates of the participating algorithms. We present the conditions for achieving a better reconstruction performance than the participating algorithms. Numerical simulations demonstrate that the proposed fusion algorithm often performs better than the participating algorithms.

Keywords. Compressed sensing; fusion; sparse signal reconstruction; multiple measurement vectors.

1. Introduction

Consider the standard compressed sensing (CS) measurement setup, where a K -sparse signal $\mathbf{x} \in \mathbb{R}^{N \times 1}$ is acquired through M linear measurements via

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{w}. \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ denotes the measurement matrix, $\mathbf{b} \in \mathbb{R}^{M \times 1}$ represents the measurement vector, and $\mathbf{w} \in \mathbb{R}^{M \times 1}$ denotes the additive measurement noise present in the system. The reconstruction problem, estimating $\mathbf{x}$ from Eq. (1) using $\mathbf{A}$ and $\mathbf{b}$, is known as single measurement vector (SMV) problem. In this work, we consider the multiple measurement vector (MMV) problem [1], where we have L measurements: $\mathbf{b}^{(1)} = \mathbf{A}\mathbf{x}^{(1)} + \mathbf{w}^{(1)}$, $\mathbf{b}^{(2)} = \mathbf{A}\mathbf{x}^{(2)} + \mathbf{w}^{(2)}$, $\dots$, $\mathbf{b}^{(L)} = \mathbf{A}\mathbf{x}^{(L)} + \mathbf{w}^{(L)}$. The vectors $\{\mathbf{x}^{(l)}\}_{l=1}^L$ are assumed to have a common sparse support-set. The problem is to estimate $\mathbf{x}^{(l)}$ ($l = 1, 2, \dots, L$). Instead of recovering the L signals individually, the attempt in the MMV problem is to simultaneously recover all the L signals. MMV problem arises in many applications such as the neuromagnetic inverse problem in magnetoencephalography (a modality for imaging the brain) [2, 3], array processing [4], nonparametric spectrum analysis of time series [5], and equalization of sparse communication channels [6].

Recently, many algorithms have been proposed to recover signal vectors with a common sparse support. Some among them are algorithms based on diversity minimization methods like $\ell_{2,1}$ minimization [7], and M-FOCUSS [1], greedy methods like M-OMP and M-ORMP [1], and Bayesian methods like MSBL [8] and T-MSBL [9].

However, it has been observed that the performance of many algorithms depends on many parameters like the dimension of the measurement vector, the sparsity level, the statistical distribution of the nonzero elements of the signal, the measurement noise power, etc. [9]. Thus, it becomes difficult to choose the best sparse reconstruction algorithm without *a priori* knowledge about these parameters.

Suppose we have the sparse signal estimates given by various algorithms. It may be possible to merge these estimates to form a more accurate estimate of the original. This idea of fusion of multiple estimators has been proposed in the context of signal denoising in [10], where fusion was performed by plain averaging. Recently, Ambat *et al* [11–15] proposed fusion of the estimates of sparse reconstruction algorithms to improve the sparse signal reconstruction performance of SMV problem.

In this paper, we propose a framework which uses several MMV reconstruction algorithms and combines their sparse signal support estimates to determine the final signal estimate. We refer to this scheme as MMV-fusion of algorithms for compressed sensing (MMV-FACS). We present an upper bound on the reconstruction error by

*For correspondence

MMV-FACS. We also present a sufficient condition for achieving a better reconstruction performance than any participating algorithm. By Monte-Carlo simulations, we show that fusion of viable algorithms leads to improved reconstruction performance for the MMV problem.

1.1 Notations

Matrices and vectors are denoted by bold upper case and bold lower case letters respectively. Sets are represented by upper case Greek alphabets and calligraphic letters. $\mathbf{A}_{\mathcal{T}}$ denotes the column submatrix of $\mathbf{A}$, where the indices of the columns are the elements of the set $\mathcal{T}$. $\mathbf{X}_{\mathcal{T},:}$ denotes the submatrix formed by those rows of $\mathbf{X}$, whose indices are listed in the set $\mathcal{T}$. $\mathbf{X}^K$ is the matrix obtained from $\mathbf{X}$ by keeping its K rows with the largest ℓ_2 -norm and by setting all other rows to zero, breaking ties lexicographically. $\text{supp}(\mathbf{X})$ denotes the set of indices of nonzero rows of $\mathbf{X}$. For a matrix $\mathbf{X}$, $\mathbf{x}^{(l)}$ denotes the l th column vector of $\mathbf{X}$. $\hat{\mathbf{X}}_i$ denotes the reconstructed matrix by the i th participating algorithm. The complement of the set $\mathcal{T}$ with respect to the set $\{1, 2, \dots, N\}$ is denoted by $\mathcal{T}^c$. For two sets $\mathcal{T}_1$ and $\mathcal{T}_2$, $\mathcal{T}_1 \setminus \mathcal{T}_2 = \mathcal{T}_1 \cap \mathcal{T}_2^c$ denotes the set difference. $|\mathcal{T}|$ denotes the cardinality of set $\mathcal{T}$. $\mathbf{A}^\dagger$ and $\mathbf{A}^T$ denote the pseudo-inverse and transpose of matrix $\mathbf{A}$, respectively. The (p, q) mixed norm of the matrix $\mathbf{X}$ is defined as

$$\|\mathbf{X}\|_{(p,q)} = \left(\sum_i \|\mathbf{X}_{i,:}\|_p^q \right)^{1/q}.$$

The Frobenius norm of matrix $\mathbf{A}$ is denoted as $\|\mathbf{A}\|_F$.

2. Problem formulation

The MMV problem involves solving the following L under-determined systems of linear equations

$$\mathbf{b}^{(l)} = \mathbf{A}\mathbf{x}^{(l)} + \mathbf{w}^{(l)}, \quad l = 1, 2, 3, \dots, L, \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ ($M \ll N$) represents the measurement matrix, $\mathbf{b}^{(l)} \in \mathbb{R}^{M \times 1}$ represents the l th measurement vector, and $\mathbf{x}^{(l)} \in \mathbb{R}^{N \times 1}$ denotes the corresponding K -sparse source vector. That is, $|\text{supp}(\mathbf{x}^{(l)})| \leq K$ and $\mathbf{x}^{(l)}$ share a common support-set for $l = 1, 2, \dots, L$. $\mathbf{w}^{(l)} \in \mathbb{R}^{M \times 1}$ represents the additive measurement noise. We can rewrite (2) as

$$\mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}, \quad (3)$$

where $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(L)}]$, $\mathbf{W} = [\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots, \mathbf{w}^{(L)}]$, and $\mathbf{B} = [\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(L)}]$. For a matrix $\mathbf{X}$, we define

$$\text{supp}(\mathbf{X}) = \bigcup_{i=1}^L \text{supp}(\mathbf{x}^{(i)}).$$

In Eq. (3), we assume that $\mathbf{X}$ is jointly K -sparse. That is, $|\text{supp}(\mathbf{X})| \leq K$. There are at most K rows in $\mathbf{X}$ that contain nonzero elements. We assume that $K < M$ and K is known a priori.

3. Fusion of algorithms for multiple measurement vector problem

In this paper, we propose to employ multiple sparse reconstruction algorithms independently to estimate $\mathbf{X}$ from Eq. (3) and fuse the resultant estimates to yield a better sparse signal estimate. Let $P \geq 2$ denote the number of different participating algorithms employed to estimate the sparse signal. Let $\hat{\mathcal{T}}_i$ denote the support-set estimated by the i th participating algorithm and let $\mathcal{T}$ denote the true support-set. Denote the union of the estimated support-sets as Γ , i.e. $\Gamma \triangleq \bigcup_{i=1}^P \hat{\mathcal{T}}_i$, assume that $R \triangleq |\Gamma| \leq M$. We hope that different participating algorithms work on different principles and the support-set estimated by each participating algorithm includes a partially correct information about the true support-set $\mathcal{T}$. It may be also observed that the union of the estimated support-sets Γ , is richer in terms of the true atoms as compared to the support-set estimated by any participating algorithm. Also note that once the support-set is estimated, the nonzero magnitudes of $\mathbf{X}$ can be estimated by solving a least-squares (LS) problem on an overdetermined system of linear equations. Hence, if we can identify all the true atoms included in the joint support-set Γ , we can achieve a better sparse signal estimate.

