

Orthogonal subspace based fast iterative thresholding algorithms for joint sparsity recovery

Ningning Han, Shidong Li, and Jian Lu *Member, IEEE*

Abstract

Sparse signal recoveries from multiple measurement vectors (MMV) with joint sparsity property have many applications in signal, image, and video processing. The problem becomes much more involved when snapshots of the signal matrix are temporally correlated. With signal's temporal correlation in mind, we provide a framework of iterative MMV algorithms based on thresholding, functional feedback and null space tuning. Convergence analysis for exact recovery is established. Unlike most of iterative greedy algorithms that select indices in a measurement/solution space, we determine indices based on an orthogonal subspace spanned by the iterative sequence. In addition, a functional feedback that controls the amount of energy relocation from the “tails” is implemented and analyzed. It is seen that the principle of functional feedback is capable to lower the number of iteration and speed up the convergence of the algorithm. Numerical experiments demonstrate that the proposed algorithm has a clearly advantageous balance of efficiency, adaptivity and accuracy compared with other state-of-the-art algorithms.

Index Terms

Multiple measurement vectors, null space tuning, thresholding, feedback, orthogonal subspace.

I. INTRODUCTION

In sparse reconstruction signal models with joint sparsity property, signals are sampled at L time instances, resulting in the multiple measurement vector (MMV) model:

$$Y = \Phi X + E, \quad (1)$$

where $Y \in \mathbb{C}^{M \times L}$ is the observation matrix containing L measurement/snapshot (column) vectors, $\Phi \in \mathbb{C}^{M \times N}$ is the measurement matrix governed by the specific physical system, and $X \in \mathbb{C}^{N \times L}$ is the underlying source signal matrix, to be recovered. $E \in \mathbb{C}^{M \times L}$ is an additive measurement noise matrix.

In this system, L measurements share the same row support and elements in each nonzero row of X are temporally correlated. The solution problem to a noiseless MMV model can be formulated as

$$\min_X \|X\|_0 \text{ s.t. } Y = \Phi X, \quad (2)$$

where $\|X\|_0 = |\text{supp}(X)|$, $\text{supp}(X) = \{1 \leq i \leq N : X_{i \cdot} \neq 0\}$, $X_{i \cdot}$ is the i -th row of X . In [1], the authors have shown that X is the unique solution of (2) if

$$\|X\|_0 < \frac{\text{spark}(\Phi) + \text{rank}(Y) - 1}{2}, \quad (3)$$

where $\text{spark}(\Phi)$ is the smallest number of linearly dependent columns of Φ .

A large majority of effective algorithms for solving (2) are based on two strategies: extending single measurement vector (SMV) algorithms or exploiting signal subspaces. Well-known algorithms of the first class include simultaneous orthogonal matching pursuit (SOMP) [2]-[5], mixed norm minimization techniques [6]-[14], simultaneous greedy algorithms [15], [16]. However, these algorithms, without exploiting subspace structures or temporal correlations, have not offered realistic improvements over performances than that of SMV cases. Recently, a multiple sparse Bayesian learning (MSBL) algorithm [17]-[21], as an extension of sparse Bayesian SMV algorithms, is seen to improve recovery performances by modeling temporal correlation of sparse vectors. Another strategy is to exploit subspace structures spanned by measurement vectors. Representative algorithms include, e.g., sequential compressive MUSIC (SeqCS-MUSIC) [22], [23], subspace-augmented MUSIC (SA-MUSIC+OSMP) [24], rank aware order recursive matching pursuit (RA-ORMP) [25], [26], [27], semi-supervised MUSIC (SS-MUSIC) [28] etc.