Since, we are estimating the support atoms only from Γ , we need to only solve the following problem which is lower dimensional as compared to the original problem (3):

$$\mathbf{B} = \mathbf{A}_{\Gamma}\mathbf{X}_{\Gamma,:} + \tilde{\mathbf{W}}, \quad (4)$$

where $\mathbf{A}_{\Gamma}$ denotes the submatrix formed by the columns of $\mathbf{A}$ whose indices are listed in Γ , $\mathbf{X}_{\Gamma,:}$ denotes the submatrix formed by the rows of $\mathbf{X}$ whose indices are listed in Γ , and $\tilde{\mathbf{W}} = \mathbf{W} + \mathbf{A}_{\Gamma^c}\mathbf{X}_{\Gamma^c,:}$. The matrix Eq. (4) represents a system of L linear equations which are overdetermined in nature. We use the method of LS to find an approximate solution to the overdetermined system of equations in Eq. (4). Let $\mathbf{V}_{\Gamma,:}$ denote the LS solution of Eq. (4). We choose the support-set estimate of MMV-FACS as the support of $\mathbf{V}^K$, i.e. indices of those rows having the largest ℓ_2 -norm. Once the nonzero rows are identified, solving the resultant overdetermined system of equations using LS we can estimate the

Algorithm 1. MMV-FACS**Inputs:** $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{B} \in \mathbb{R}^{M \times L}$, K , and $\{\hat{\mathcal{T}}_i\}_{i=1:P}$.**Assumption:** $|\bigcup_{i=1}^P \hat{\mathcal{T}}_i| \leq M$.**Initialization:** $\mathbf{V} = \mathbf{0} \in \mathbb{R}^{N \times 1}$.**Fusion:**

- 1: $\Gamma = \bigcup_{i=1}^P \hat{\mathcal{T}}_i$.
- 2: $\mathbf{V}_{\Gamma,:} = \mathbf{A}_{\Gamma,:}^\dagger \mathbf{B}$, $\mathbf{V}_{\Gamma^c,:} = \mathbf{0}$.
- 3: $\hat{\mathcal{T}} = \text{supp}(\mathbf{V}^K)$.

Outputs: $\hat{\mathcal{T}}$ and $\hat{\mathbf{X}}$ (where $\hat{\mathbf{X}}_{\hat{\mathcal{T}},:} = \mathbf{A}_{\hat{\mathcal{T}},:}^\dagger \mathbf{B}$ and $\hat{\mathbf{X}}_{\hat{\mathcal{T}}^c,:} = \mathbf{0}$).

nonzero entries of $\hat{\mathbf{X}}$. MMV-FACS is summarized in algorithm 1.

4. Theoretical studies of MMV-FACS

In this section, we will theoretically analyse the performance of MMV-FACS. We consider the general case for an arbitrary signal matrix. We also study the average case performance of MMV-FACS subsequently.

The performance analysis is characterized by signal-to-reconstruction-error-ratio (SRER) extended for MMV which is defined as

$$\text{SRER} \triangleq \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2}, \quad (5)$$

where $\mathbf{X}$ and $\hat{\mathbf{X}}$ denote the actual and reconstructed signal matrices respectively.

Lemma 1 Suppose that $\mathbf{A}$ satisfies the relation, for some constant $\delta_{R+K} \in (0, 1)$,

$$\|\mathbf{A}\mathbf{X}\|_F \leq \sqrt{1 + \delta_{R+K}} \|\mathbf{X}\|_F,$$

where $\|\mathbf{X}\|_0 \leq R + K$ and $\delta_{R+K} \in (0, 1)$. Here $\|\mathbf{X}\|_0$ denotes the number of nonzero rows of the matrix $\mathbf{X}$. Then, for every matrix $\mathbf{X}$,

$$\|\mathbf{A}\mathbf{X}\|_F \leq \sqrt{1 + \delta_{R+K}} \left[\|\mathbf{X}\|_F + \frac{1}{\sqrt{R+K}} \|\mathbf{X}\|_{2,1} \right].$$

Proof Proof is given in Appendix. $\square$

Lemma 2 Consider $\mathbf{A} \in \mathbb{R}^{M \times N}$ and let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two subsets of $\{1, 2, \dots, N\}$, such that $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$. Assume that $\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|} \leq 1$ and let $\mathbf{Y}$ be any matrix, such that $\text{span}(\mathbf{Y}) \in \text{span}(\mathbf{A}_{\mathcal{T}_1})$ and $\mathbf{R} = \mathbf{Y} - \mathbf{A}_{\mathcal{T}_2} \mathbf{A}_{\mathcal{T}_2}^\dagger \mathbf{Y}$. Then we have

$$\left(1 - \frac{\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}{1 - \delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}} \right) \|\mathbf{Y}\|_2 \leq \|\mathbf{R}\|_2 \leq \|\mathbf{Y}\|_2.$$

Proof Proof is given in Appendix. $\square$

4.1 Performance analysis for arbitrary signals under measurement perturbations

We analyse the performance of MMV-FACS for arbitrary signals and give an upper bound on the reconstruction error in theorem 1. We also derive a sufficient condition to get an improved performance of MMV-FACS scheme over any given participating algorithm.

Theorem 1 Let $\mathbf{X}$ be an arbitrary signal with $\mathcal{T} = \text{supp}(\mathbf{X}^K)$. Consider the MMV-FACS setup discussed in section 3, and assume that the measurement matrix $\mathbf{A}$ satisfies RIP with RIC δ_{R+K} . We have the following results:

(i) Upper bound on reconstruction error: We have

$$\begin{aligned} \|\mathbf{X} - \hat{\mathbf{X}}\|_F &\leq C_1 \|\mathbf{X} - \mathbf{X}^K\|_F + C_2 \|\mathbf{X} - \mathbf{X}^K\|_{2,1} + C_3 \\ &\quad \|\mathbf{X}_{\Gamma^c,:}\|_F + v \|\mathbf{W}\|_F, \end{aligned}$$

where

$$\begin{aligned} C_1 &= (1 + v\sqrt{1 + \delta_{R+K}}), \quad C_2 = \frac{v\sqrt{1 + \delta_{R+K}}}{\sqrt{R+K}}, \\ C_3 &= \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2}, \quad \text{and } v = \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2}. \end{aligned}$$

(ii) SRER gain:

For $\|\mathbf{X}_{\hat{\mathcal{T}}^c,:}\|_F \neq 0$ and $\|\mathbf{X}_{\Gamma^c,:}\|_F \neq 0$, MMV-FACS provides at least SRER gain of

$$\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i} \right)^2$$

over the i th participating algorithm if

$$\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)},$$

where $\eta_i = \frac{\|\mathbf{X}_{\Gamma^c,:}\|_F}{\|\mathbf{X}_{\hat{\mathcal{T}}^c,:}\|_F}$, $\zeta = \frac{\|\mathbf{W}\|_F}{\|\mathbf{X}_{\Gamma^c,:}\|_F}$, and

$$\begin{aligned} \xi &= (3\sqrt{1 + \delta_{R+K}} + 1) \frac{\|\mathbf{X} - \mathbf{X}^K\|_F}{3\|\mathbf{X}_{\Gamma^c,:}\|_F} \\ &\quad + \frac{\sqrt{1 + \delta_{R+K}}}{\sqrt{R+K}} \frac{\|\mathbf{X} - \mathbf{X}^K\|_{2,1}}{\|\mathbf{X}_{\Gamma^c,:}\|_F}. \end{aligned}$$

Proof (i) We have

$$\|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq \|\mathbf{X} - \mathbf{X}^K\|_F + \|\mathbf{X}^K - \hat{\mathbf{X}}\|_F. \quad (6)$$

Consider

$$\begin{aligned} \|\mathbf{X}^K - \hat{\mathbf{X}}\|_F &\leq \|(\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \hat{\mathbf{X}}_{\hat{\mathcal{T}},:}\|_F + \|(\mathbf{X}^K)_{\hat{\mathcal{T}}^c,:} - \hat{\mathbf{X}}_{\hat{\mathcal{T}}^c,:}\|_F \\ &\leq \|(\mathbf{X}^K)_{\hat{\mathcal{T}},:} - \hat{\mathbf{X}}_{\hat{\mathcal{T}},:}\|_F + \|(\mathbf{X}^K)_{\hat{\mathcal{T}}^c,:}\|_F \quad (\because \hat{\mathbf{X}}_{\hat{\mathcal{T}}^c,:} = \mathbf{0}). \end{aligned} \quad (7)$$

Using the relations $\hat{\mathbf{X}}_{\hat{\mathcal{T}},:} = \mathbf{A}_{\hat{\mathcal{T}},:}^\dagger \mathbf{B}$ (from algorithm 1) and $\mathbf{A}_{\hat{\mathcal{T}},:}^\dagger \mathbf{A}_{\hat{\mathcal{T}},:} = \mathbf{I}$, we get

$$\begin{aligned}
& \|(\mathbf{X}^K)_{\hat{T},:} - \hat{\mathbf{X}}_{\hat{T},:}\|_F \\
&= \|(\mathbf{X}^K)_{\hat{T},:} - \mathbf{A}_{\hat{T}}^\dagger \mathbf{B}\|_F \\
&= \|(\mathbf{X}^K)_{\hat{T},:} - \mathbf{A}_{\hat{T}}^\dagger (\mathbf{A}\mathbf{X} + \mathbf{W})\|_F \quad (\cdot: \mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}) \\
&= \|(\mathbf{X}^K)_{\hat{T},:} - \mathbf{A}_{\hat{T}}^\dagger (\mathbf{A}\mathbf{X}^K + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W})\|_F \\
&= \|(\mathbf{X}^K)_{\hat{T},:} - \mathbf{A}_{\hat{T}}^\dagger [\mathbf{A}_{\hat{T}}(\mathbf{X}^K)_{\hat{T},:} + \mathbf{A}_{\hat{T}^c}(\mathbf{X}^K)_{\hat{T}^c,:} + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W}]\|_F \\
&= \|\mathbf{A}_{\hat{T}}^\dagger \mathbf{A}_{\hat{T}^c}(\mathbf{X}^K)_{\hat{T}^c,:} + \mathbf{A}_{\hat{T}}^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{A}_{\hat{T}}^\dagger \mathbf{W}\|_F \\
&\leq \|(\mathbf{A}_{\hat{T}}^H \mathbf{A}_{\hat{T}})^{-1} \mathbf{A}_{\hat{T}}^H \mathbf{A}_{\hat{T}^c}(\mathbf{X}^K)_{\hat{T}^c,:}\|_F + \|\mathbf{A}_{\hat{T}}^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{A}_{\hat{T}}^\dagger \mathbf{W}\|_F.
\end{aligned} \tag{8}$$

Let $\mathbf{x}^{(i)}$ denote the i th column of matrix $\mathbf{X}$ and $\mathbf{w}^{(i)}$ denote the i th column of matrix $\mathbf{W}$, $i = 1, 2, \dots, L$. Now from proposition 3.1 and corollary 3.3 of [16], we obtain the following relations.

$$\|\mathbf{A}_{\hat{T}}^\dagger \mathbf{w}^{(i)}\|_2 \leq \frac{\|\mathbf{w}^{(i)}\|_2}{\sqrt{1 - \delta_{R+K}}}. \tag{9}$$

$$\|\mathbf{A}_{\hat{T}}^\dagger \mathbf{A}(\mathbf{x}^{(i)} - (\mathbf{x}^{(i)})^K)\|_2 \leq \frac{\|\mathbf{A}(\mathbf{x}^{(i)} - (\mathbf{x}^{(i)})^K)\|_2}{\sqrt{1 - \delta_{R+K}}}. \tag{10}$$

$$\|(\mathbf{A}_{\hat{T}}^H \mathbf{A}_{\hat{T}})^{-1} \mathbf{A}_{\hat{T}}^H \mathbf{A}_{\hat{T}^c}((\mathbf{x}^{(i)})^K)_{\hat{T}^c,:}\|_2 \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|((\mathbf{x}^{(i)})^K)_{\hat{T}^c,:}\|_2. \tag{11}$$

□

Consider Eq. (9), we get

$$\|\mathbf{A}_{\hat{T}}^\dagger \mathbf{w}^{(i)}\|_2^2 \leq \frac{\|\mathbf{w}^{(i)}\|_2^2}{1 - \delta_{R+K}} \quad \forall i = 1, 2, \dots, L.$$

Summing the above equation over $i = 1, 2, \dots, L$, we obtain

$$\begin{aligned}
\sum_{i=1}^L \|\mathbf{A}_{\hat{T}}^\dagger \mathbf{w}^{(i)}\|_2^2 &\leq \frac{1}{1 - \delta_{R+K}} \sum_{i=1}^L \|\mathbf{w}^{(i)}\|_2^2 \\
\|\mathbf{A}_{\hat{T}}^\dagger \mathbf{W}\|_F^2 &\leq \frac{1}{1 - \delta_{R+K}} \|\mathbf{W}\|_F^2 \\
\|\mathbf{A}_{\hat{T}}^\dagger \mathbf{W}\|_F &\leq \frac{1}{\sqrt{1 - \delta_{R+K}}} \|\mathbf{W}\|_F.
\end{aligned} \tag{12}$$

Similarly, summing the relations in Eqs. (10) and (11), we obtain

$$\|\mathbf{A}_{\hat{T}}^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F \leq \frac{\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F}{\sqrt{1 - \delta_{R+K}}}. \tag{13}$$

$$\|(\mathbf{A}_{\hat{T}}^H \mathbf{A}_{\hat{T}})^{-1} \mathbf{A}_{\hat{T}}^H \mathbf{A}_{\hat{T}^c}(\mathbf{X}^K)_{\hat{T}^c,:}\|_F \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F. \tag{14}$$

Substituting Eqs. (12), (13) and (14) in Eq. (8), we get

$$\begin{aligned}
\|(\mathbf{X}^K)_{\hat{T},:} - \hat{\mathbf{X}}_{\hat{T},:}\|_F &\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F \\
&\quad + \frac{\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F}{\sqrt{1 - \delta_{R+K}}} + \frac{\|\mathbf{W}\|_F}{\sqrt{1 - \delta_{R+K}}} \\
&\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F + \frac{1}{1 - \delta_{R+K}} \\
&\quad (\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{W}\|_F).
\end{aligned} \tag{15}$$

Substituting Eq. (15) in Eq. (7), we get

$$\|\mathbf{X}^K - \hat{\mathbf{X}}\|_F \leq \frac{1}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F + \frac{1}{1 - \delta_{R+K}} (\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{W}\|_F). \tag{16}$$

Next, we will find an upper bound for $\|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F$.