In this report, we provide a computationally efficient “greedy” algorithm for joint sparsity signal recoveries from their multiple measurement vectors. The proposed algorithm combines procedures of hard thresholding (HT), functional feedback (f -FB) for “tail” energy shrinkage and enhanced feasibility, the null space tuning (NST), and a novel variable selection mechanism. The novel criterion of variable selection is based on estimations of significant coefficients in an orthogonal subspace of the iterative

This work was supported by the National Natural Science Foundation of China under grants 61972265, 11871348 and 61373087, by the Natural Science Foundation of Guangdong Province of China under grant 2020B1515310008, by the Educational Commission of Guangdong Province of China under grant 2019KZDZX1007, and by the Guangdong Key Laboratory of Intelligent Information Processing, China, and the NSF of USA (DMS-1615288).

Ningning Han (ningninghan@szu.edu.cn) and Jian Lu (*corresponding author*, jianlu@szu.edu.cn) are with Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen, 518060.

Shidong Li (shidong@sfsu.edu) is with Department of Mathematics, San Francisco State University, San Francisco, CA94132.

sequence. The cardinality of selected variables is determined by the feedback function f . Experimental results show that the proposed algorithm provides superior performances in terms of the efficiency and the critical sparsity (i.e., the maximum sparsity level at which the perfect recovery is guaranteed [29]). In fact, the rate of successful recovery of our algorithm has broken through the algebraic upper bound given in (3).

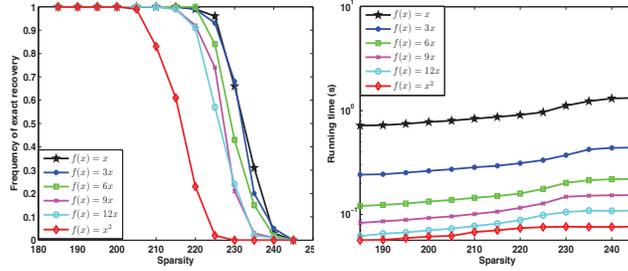


Fig. 1. Left: Frequency of exact recovery as a function of sparsity; right: running time as a function of sparsity.

II. ORTHOGONAL SUBSPACE NST+HT+ f -FB ALGORITHM

A. Notations

A submatrix of Φ with columns indexed by a set I is denoted by Φ_I and a submatrix of Φ with rows indexed by a set J is denoted by $\Phi_{(J)}$. We denote the i -th row and the j -th column of a matrix Φ by $\Phi_{i,\cdot}$ and $\Phi_{\cdot,j}$, respectively. $T\Delta T'$ is the symmetric difference of T and T' , i.e., $T\Delta T' = (T \setminus T') \cup (T' \setminus T)$. $\mathbf{H}_T(X)$ is a linear operator that sets all but elements belong to rows indexed by T of X to zero.

Algorithm 1 OSNST+HT+ f -FB

Input: $\Phi, Y, \epsilon, f(\cdot), K$;

Output: W ;

Initialize: $k = 1, W^0 = 0$;

While $\|Y - \Phi W^{k-1}\|_2 > \epsilon$ and $k < K$ **do**

$$X^k = W^{k-1} + \Phi^*(\Phi\Phi^*)^{-1}(Y - \Phi W^{k-1});$$

$$Q^k = \text{orth}(X^k);$$

$$T_k = \{\text{Indices of } f(k) \text{ largest } \|Q_{i,\cdot}^k\|_2\};$$

$$W_{T_k}^k = X_{T_k}^k + (\Phi_{T_k}^* \Phi_{T_k})^{-1} \Phi_{T_k}^* \Phi_{T_k^c} X_{T_k^c}^k;$$

$$W_{T_k^c}^k = 0;$$

$$k = k + 1;$$

end while;

B. Algorithm framework

The iterative framework of approximation and null space tuning (NST) algorithms is as follows

$$\begin{cases} W^k = \mathbb{D}(X^k), \\ X^{k+1} = X^k + \mathbb{P}(W^k - X^k). \end{cases}$$

Here $\mathbb{D}(X^k)$ approximates the desired solution by various principles, and $\mathbb{P} := I - \Phi^*(\Phi\Phi^*)^{-1}\Phi$ is the orthogonal projection onto $\ker(\Phi)$.