Define $\hat{T}_\Delta \triangleq \Gamma \setminus \hat{T}$. That is, $\hat{T}_\Delta$ is the set formed by the atoms in Γ which are discarded by algorithm 1. Since $\hat{T} \subset \Gamma$, we have $\hat{T}^c = \Gamma^c \cup \hat{T}_\Delta$ and hence, we obtain

$$\|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F \leq \|(\mathbf{X}^K)_{\Gamma^c,:}\|_F + \|(\mathbf{X}^K)_{\hat{T}_\Delta,:}\|_F. \tag{17}$$

□ We also have

$$\begin{aligned}
\|(\mathbf{X}^K)_{\hat{T}_\Delta,:}\|_F &\leq \|(\mathbf{V}_{\Gamma,:})_{\hat{T}_\Delta,:}\|_F + \|(\mathbf{V}_{\Gamma,:} - (\mathbf{X}^K)_{\Gamma,:})_{\hat{T}_\Delta,:}\|_F \\
&\leq \|(\mathbf{V}_{\Gamma,:})_{\hat{T}_\Delta,:}\|_F + \|\mathbf{V}_{\Gamma,:} - (\mathbf{X}^K)_{\Gamma,:}\|_F.
\end{aligned} \tag{18}$$

Note that $(\mathbf{V}_{\Gamma,:})_{\hat{T}_\Delta,:}$ contains the K -rows of $\mathbf{V}_{\Gamma,:}$ with highest row ℓ_2 -norm. Therefore, using $|\hat{T}| = |\mathcal{T}| = K$, we get

$$\begin{aligned}
\|(\mathbf{V}_{\Gamma,:})_{\hat{T}_\Delta,:}\|_F &\leq \|(\mathbf{V}_{\Gamma,:})_{\Gamma \setminus \mathcal{T},:}\|_F \\
&= \|\mathbf{V}_{\Gamma \setminus \mathcal{T},:} - (\mathbf{X}^K)_{\Gamma \setminus \mathcal{T},:}\|_F \quad (\cdot: (\mathbf{X}^K)_{\Gamma \setminus \mathcal{T},:} = 0) \\
&\leq \|(\mathbf{V}_{\Gamma,:} - (\mathbf{X}^K)_{\Gamma,:})_{\Gamma \setminus \mathcal{T},:}\|_F.
\end{aligned} \tag{19}$$

Substituting Eq. (19) in Eq. (18), we get

$$\|(\mathbf{X}^K)_{\hat{T}_\Delta,:}\|_F \leq 2 \|(\mathbf{V}_{\Gamma,:} - (\mathbf{X}^K)_{\Gamma,:})_{\Gamma \setminus \mathcal{T},:}\|_F. \tag{20}$$

Now, consider

$$\begin{aligned}
& \|\mathbf{V}_{\Gamma,:} - (\mathbf{X}^K)_{\Gamma,:}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger \mathbf{B} - (\mathbf{X}^K)_{\Gamma,:}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger (\mathbf{A}\mathbf{X} + \mathbf{W}) - (\mathbf{X}^K)_{\Gamma,:}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger (\mathbf{A}\mathbf{X}^K + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W}) - (\mathbf{X}^K)_{\Gamma,:}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger (\mathbf{A}_\Gamma (\mathbf{X}^K)_{\Gamma,:} + \mathbf{A}_{\Gamma^c} (\mathbf{X}^K)_{\Gamma^c,:} + \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{W}) - (\mathbf{X}^K)_{\Gamma,:}\|_F \\
&= \|\mathbf{A}_\Gamma^\dagger \mathbf{A}_{\Gamma^c} (\mathbf{X}^K)_{\Gamma^c,:} + \mathbf{A}_\Gamma^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K) + \mathbf{A}_\Gamma^\dagger \mathbf{W}\|_F \quad (\because \mathbf{A}_\Gamma^\dagger \mathbf{A}_\Gamma = \mathbf{I}) \\
&\leq \|\mathbf{A}_\Gamma^\dagger \mathbf{A}_{\Gamma^c} (\mathbf{X}^K)_{\Gamma^c,:}\|_F + \|\mathbf{A}_\Gamma^\dagger \mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{A}_\Gamma^\dagger \mathbf{W}\|_F.
\end{aligned} \tag{21}$$

Using Eqs. (12), (13) and (14) in Eq. (21), we get

$$\begin{aligned}
\|\mathbf{V}_{\Gamma,:} - (\mathbf{X}^K)_{\Gamma,:}\|_F &\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\Gamma^c,:}\|_F \\
&\quad + \frac{\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F}{\sqrt{1 - \delta_{R+K}}} + \frac{\|\mathbf{W}\|_F}{\sqrt{1 - \delta_{R+K}}} \\
&\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\Gamma^c,:}\|_F \\
&\quad + \frac{\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F}{1 - \delta_{R+K}} + \frac{\|\mathbf{W}\|_F}{1 - \delta_{R+K}}. \\
&\quad (\because 0 < 1 - \delta_{R+K} < 1).
\end{aligned} \tag{22}$$

Using Eqs. (20) and (22) in Eq. (17), we get

$$\begin{aligned}
\|(\mathbf{X}^K)_{\hat{T}^c,:}\|_F &\leq \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \|(\mathbf{X}^K)_{\Gamma^c,:}\|_F + \frac{2}{1 - \delta_{R+K}} \\
&\quad \left(\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{W}\|_F \right).
\end{aligned} \tag{23}$$

Let $\mathbf{x}_1^{(i)}$ denote the i th column of matrix $\mathbf{X}^K$. Then, we have

$$\begin{aligned}
\|\mathbf{A}\mathbf{x}_1^{(i)}\|_2^2 &\leq (1 + \delta_{R+K}) \|\mathbf{x}_1^{(i)}\|_2^2 \quad (\because \mathbf{A} \text{ satisfies RIP}) \\
\sum_{i=1}^L \|\mathbf{A}\mathbf{x}_1^{(i)}\|_2^2 &\leq \sum_{i=1}^L (1 + \delta_{R+K}) \|\mathbf{x}_1^{(i)}\|_2^2 \\
\|\mathbf{A}\mathbf{X}^K\|_F^2 &\leq (1 + \delta_{R+K}) \|\mathbf{X}^K\|_F^2.
\end{aligned} \tag{24}$$

Using lemma 1 and Eq. (24), we get

$$\begin{aligned}
\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F &\leq \sqrt{1 + \delta_{R+K}} \\
&\quad \left[\|\mathbf{X} - \mathbf{X}^K\|_F + \frac{1}{\sqrt{R+K}} \|(\mathbf{X} - \mathbf{X}^K)\|_{2,1} \right].
\end{aligned} \tag{25}$$

Substituting Eq. (23) in Eq. (16), we get

$$\begin{aligned}
\|\mathbf{X}^K - \hat{\mathbf{X}}\|_F &\leq \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|(\mathbf{X}^K)_{\Gamma^c,:}\|_F + \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2} \\
&\quad \left[\|\mathbf{A}(\mathbf{X} - \mathbf{X}^K)\|_F + \|\mathbf{W}\|_F \right] \\
&\leq v\sqrt{1 + \delta_{R+K}} \|\mathbf{X} - \mathbf{X}^K\|_F + \frac{v\sqrt{1 + \delta_{R+K}} \|\mathbf{X} - \mathbf{X}^K\|_{2,1}}{\sqrt{R+K}} \\
&\quad + \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|\mathbf{X}_{\Gamma^c,:}\|_F + v\|\mathbf{W}\|_F, \\
&\quad (\text{using Eq. (25)}),
\end{aligned} \tag{26}$$

where

$$v = \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2}.$$

Substituting Eq. (26) in Eq. (6) and using the definitions of C_1 , C_2 , and C_3 , we get

$$\begin{aligned}
\|\mathbf{X} - \hat{\mathbf{X}}\|_F &\leq C_1 \|\mathbf{X} - \mathbf{X}^K\|_F + C_2 \|\mathbf{X} - \mathbf{X}^K\|_{2,1} + C_3 \|\mathbf{X}_{\Gamma^c,:}\|_F \\
&\quad + v\|\mathbf{W}\|_F.
\end{aligned} \tag{27}$$

(ii) Using Eq. (27) and the definitions of ζ and η_i , we get

$$\begin{aligned}
\|\mathbf{X} - \hat{\mathbf{X}}\|_F &\leq \frac{1 + \delta_{R+K} + 3\zeta + 3\zeta}{(1 - \delta_{R+K})^2} \|\mathbf{X}_{\Gamma^c,:}\|_F \\
&= \frac{1 + \delta_{R+K} + 3\zeta + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \|(\mathbf{X} - \hat{\mathbf{X}}_i)_{\hat{T}_i^c,:}\|_F \\
&\quad (\because (\hat{\mathbf{X}}_i)_{\hat{T}_i^c,:} = \mathbf{0}) \\
&\leq \frac{1 + \delta_{R+K} + 3\zeta + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \|\mathbf{X} - \hat{\mathbf{X}}_i\|_F.
\end{aligned}$$

Hence, we obtain the relation for SRER for MMV-FACS, in case of arbitrary signals, as

$$\begin{aligned}
\text{SRER}|_{\text{MMV-FACS}} &= \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2} \\
&\geq \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}_i\|_F^2} \times \left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\zeta)\eta_i} \right)^2.
\end{aligned}$$

Hence, MMV-FACS provides at least SRER gain of

$$\begin{aligned}
&\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\zeta)\eta_i} \right)^2 \text{ over } i\text{th algorithm if} \\
&\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\zeta)}. \\
&\text{Note that} \\
&\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\zeta)} < 1.
\end{aligned} \quad \square$$

4.2 Exactly K -sparse matrix

Theorem 1 considered the case when $\mathbf{X}$ is an arbitrary matrix. If $\mathbf{X}$ is a K -sparse matrix, then we have $\mathbf{X} = \mathbf{X}^K$ and $\zeta = 0$. Thus, it follows from theorem 1 that MMV-FACS provides at least SRER gain of

$$\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)\eta_i} \right)^2$$

over i th participating algorithm if

$$\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)}.$$

Thus, the improvement in the SRER gain provided by MMV-FACS over the i th algorithm for a K -sparse matrix is greater than that of an arbitrary matrix by a factor of

$$\left(1 + \frac{3\zeta}{(1 + \delta_{R+K} + 3\zeta)} \right)^2.$$

The second part of theorem 1 considers the case when $\|\mathbf{X}_{\hat{\mathcal{T}}_i, :}\|_F \neq 0$ and $\|\mathbf{X}_{\Gamma^c, :}\|_F \neq 0$. If $\|\mathbf{X}_{\hat{\mathcal{T}}_i, :}\|_F \neq 0$, then $\hat{\mathcal{T}}_i \not\subseteq \mathcal{T}$. Also, $\|\mathbf{X}_{\Gamma^c, :}\|_F = 0$ implies $\mathcal{T} \subseteq \Gamma$. Suppose $\|\mathbf{X}_{\hat{\mathcal{T}}_i, :}\|_F = 0$, then the support-set is correctly estimated by i th algorithm and further performance improvement is not possible by MMV-FACS. Hence, we consider the case where $\|\mathbf{X}_{\Gamma^c, :}\|_F = 0$, and derive the condition for exact reconstruction by MMV-FACS in the following proposition.

Proposition 1 Assume that $\|\mathbf{X}_{\Gamma^c, :}\|_F = 0$ and all other conditions in theorem 1 hold good. Then, in clean measurement case ($\mathbf{W} = \mathbf{0}$), MMV-FACS estimates the support-set correctly and provides exact reconstruction.

Proof We have

$$\mathbf{X}_{\Gamma^c, :} = \mathbf{0} \Rightarrow \mathcal{T} \subset \Gamma. \quad (28)$$

$$\mathbf{B} = \mathbf{A}_{\mathcal{T}} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{W}. \quad (29)$$

From algorithm 1, we have $\mathbf{V} \in \mathbb{R}^{N \times L}$, where $\mathbf{V}_{\Gamma^c, :} = \mathbf{0}$, and

$$\begin{aligned} \mathbf{V}_{\Gamma, :} &= \mathbf{A}_{\Gamma}^{\dagger} \mathbf{B} \\ &= \mathbf{A}_{\Gamma}^{\dagger} (\mathbf{A}_{\mathcal{T}} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{W}) \text{ (using Eq. (29))} \\ &= \mathbf{A}_{\Gamma}^{\dagger} (\mathbf{A}_{\Gamma} \mathbf{X}_{\Gamma, :} + \mathbf{W}) \text{ (using Eq. (28))} \\ &= \mathbf{X}_{\Gamma, :} + \mathbf{A}_{\Gamma}^{\dagger} \mathbf{W}. \end{aligned}$$

If $\mathbf{W} = \mathbf{0}$, then $\mathbf{V}_{\Gamma, :} = \mathbf{X}_{\Gamma, :}$ and $\mathbf{V} = \mathbf{X}$ ($\because \mathcal{T} \subset \Gamma$). Thus, MMV-FACS estimates the support-set correctly from $\mathbf{V}$. $\square$

In practice, the original signal is not known and hence, it is not possible to evaluate the performance w.r.t. the true signal. Hence, in applications, the decrease in energy of the residual is often treated as a measure of performance improvement. Proposition 2 gives a sufficient condition for decrease in the energy of the residual matrix obtained by MMV-FACS over the i th participating algorithm.

Proposition 2 For a K -sparse matrix $\mathbf{X}$, let $\mathbf{R} = \mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{B}$ and $\mathbf{R}_i = \mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{B}$ represent the residue matrix of MMV-FACS and i th algorithm respectively. Assume that

$$\frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \leq \left(\frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta \right)$$

is satisfied then we have,

$$\|\mathbf{R}\|_F \leq \|\mathbf{R}_i\|_F.$$

Proof We have

$$\begin{aligned} \|\mathbf{R}\|_F &= \|\mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{B}\|_F \\ &= \|\mathbf{A} \mathbf{X} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} (\mathbf{A} \mathbf{X} + \mathbf{W})\|_F \quad (\because \mathbf{B} = \mathbf{A} \mathbf{X} + \mathbf{W}) \\ &= \|\mathbf{A}_{\hat{\mathcal{T}}} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} (\mathbf{A}_{\hat{\mathcal{T}}} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :} + \mathbf{W})\|_F \\ &= \|\mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{W}\|_F \quad (\because \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{A}_{\hat{\mathcal{T}}} = \mathbf{I}) \\ &\leq \|\mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :}\|_F + \|\mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}} \mathbf{A}_{\hat{\mathcal{T}}}^{\dagger} \mathbf{W}\|_F \\ &\leq \|\mathbf{A}_{\hat{\mathcal{T}}^c} \mathbf{X}_{\hat{\mathcal{T}}^c, :}\|_F + \|\mathbf{W}\|_F \quad (\text{using lemma 2}) \\ &= \|\mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}, :}\|_F + \|\mathbf{W}\|_F \quad (\because \mathcal{T} = \text{supp}(\mathbf{X})) \\ &\leq \sqrt{1 + \delta_{R+K}} \|\mathbf{X}_{\hat{\mathcal{T}}^c, :}\|_F + \|\mathbf{W}\|_F \quad (\because |\mathcal{T} \setminus \hat{\mathcal{T}}| \leq K \text{ and } \delta_K \leq \delta_{R+K}). \end{aligned}$$

Using Eq. (23), we have

$$\begin{aligned} \|\mathbf{R}\|_F &\leq \sqrt{1 + \delta_{R+K}} \left(\frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \|\mathbf{X}_{\Gamma^c, :}\|_F + \frac{2}{1 - \delta_{R+K}} \|\mathbf{W}\|_F \right) \\ &\quad + \|\mathbf{W}\|_F \\ &\leq \frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} \left(1 + \delta_{R+K} + 2\zeta + \frac{1 - \delta_{R+K}}{\sqrt{1 + \delta_{R+K}}} \zeta \right) \|\mathbf{X}_{\Gamma^c, :}\|_F \\ &\leq \frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \|\mathbf{X}_{\Gamma^c, :}\|_F. \end{aligned} \quad (30)$$