Since the sequence $\{X^k\}$ is always feasible (i.e., $Y = \Phi X^k$) under the NST principle, one may split Y as

$$Y = \Phi X = \Phi_{T_k} X_{(T_k)}^k + \Phi_{T_k^c} X_{(T_k^c)}^k,$$

where T_k includes indices of $f(k)$ largest $\|Q_{i,\cdot}^k\|_2$ ($i \in \{1, \dots, N\}$), $f(\cdot) \geq 0$ is a non-decreasing function and columns of Q^k are an orthonormal basis for the column space of X^k , i.e., $Q^k = \text{orth}(X^k)$. The mechanism of feedback is to feed the contribution of $\Phi_{T_k^c} X_{(T_k^c)}^k$ to Y back to $\text{im}(\Phi_{T_k})$, the image of Φ_{T_k} . A straightforward way is to set

$$\Lambda^k = \arg \min_{\Lambda} \|\Phi_{T_k} \Lambda - \Phi_{T_k^c} X_{(T_k^c)}^k\|_2,$$

which has the best/least-square solution

$$\Lambda^k = (\Phi_{T_k}^* \Phi_{T_k})^{-1} \Phi_{T_k}^* \Phi_{T_k^c} X_{(T_k^c)}^k.$$

The orthogonal subspace iterative thresholding algorithm with functional feedback and null space tuning (OSNST+HT+ f -FB) is then established in *Algorithm 1*.

C. Convergence analysis

In this paper, we assume the number of snapshots is smaller than the dimension of measurement, i.e., $L < M$, and the measurement matrix Y is full column rank, i.e., $\text{rank}(Y) = L$. We now turn to the convergence of OSNST+HT+f-FB.

Definition 1. [30]. For each integer $s = 1, 2, \dots$, the restricted isometry constant (RIC) δ_s of a matrix Φ is defined as the smallest number δ_s such that

$$(1 - \delta_s)\|X\|_F^2 \leq \|\Phi X\|_F^2 \leq (1 + \delta_s)\|X\|_F^2$$

holds for all s row-sparse matrix X . Equivalently, it is given by

$$\delta_s = \max_{|S| \leq s} \|I - \Phi_S^* \Phi_S\|_2.$$

Definition 2. [31]. For each integer $s = 1, 2, \dots$ the preconditioned restricted isometry constant γ_s of a matrix A is defined as the smallest number γ_s such that

$$(1 - \gamma_s)\|X\|_F^2 \leq \|(\Phi\Phi^*)^{-\frac{1}{2}}\Phi X\|_F^2$$

holds for all s row-sparse matrix X . In fact, the preconditioned restricted isometry constant γ_s represents the restricted isometry property of the preconditioned matrix $(\Phi\Phi^*)^{-\frac{1}{2}}\Phi$. Since

$$\|(\Phi\Phi^*)^{-\frac{1}{2}}\Phi X\|_F \leq \|(\Phi\Phi^*)^{-\frac{1}{2}}\Phi\|_2 \|X\|_F = \|X\|_F,$$

γ_s is actually the smallest number such that, for all s row-sparse matrix X ,

$$(1 - \gamma_s)\|X\|_F^2 \leq \|(\Phi\Phi^*)^{-\frac{1}{2}}\Phi X\|_F^2 \leq (1 + \gamma_s)\|X\|_F^2.$$

It indicates $\gamma_s(\Phi) = \delta_s((\Phi\Phi^*)^{-\frac{1}{2}}\Phi)$. Equivalently, it is given by

$$\gamma_s = \max_{|S| \leq s} \|I - \Phi_S^*(\Phi\Phi^*)^{-1}\Phi_S\|_2.$$

Definition 3. Let the feasible solution space of (2) be $\mathcal{X} = \{X \in \mathbb{C}^{N \times L} : Y = \Phi X\}$. Define the modified matrix condition number of \mathcal{X} by $\alpha = \max_{X \in \mathcal{X}} \frac{\sigma_{\max}(X)}{\sigma_{\min}(X)}$, where $\sigma_{\min}(X)$ and $\sigma_{\max}(X)$ denote the smallest and the largest nonzero singular values of X , respectively.

Lemma 4. Let $U, V \in \mathbb{C}^{N \times L}$ with $|supp(U) \cup supp(V)| \leq t$, then $|\langle U, (I - \Phi^*\Phi)V \rangle| \leq \delta_t \|U\|_F \|V\|_F$. Suppose $|R \cup supp(V)| \leq t$, then $\|[(I - \Phi^*\Phi)V]_{(R)}\|_F \leq \delta_t \|V\|_F$.

Proof. Let $T = supp(U) \cup supp(V)$, we then have

$$\begin{aligned} |\langle U, (I - \Phi^*\Phi)V \rangle| &= |\langle U, V \rangle - \langle \Phi U, \Phi V \rangle| \\ &= |\langle U_{(T)}, V_{(T)} \rangle - \langle \Phi_T U_{(T)}, \Phi_T V_{(T)} \rangle| \\ &= |\langle U_{(T)}, (I - \Phi_T^* \Phi_T) V_{(T)} \rangle| \\ &\leq \|U_{(T)}\|_F \|(I - \Phi_T^* \Phi_T) V_{(T)}\|_F \\ &\leq \|U_{(T)}\|_F \|I - \Phi_T^* \Phi_T\|_2 \|V_{(T)}\|_F \\ &\leq \delta_t \|U\|_F \|V\|_F. \end{aligned}$$

The first and the second inequalities are due to the Cauchy-Schwarz inequality, and the sub-multiplicativity of matrix norms, respectively. The last step is by *Definition 1*. It then follows that

$$\|[(I - \Phi^*\Phi)V]_{(R)}\|_F^2 = \langle (\mathbf{H}_R ((I - \Phi^*\Phi)V), (I - \Phi^*\Phi)V \rangle \leq \delta_t \|[(I - \Phi^*\Phi)V]_{(R)}\|_F \|V\|_F.$$

Therefore, $\|[(I - \Phi^*\Phi)V]_{(R)}\|_F \leq \delta_t \|V\|_F$. \square

Remark 5. Let γ_t be the P-RIP constant of Φ and $U, V \in \mathbb{C}^{N \times L}$ with $|supp(U) \cup supp(V)| \leq t$, then $|\langle U, (I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)V \rangle| \leq \gamma_t \|U\|_F \|V\|_F$. Suppose $|R \cup supp(V)| \leq t$, then $\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)V]_{(R)}\|_F \leq \gamma_t \|V\|_F$.

Lemma 6. For $E \in \mathbb{C}^{M \times L}$, $\|[\Phi^*(\Phi\Phi^*)^{-1}E]_{(T)}\|_F \leq \sqrt{1 + \theta_t} \|E\|_F$, where $\theta_t = \delta_t((\Phi\Phi^*)^{-1}\Phi)$ and $\delta_t((\Phi\Phi^*)^{-1}\Phi)$ is RIC of matrix $(\Phi\Phi^*)^{-1}\Phi$.

Proof.

$$\begin{aligned} \|[\Phi^*(\Phi\Phi^*)^{-1}E]_{(T)}\|_F^2 &= \langle \Phi^*(\Phi\Phi^*)^{-1}E, \mathbf{H}_T(\Phi^*(\Phi\Phi^*)^{-1}E) \rangle \\ &= \langle E, (\Phi\Phi^*)^{-1}\Phi \mathbf{H}_T(\Phi^*(\Phi\Phi^*)^{-1}E) \rangle \\ &\leq \|E\|_F \sqrt{1 + \theta_t} \|[\Phi^*(\Phi\Phi^*)^{-1}E]_{(T)}\|_F. \end{aligned}$$