Now, consider

$$\begin{aligned} \|\mathbf{R}_i\|_F &= \|\mathbf{B} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{B}\|_F \\ &= \|\mathbf{A} \mathbf{X} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} (\mathbf{A} \mathbf{X} + \mathbf{W})\|_F \quad (\because \mathbf{B} = \mathbf{A} \mathbf{X} + \mathbf{W}) \\ &= \|\mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} (\mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{X}_{\hat{\mathcal{T}}_i, :} + \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} + \mathbf{W})\|_F \\ &= \|\mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} + \mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{W}\|_F \\ &\geq \|\mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{A}_{\hat{\mathcal{T}}_i^c} \mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F - \|\mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{W}\|_F \\ &\quad (\text{using reverse triangle inequality}) \\ &= \|\mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i, :} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i, :}\|_F - \|\mathbf{W} - \mathbf{A}_{\hat{\mathcal{T}}_i} \mathbf{A}_{\hat{\mathcal{T}}_i}^{\dagger} \mathbf{W}\|_F \\ &\geq \left(1 - \frac{\delta_{R+K}}{1 - \delta_{R+K}} \right) \|\mathbf{A}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i} \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}_i, :}\|_F - \|\mathbf{W}\|_F \\ &\quad (\text{using lemma 2 and } \delta_{2K} \leq \delta_{R+K}) \\ &\geq \frac{1 - 2\delta_{R+K}}{1 - \delta_{R+K}} \sqrt{1 - \delta_{R+K}} \|\mathbf{X}_{\hat{\mathcal{T}}_i^c, :}\|_F - \|\mathbf{W}\|_F \quad (\because |\mathcal{T} \setminus \hat{\mathcal{T}}_i| \leq K \\ &\quad \text{and } \delta_K \leq \delta_{R+K}) \\ &= \left(\frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta \right) \|\mathbf{X}_{\Gamma^c, :}\|_F. \end{aligned} \quad (31)$$

From Eqs. (30) and (31), we get a sufficient condition for $\|\mathbf{R}\|_F \leq \|\mathbf{R}_i\|_F$ as

$$\frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \leq \left(\frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta \right). \quad (32)$$

Thus, if Eq. (32) is satisfied, MMV-FACS produces a smaller residual matrix (in the Frobenius norm sense) than that of the i th participating algorithm. $\square$

4.3 Average case analysis

Intuitively, we expect multiple measurement vector problem to perform better than the single measurement vector case. However, if each measurement vector is the same, i.e. in the worst case, we have $\mathbf{x}^{(i)} = \mathbf{c}$, $\forall i = 1, \dots, L$, then we do not have any additional information on $\mathbf{X}$ than that provided by a single measurement vector $\mathbf{x}^{(1)}$. So far, we have carried out only the worst case analysis, i.e., conditions under which the algorithm is able to recover any joint sparse matrix $\mathbf{X}$. This approach does not provide insight into the superiority of sparse signal reconstruction with multiple measurement vectors compared to the single measurement vector case.

To notice a performance gain with multiple measurement vectors, next, we proceed with an average case analysis. Here, we impose a probability model on the K sparse $\mathbf{X}$ as suggested by Remi *et al* [17]. In particular, on the support-set $\mathcal{T}$, we impose that $\mathbf{X}_{\mathcal{T},:} = \Sigma \Phi$, where Σ is a $K \times K$ diagonal matrix with positive diagonal entries and Φ is a $K \times L$ random matrix with i.i.d. Gaussian entries. Our goal is to show that under this signal model, the typical behaviour of MMV-FACS is better than in the worst case.

Theorem 2 Consider the MMV-FACS setup discussed in section 3. Assume a Gaussian signal model, i.e., $\mathbf{X}_{\mathcal{T},:} = \Sigma \Phi$, where Σ is a $K \times K$ diagonal matrix with positive diagonal entries and Φ is a $K \times L$ random matrix with i.i.d. Gaussian entries. Let $\mathbf{e}_i$ denote a $|\Gamma| \times 1$ vector with a “1” in the i th coordinate and “0” elsewhere. Let

$$\eta = \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2$$

and

$$\gamma = \frac{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \frac{\eta}{C_2(L)}}{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}.$$

where $C_2(L) = \mathbb{E} \|\mathbf{Z}\|_2$ with $\mathbf{Z} = (Z_1, \dots, Z_L)$ being a vector of independent standard normal variables. Assume that $\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 > \frac{\eta}{C_2(L)}$. Let Θ denote the event that MMV-FACS picks all correct indices from the union-set Γ . Then, we have

$$P(\Theta) \geq 1 - K \exp(-2A_2(L)\gamma^2),$$

where

$$A_2(L) = \left(\frac{\ddot{\Gamma}(\frac{L+1}{2})}{\ddot{\Gamma}(\frac{L}{2})} \right)^2 \approx \frac{L}{2},$$

$\ddot{\Gamma}(\cdot)$ denotes the Gamma function.

Proof We have

$$\begin{aligned} P(\Theta) &= P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{B} \right\|_2 > \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{B} \right\|_2\right) \\ &= P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger (\mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} + \mathbf{W}) \right\|_2 > \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger (\mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} + \mathbf{W}) \right\|_2\right) \\ &\geq P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left(\left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 \right) > \max_{j \in (\Gamma \setminus \mathcal{T})} \left(\left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 + \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 \right)\right) \\ &\quad \text{(using reverse triangle inequality and triangle inequality respectively)} \\ &= P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 > \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{W} \right\|_2\right) \\ &= P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 > \eta\right) \\ &= 1 - P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq \eta\right) \\ &\geq 1 - P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq C\right) \\ &\quad - P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \geq C - \eta\right). \end{aligned} \quad (33)$$

Now, let us derive an upper bound for $P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq C\right)$. Influenced by the concentration of measure results in [17], we set

$$C = (1 - \epsilon_1) C_2(L) \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2, \quad (34)$$

where $0 < \epsilon_1 < 1$.

Using Eq. (5.5) in [17], we get,

$$P\left(\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_{\mathcal{T},:} \right\|_2 \leq C\right) \leq |\mathcal{T}| \exp(-A_2(L)\epsilon_1^2). \quad (35)$$

To bound the second probability, consider

$$\begin{aligned} &P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \Phi \right\|_2 \geq C - \eta\right) \\ &= P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \Phi \right\|_2 \geq (C - \eta) \frac{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}\right). \end{aligned}$$

Let

$$1 + \epsilon_2 = \frac{C - \eta}{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}. \quad (36)$$

Using Eq. (5.3) in [17]

$$\begin{aligned}
P\left(\max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \Phi \right\|_2 \right. \\
\geq (1 + \epsilon_2) C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 \Big) \\
\leq |\mathcal{T}| \exp(-A_2(L) \epsilon_2^2). \quad (37)
\end{aligned}$$

For the above inequality to hold, it is required that $\epsilon_2 > 0$. By setting $\epsilon_2 = \epsilon_1$, and using Eqs. (34) and (36), we get

$$\epsilon_1 = \frac{(1 - \epsilon_1) C_2(L) \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \eta}{C_2(L) \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2} - 1.$$

Now, solving for ϵ , we get

$$\epsilon_1 = \frac{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 - \frac{\eta}{C_2(L)}}{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| \mathbf{e}_i^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| \mathbf{e}_j^T \mathbf{A}_\Gamma^\dagger \mathbf{A}_\mathcal{T} \Sigma \right\|_2}.$$

Clearly $\epsilon_1 < 1$ and by the assumption in the theorem $\epsilon_1 > 0$. Hence, we have $0 < \epsilon_1 < 1$. Also, note that $\gamma = \epsilon_1$. Substituting Eqs. (35) and (37) in Eq. (33), we get

$$P(\Theta) \geq 1 - K \exp(-2A_2(L)\gamma^2).$$

□

Since $A_2(L) \approx \frac{L}{2}$, the probability that MMV-FACS selects all correct indices from the union set increases as L increases. Thus, more than one measurement vector improves the performance.