Applying *Definition 1* to the matrix $\Phi^*(\Phi\Phi^*)^{-1}$ obtains the last step. Hence, for all $E \in \mathbb{C}^{M \times L}$, we have $\|[\Phi^*(\Phi\Phi^*)^{-1}E]_{(T)}\|_F \leq \sqrt{1 + \theta_t} \|E\|_F$. \square

Lemma 7. Let $Y = \Phi X + E$, where $X \in \mathbb{C}^{N \times L}$ is s row-sparse with $S = \text{supp}(X)$ and $E \in \mathbb{C}^{M \times L}$ is the measurement error. If $\tilde{W} \in \mathbb{C}^{N \times L}$ is \tilde{s} row-sparse, $\tilde{X} = \tilde{W} + \Phi^*(\Phi\Phi^*)^{-1}(Y - \Phi\tilde{W})$, $\tilde{Q} = \text{orth}(\tilde{X})$, and T is an index set of $t \geq s$ largest $\|\tilde{Q}_i\|_2$, then

$$\|X_{(T^c)}\|_F \leq \sqrt{2}\alpha(\gamma_{s+\tilde{s}+t}\|X - \tilde{W}\|_F + \sqrt{1 + \theta_{t+s}}\|E\|_F),$$

where $\theta_{t+s}(\Phi) = \delta_{t+s}((\Phi\Phi^*)^{-1}\Phi)$.

Proof. Since $\text{rank}(Y) = L$ and $Y = \Phi\tilde{X}$, it is obvious that $\text{rank}(\tilde{X}) = L$. Consequently, the singular value decomposition of \tilde{X} can be denoted as $\tilde{X} = \tilde{U}_\ell \tilde{\Sigma}_{(\ell)} \tilde{V}^*$, where \tilde{U}_ℓ is the first L columns of \tilde{U} and $\tilde{\Sigma}_{(\ell)}$ denotes the first L rows of $\tilde{\Sigma}$. Since \tilde{U}_ℓ can be regarded as an orthonormal basis for the range of \tilde{X} , without loss of generality, let $\tilde{Q} = \tilde{U}_\ell$, we have

$$\|[\tilde{X}\tilde{V}\tilde{\Sigma}_{(\ell)}^{-1}]_{(T)}\|_F \geq \|[\tilde{X}\tilde{V}\tilde{\Sigma}_{(\ell)}^{-1}]_{(S)}\|_F.$$

It then follows that

$$\tilde{\sigma}_{\min}^{-1}\|\tilde{X}_{(T)}\|_F \geq \tilde{\sigma}_{\max}^{-1}\|\tilde{X}_{(S)}\|_F,$$

where $\tilde{\sigma}_{\min}$ and $\tilde{\sigma}_{\max}$ denote the smallest and the largest singular value of $\tilde{\Sigma}_{(\ell)}$. Eliminating the common terms over $T \cap S$, we obtain

$$\tilde{\sigma}_{\min}^{-1}\|[\tilde{W} + \Phi^*(\Phi\Phi^*)^{-1}(Y - \Phi\tilde{W})]_{(T \setminus S)}\|_F \geq \tilde{\sigma}_{\max}^{-1}\|[\tilde{W} + \Phi^*(\Phi\Phi^*)^{-1}(Y - \Phi\tilde{W})]_{(S \setminus T)}\|_F.$$

For the left hand,

$$\begin{aligned} & \tilde{\sigma}_{\min}^{-1}\|[\tilde{W} + \Phi^*(\Phi\Phi^*)^{-1}(Y - \Phi\tilde{W})]_{(T \setminus S)}\|_F \\ &= \tilde{\sigma}_{\min}^{-1}\|[\tilde{W} - X + \Phi^*(\Phi\Phi^*)^{-1}(\Phi X + E - \Phi\tilde{W})]_{(T \setminus S)}\|_F \\ &= \tilde{\sigma}_{\min}^{-1}\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)(\tilde{W} - X) + \Phi^*(\Phi\Phi^*)^{-1}E]_{(T \setminus S)}\|_F. \end{aligned}$$