5. Numerical experiments and results

We conducted numerical experiments using synthetic data and real signals to evaluate the performance of MMV-FACS. The performance is evaluated using Average SRER (ASRER), which is defined as

$$\text{ASRER} = \frac{\sum_{j=1}^{n_{\text{trials}}} \|\mathbf{X}_j\|_F^2}{\sum_{j=1}^{n_{\text{trials}}} \|\mathbf{X}_j - \hat{\mathbf{X}}_j\|_F^2}, \quad (38)$$

where $\mathbf{X}_j$ and $\hat{\mathbf{X}}_j$ denote the actual and reconstructed jointly sparse signal matrix in the j th trial respectively, and n_{trials} denotes the total number of trials.

5.1 Synthetic sparse signals

For noisy measurement simulations, we define the signal-to-measurement-noise-ratio (SMNR) as

$$\text{SMNR} \triangleq \frac{\mathbb{E}\{\|\mathbf{x}^{(i)}\|_2^2\}}{\mathbb{E}\{\|\mathbf{w}^{(i)}\|_2^2\}},$$

where $\mathbb{E}\{\cdot\}$ denotes the mathematical expectation operator. The simulation setup is described below.

5.1a Experimental setup: Following steps are involved in the simulation:

- (i) Generate elements of $\mathbf{A}_{M \times N}$ independently from $\mathcal{N}(0, \frac{1}{M})$ and normalize each column norm to unity.
- (ii) Choose K nonzero locations uniformly at random from the set $\{1, 2, \dots, N\}$ and fill those K rows of $\mathbf{X}$ based on the choice of signal characteristics:
  - (a) *Gaussian sparse signal matrix:* Nonzero values independently from $\mathcal{N}(0, 1)$.
  - (b) *Rademacher sparse signal matrix:* Nonzero values are set to $+1$ or -1 with probability $\frac{1}{2}$.

Remaining $N - K$ rows of $\mathbf{X}$ are set to zero.

- (iii) The MMV measurement matrix $\mathbf{B}$ is computed as $\mathbf{B} = \mathbf{A}\mathbf{X} + \mathbf{W}$, where the columns of $\mathbf{W}$, $\mathbf{w}^{(i)}$'s are independent and their elements are i.i.d. and Gaussian with variance determined from the specified SMNR.
- (iv) Apply the MMV sparse recovery method.
- (v) Repeat steps (i)–(iv), S times.
- (vi) Find ASRER using Eq. (38).

5.1b Results and discussions: We used M-OMP, M-SP, M-BPDN [18], and M-FOCUSS [1] as the participating algorithms in MMV-FACS. The software code for M-BPDN was taken from SPGL1 software package [19]. Since M-FOCUSS and M-BPDN algorithms may not yield an exact K -sparse solution, we estimate the support-set as the indices of the K rows with largest ℓ_2 norm. We fixed the sparse signal dimension $N = 500$ and sparsity level $K = 20$ in the simulation. The results were calculated by averaging over 1000 trials ($S = 1000$). We use an oracle estimator for performance benchmarking. The oracle estimator is aware of the true support-set and finds the nonzero entries of the sparse matrix by solving LS.

The empirical performance of MMV reconstruction algorithms for different values of M is shown in figure 1. The simulation parameters are $L = 20$, $\text{SMNR} = 20$ dB and $\mathbf{X}$ is chosen as Gaussian sparse signal matrix. For $M = 35$, MMV-FACS (M-BPDN, M-FOCUSS) gave 10.67 and 4.27 dB improvement over M-BPDN and M-FOCUSS respectively.

Figure 2 depicts the performance of Rademacher sparse signal matrix for different values of M , where we set $L = 20$ and $\text{SMNR} = 20$ dB. We again observe similar performance improvement as in the case of Gaussian sparse signal matrix. For example, for $M = 35$, MMV-FACS(M-

Figure 1. Performance of MMV-FACS, averaged over 1000 trials, for Gaussian sparse signal matrices with SMNR = 20 dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$, and number of measurement vectors $L = 20$.

OMP, M-BPDN) showed 7.56 dB and 4.32 dB over M-OMP and M-BPDN respectively.

A comparison of MMV reconstruction techniques is shown in figure 3 for Gaussian sparse signal matrix for different values of L , where we set $M = 50$ and SMNR = 20 dB. It may be observed that MMV-FACS gave a significant performance improvement over the participating algorithms. Specifically, MMV-FACS(M-OMP, M-SP) improved the performance by 5.77 and 4.94 dB over M-OMP and M-SP respectively.

To show the dependency of recovery performance on SMNR, we conducted simulations for different values of SMNR. Figure 4 illustrates the performance for Gaussian sparse signal matrix where $L = 10$ and $M = 45$. An additional ASRR improvement of 2.51 and 2.08 dB were achieved as compared to M-OMP and M-FOCUSS respectively for SMNR = 10 dB. This shows the robustness of MMV-FACS to noisy measurements.

From the above simulation results, it can be seen that MMV-FACS improved the sparse signal recovery compared to participating algorithms.

5.1c Reproducible research: We provide necessary Matlab codes to reproduce all the figures, publicly downloadable

from <http://www.ece.iisc.ernet.in/~sslab/Public/MMVFACS.tar.gz>.

5.2 Real compressible signals

To evaluate the performance of MMV-FACS on compressible signals and real world data, we used the data set “05091.dat” from MIT-BIH atrial fibrillation database [20]. The recording is for 10 h duration, and contains two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ± 10 mv. We selected the first 250 time points of the recording as the data set used in our experiment. We used a randomly generated Gaussian sensing matrix of size $M \times 250$ with different values of M in the experiment. We assumed sparsity level $K = 50$ and used M-OMP and M-SP as the participating algorithms. The reconstruction results are shown in figure 5.

Similar to synthetic signals, MMV-FACS shows a better ASRR compared to the participating algorithms M-OMP and M-SP. This demonstrates the advantage of MMV-FACS in real-life applications, requiring fewer measurement samples to yield an approximate reconstruction.

Figure 2. Performance of MMV-FACS, averaged over 1000 trials, for Rademacher sparse signal matrices with SMNR = 20 dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$ and number of measurement vectors $L = 20$.

Figure 3. Performance of MMV-FACS, averaged over 1000 trials, for Gaussian sparse signal matrices with SMNR = 20 dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$, and number of measurements $M = 50$.

6. Conclusions

In this paper, we extended FACS to the MMV case and showed that MMV-FACS improves sparse signal matrix reconstruction. Using RIP, we theoretically analysed the proposed scheme and derived sufficient conditions for the performance improvement over the participating

algorithm. Using Monte-Carlo simulations, we showed the performance improvement of the proposed scheme over the participating methods. Though this paper discusses only the extension of FACS for MMV problem, a similar approach can be used to extend the other fusion algorithms developed by Ambat *et al* [14, 15, 21].

$$\arg \max_{\mathbf{X} \in Q} \|\mathbf{AX}\|_F \leq \arg \max_{\mathbf{X} \in S} \|\mathbf{AX}\|_F.$$

To prove this, it is sufficient to ensure that $Q \subset S$.