The right hand satisfies

$$\begin{aligned} & \tilde{\sigma}_{\max}^{-1}\|[\tilde{W} + \Phi^*(\Phi\Phi^*)^{-1}(Y - \Phi\tilde{W})]_{(S \setminus T)}\|_F \\ &= \tilde{\sigma}_{\max}^{-1}\|[\tilde{W} + \Phi^*(\Phi\Phi^*)^{-1}(\Phi X + E - \Phi\tilde{W}) + X - X]_{(S \setminus T)}\|_F \\ &\geq \tilde{\sigma}_{\max}^{-1}\|X_{(S \setminus T)}\|_F - \tilde{\sigma}_{\max}^{-1}\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)(\tilde{W} - X) + \Phi^*(\Phi\Phi^*)^{-1}E]_{(S \setminus T)}\|_F. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} & \tilde{\sigma}_{\max}^{-1}\|X_{(S \setminus T)}\|_F \\ &\leq \tilde{\sigma}_{\max}^{-1}\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)(\tilde{W} - X) + \Phi^*(\Phi\Phi^*)^{-1}E]_{(S \setminus T)}\|_F \\ &+ \tilde{\sigma}_{\min}^{-1}\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)(\tilde{W} - X) + \Phi^*(\Phi\Phi^*)^{-1}E]_{(T \setminus S)}\|_F \\ &\leq \sqrt{2}\tilde{\sigma}_{\min}^{-1}\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)(\tilde{W} - X) + \Phi^*(\Phi\Phi^*)^{-1}E]_{(T \Delta S)}\|_F \\ &\leq \sqrt{2}\tilde{\sigma}_{\min}^{-1}\|[(I - \Phi^*(\Phi\Phi^*)^{-1}\Phi)(\tilde{W} - X)]_{(T \Delta S)}\|_F + \sqrt{2}\tilde{\sigma}_{\min}^{-1}\|[\Phi^*(\Phi\Phi^*)^{-1}E]_{(T \Delta S)}\|_F \\ &\leq \sqrt{2}\tilde{\sigma}_{\min}^{-1}(\gamma_{s+\tilde{s}+t}\|X - \tilde{W}\|_F + \sqrt{1 + \theta_{t+s}}\|E\|_F). \end{aligned}$$

The last step is due to *Remark 5* and *Lemma 6*. In view of *Definition 3*, we derive

$$\|X_{(S \setminus T)}\|_F \leq \sqrt{2}\alpha(\gamma_{s+\tilde{s}+t}\|X - \tilde{W}\|_F + \sqrt{1 + \theta_{t+s}}\|E\|_F). \quad \square$$

Lemma 8. Let $Y = \Phi X + E$, where $X \in \mathbb{C}^{N \times L}$ is s row-sparse signal matrix, and $E \in \mathbb{C}^{M \times L}$ is the measurement error. Let $S = \text{supp}(X)$ be the index set of the s sparse rows of X . Denote by $\tilde{Q} = \text{orth}(\tilde{X})$ the orthogonal basis of the row-space of X , and T the index set of $t \geq s$ largest values of $\|\tilde{Q}_i\|_2$. If \overline{W} is the feedback of \tilde{X} given by $\overline{W}_{(T)} = \tilde{X}_{(T)} + (\Phi_T^* \Phi_T)^{-1} \Phi_T^* \Phi_{T^c} \tilde{X}_{(T^c)}$ and $\overline{W}_{(T^c)} = 0$, then

$$\|(X - \overline{W})\|_F \leq \frac{\|X_{(T^c)}\|_F}{\sqrt{1 - \delta_{s+t}^2}} + \frac{\sqrt{1 + \delta_t}\|E\|_F}{1 - \delta_{s+t}}.$$

Proof. For any $Z \in \mathbb{C}^{N \times L}$ supported on T ,

$$\begin{aligned} & \langle \Phi \overline{W} - Y, \Phi Z \rangle \\ &= \langle \Phi_T \tilde{X}_{(T)} + \Phi_T (\Phi_T^* \Phi_T)^{-1} \Phi_T^* \Phi_{T^c} \tilde{X}_{(T^c)} - Y, \Phi_T Z_{(T)} \rangle \\ &= \langle \Phi_T^* (\Phi_T \tilde{X}_{(T)} + \Phi_{T^c} \tilde{X}_{(T^c)} - Y), Z_{(T)} \rangle \\ &= \langle \Phi_T^* (\Phi \tilde{X} - Y), Z_{(T)} \rangle \\ &= 0. \end{aligned}$$

The last step is due to the feasibility of \tilde{X} . The inner product can also be written as $\langle \Phi \overline{W} - Y, \Phi Z \rangle = \langle (\Phi \overline{W} - \Phi X - E), \Phi Z \rangle = 0$. Therefore, $\langle (\overline{W} - X), \Phi^* \Phi Z \rangle = \langle E, \Phi Z \rangle$, $\forall Z \in \mathbb{C}^{N \times L}$ supported on T . Since $(\overline{W} - X)_T$ is supported on T , one has $\langle (\overline{W} - X), \Phi^* \Phi_T (\overline{W} - X)_{(T)} \rangle = \langle E, \Phi_T (\overline{W} - X)_{(T)} \rangle$.

Consequently,

$$\begin{aligned} \|(\overline{W} - X)_{(T)}\|_F^2 &= \langle (\overline{W} - X), \mathbf{H}_T(\overline{W} - X) \rangle \\ &= |\langle (X - \overline{W}), (I - \Phi^* \Phi) \mathbf{H}_T(X - \overline{W}) \rangle| + |\langle E, \Phi \mathbf{H}_T(X - \overline{W}) \rangle| \\ &\leq \delta_{s+t} \|X - \overline{W}\|_F \|(X - \overline{W})_{(T)}\|_F + \sqrt{1 + \delta_t} \|E\|_F \|(X - \overline{W})_{(T)}\|_F. \end{aligned}$$

The last step is due to *Lemma 4* and *Definition 1*. We can obtain

$$\|(X - \overline{W})_{(T)}\|_F \leq \delta_{s+t} \|X - \overline{W}\|_F + \sqrt{1 + \delta_t} \|E\|_F.$$

It then follows that

$$\begin{aligned} \|(X - \overline{W})\|_F^2 &= \|(X - \overline{W})_{(T)}\|_F^2 + \|(X - \overline{W})_{(T^c)}\|_F^2 \\ &\leq (\delta_{s+t} \|X - \overline{W}\|_F + \sqrt{1 + \delta_t} \|E\|_F)^2 + \|X_{(T^c)}\|_F^2. \end{aligned}$$

This in turn implies $p(\|X - \widetilde{W}\|_F) \leq 0$, where $p(\cdot)$ is a quadratic polynomial, defined by

$$p(x) = (1 - \delta_{s+t}^2)x^2 - 2\delta_{s+t}\sqrt{1 + \delta_t}\|E\|_F x - (1 + \delta_t)\|E\|_F^2 - \|X_{(T^c)}\|_F^2.$$

Since $(1 - \delta_{s+t}^2) \geq 0$, it means that $\|(X - \overline{W})\|_F$ is smaller than the largest root of $p(\cdot)$

$$\begin{aligned} \|(X - \overline{W})\|_F &\leq \frac{\delta_{s+t}\sqrt{1 + \delta_t}\|E\|_F + \sqrt{(1 + \delta_t)\|E\|_F^2 + (1 - \delta_{s+t}^2)\|X_{(T^c)}\|_F^2}}{1 - \delta_{s+t}^2} \\ &\leq \frac{\|X_{(T^c)}\|_F}{\sqrt{1 - \delta_{s+t}^2}} + \frac{\sqrt{1 + \delta_t}\|E\|_F}{1 - \delta_{s+t}}. \end{aligned}$$