Consider a matrix $\mathbf{X} \in Q$. Partition the support of $\mathbf{X}$ into sets of size $R + K$. Let set $\mathcal{I}_0$ contain the indices of the $R + K$ rows of $\mathbf{X}$, which have largest row ℓ_2 -norm, breaking ties lexicographically. Let set $\mathcal{I}_1$ contain the indices of the next largest (row ℓ_2 -norm) $R + K$ rows and so on. The final block $\mathcal{I}_J$ may have lesser than $R + K$ components. This partition gives rise to the following decomposition:

$$\mathbf{X} = \mathbf{X}|_{\mathcal{I}_0} + \sum_{j=1}^J \mathbf{X}|_{\mathcal{I}_j} = \lambda_0 \mathbf{Y}_0 + \sum_{j=1}^J \lambda_j \mathbf{Y}_j,$$

where $\lambda_j = \|\mathbf{X}|_{\mathcal{I}_j}\|_F$ and $\mathbf{Y}_j = \lambda_j^{-1} \mathbf{X}|_{\mathcal{I}_j}$.

By construction, each matrix $\mathbf{Y}_j$ belongs to S because it is $R + K$ sparse and has unit Frobenius norm. We will show that $\sum_j \lambda_j \leq 1$. This implies that $\mathbf{X}$ can be written as a convex combination of matrices from the set S . As a result $\mathbf{X} \in S$. Therefore, $Q \subset S$.

Fix some j in the range $\{1, 2, \dots, J\}$. Then, $\mathcal{I}_j$ contains at most $R + K$ elements and $\mathcal{I}_{j-1}$ contains exactly $R + K$ elements. Therefore,

$$\lambda_j = \|\mathbf{X}|_{\mathcal{I}_j}\|_F \leq \sqrt{R+K} \|\mathbf{X}|_{\mathcal{I}_j}\|_{2,\infty} \leq \sqrt{R+K} \cdot \frac{1}{R+K} \|\mathbf{X}|_{\mathcal{I}_{j-1}}\|_{2,1}.$$

The last inequality holds because the row ℓ_2 -norm of $\mathbf{X}$ on the set $\mathcal{I}_{j-1}$ dominates its largest row ℓ_2 -norm in $\mathcal{I}_j$. Summing these relations, we get

$$\sum_{j=1}^J \lambda_j \leq \frac{1}{\sqrt{R+K}} \sum_{j=1}^J \|\mathbf{X}|_{\mathcal{I}_{j-1}}\|_{2,1} \leq \frac{1}{\sqrt{R+K}} \|\mathbf{X}\|_{2,1}.$$

Also, we have $\lambda_0 = \|\mathbf{X}|_{\mathcal{I}_0}\|_F \leq \|\mathbf{X}\|_F$. Since $\mathbf{X} \in Q$, we conclude that

$$\sum_{j=0}^J \lambda_j \leq \left[\|\mathbf{X}\|_F + \frac{1}{\sqrt{R+K}} \|\mathbf{X}\|_{2,1} \right] \leq 1.$$

□

Proof of lemma 2

Let $\mathbf{y}^{(i)}$ denote the i th column of matrix $\mathbf{Y}$ and $\mathbf{r}^{(i)}$ denote the i th column of matrix $\mathbf{R}$, $i = 1, 2, \dots, L$. Then, we have from lemma 2 of [22]

$$\left(1 - \frac{\delta_{|\mathcal{I}_1|+|\mathcal{I}_2|}}{1 - \delta_{|\mathcal{I}_1|+|\mathcal{I}_2|}}\right) \|\mathbf{y}^{(i)}\|_2 \leq \|\mathbf{r}^{(i)}\|_2 \leq \|\mathbf{y}^{(i)}\|_2.$$

Summing the above relation, we obtain

$$\left(1 - \frac{\delta_{|\mathcal{I}_1|+|\mathcal{I}_2|}}{1 - \delta_{|\mathcal{I}_1|+|\mathcal{I}_2|}}\right) \sum_{i=1}^L \|\mathbf{y}^{(i)}\|_2^2 \leq \sum_{i=1}^L \|\mathbf{r}^{(i)}\|_2^2 \leq \sum_{i=1}^L \|\mathbf{y}^{(i)}\|_2^2.$$

Equivalently, we have

$$\left(1 - \frac{\delta_{|\mathcal{I}_1|+|\mathcal{I}_2|}}{1 - \delta_{|\mathcal{I}_1|+|\mathcal{I}_2|}}\right) \|\mathbf{Y}\|_F \leq \|\mathbf{R}\|_F \leq \|\mathbf{Y}\|_F.$$

□

References

- [1] Cotter S F, Rao B D, Engan K and Kreutz-Delgado K 2005 Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* 53(7): 2477–2488
- [2] Gorodnitsky I F, George J S and Rao B D 1995 Neuro-magnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm. *J. Electroencephal. Clin. Neurophysiol.* 95(4): 231–251
- [3] Gorodnitsky I F and Rao B D 1997 Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Trans. Signal Process.* 45(3): 600–616
- [4] Malioutov D, Cetin M and Willsky A S 2005 A Sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* 53(8): 3010–3022
- [5] Stoica P and Moses R L 1997 *Introduction to spectral analysis*. Upper Saddle River, N.J.: Prentice Hall
- [6] Cotter S F and Rao B D 2002 Sparse channel estimation via matching pursuit with application to equalization. *IEEE Trans. Commun.* 50(3): 374–377
- [7] Tropp J A 2006 Algorithms for simultaneous sparse approximation. Part II: Convex relaxation. *Signal Process.* 86(3): 589–602 [Sparse Approximations in Signal and Image Processing Sparse Approximations in Signal and Image Processing]
- [8] Wipf D P and Rao B D 2007 An empirical bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* 55(7): 3704–3716
- [9] Zhang Z and Rao B D 2011 Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning. *IEEE J. Sel. Topics Signal Process.* 5(5): 912–926
- [10] Elad M and Yavneh I (2009) A plurality of sparse representations is better than the sparsest one alone. *IEEE Trans. Inf. Theory* 55(10): 4701–4714
- [11] Ambat S K, Chatterjee S and K V S Hari 2013a Fusion of algorithms for compressed sensing. *IEEE Trans. Signal Process.* 61(14): 3699–3704
- [12] Ambat S K, Chatterjee S and K V S Hari 2013b Fusion of algorithms for compressed sensing. In: *Acoustics, speech and signal processing (ICASSP), 2013 IEEE International Conference on*, pp. 5860–5864
- [13] Ambat S K, Chatterjee S and K V S Hari 2014 Progressive fusion of reconstruction algorithms for low latency applications in compressed sensing. *Signal Process.* 97:146–151

- [14] Ambat S K, Saikat Chatterjee and K V S Hari 2012 fusion of greedy pursuits for compressed sensing signal reconstruction. In: *20th European Signal Processing Conference 2012 (EUSIPCO 2012)*, Bucharest, Romania
- [15] Ambat S K, Saikat Chatterjee and K V S Hari 2014a A committee machine approach for compressed sensing signal reconstruction. *IEEE Trans. Signal Process.* 62(7): 1705–1717
- [16] Needell D and Tropp J A 2009 CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* 26(3): 301–321
- [17] Gribonval R, Rauhut H, Schnass K and Vandergheynst P 2008 Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms. *J. Fourier Anal. Appl.* 14(5): 655–687
- [18] van den Berg E and Friedlander M 2009 Probing the Pareto Frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* 31(2): 890–912
- [19] van den Berg E and Friedlander M P 2007 SPGL1: A solver for large-scale sparse reconstruction. <http://www.cs.ubc.ca/labs/scl/spgl1>
- [20] Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C-K and Stanley H E 2000 PhysioBank, physiotookit, and physioNet: Components of a new research resource for complex physiologic signals. *Circulation* 101(23): e215–e220. *Circulation electronic pages*: <http://circ.ahajournals.org/cgi/content/full/101/23/e215> PMID:1085218; doi:10.1161/01.CIR.101.23.e215
- [21] Ambat S K, Saikat Chatterjee and K V S Hari 2014b Progressive fusion of reconstruction algorithms for low latency applications in compressed sensing. *Signal Process.* 97: 146–151
- [22] Dai W and Milenkovic O 2009 Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory* 55(5): 2230–2249