□

Theorem 9. *Let $Y = \Phi X + E$, where X is the s row-sparse signal matrix. Then the sequence $\{W^k\}$ produced by OSNST+HT+f-FB satisfies*

$$\|(X - W^k)\|_F \leq \rho_{s+f(k)+f(k-1)}^k \|X - W^0\|_F + \frac{\kappa_{s+f(k)+f(k-1)}(1 - \rho_{s+f(k)+f(k-1)}^k)}{1 - \rho_{s+f(k)+f(k-1)}} \|E\|_F,$$

where $\rho_\ell = \sqrt{\frac{2\alpha^2\gamma_\ell^2}{1 - \delta_\ell^2}}$ and $\kappa_\ell = (\frac{\sqrt{1 + \delta_\ell}}{1 - \delta_\ell} + \frac{\sqrt{2\alpha^2(1 + \theta_\ell)}}{\sqrt{1 - \delta_\ell^2}})$.

Proof. Applying *Lemma 7* to $\widetilde{W} = W^{k-1}$ and $T = T_k$ gives

$$\|X_{(T_k^c)}\|_F \leq \sqrt{2}\alpha(\gamma_{s+f(k-1)+f(k)}\|X - W^{k-1}\|_F + \sqrt{1 + \theta_{s+f(k)}}\|E\|_F),$$

and setting $\overline{W} = W^k$ and $T = T_k$ in *Lemma 8* obtains

$$\|(X - W^k)\|_F \leq \frac{\|X_{(T_k^c)}\|_F}{\sqrt{(1 - \delta_{s+f(k)})}} + \frac{\sqrt{1 + \delta_{f(k)}}\|E\|_F}{1 - \delta_{s+f(k)}}.$$

Combining these two inequalities, we have

$$\|(X - W^k)\|_F \leq \sqrt{\frac{2\alpha^2\gamma_{s+f(k)+f(k-1)}^2}{(1 - \delta_{s+f(k)})}} \|X - W^{k-1}\|_F + (\frac{\sqrt{1 + \delta_{f(k)}}}{1 - \delta_{s+f(k)}} + \frac{\sqrt{2\alpha^2(1 + \theta_{s+f(k)})}}{\sqrt{1 - \delta_{s+f(k)}^2}}) \|E\|_F.$$

Since δ_ℓ and γ_ℓ are all non-decreasing [30], ρ_ℓ and κ_ℓ are also all non-decreasing as ℓ increases for all integer ℓ . Note that $f(\ell)$ is also a nondecreasing function, it then follows that

$$\|(X - W^k)\|_F \leq \rho_{s+f(k)+f(k-1)}^k \|X - W^0\|_F + \frac{\kappa_{s+f(k)+f(k-1)}(1 - \rho_{s+f(k)+f(k-1)}^k)}{1 - \rho_{s+f(k)+f(k-1)}} \|E\|_F.$$

□

Consequently, if the RIP and the P-RIP of the matrix Φ obeys $2\alpha^2\gamma_{s+f(k)+f(k-1)}^2 + \delta_{s+f(k)+f(k-1)}^2 < 1$, the OSNST+HT+FB algorithm is guaranteed to converge.

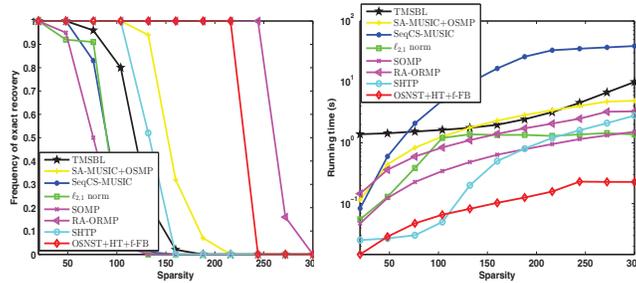


Fig. 2. Left: Frequency of exact recovery as a function of sparsity; right: running time as a function of sparsity.

III. EXPERIMENTS

In this experiment, the measurement matrix Φ is an 300×1000 Gaussian random matrix and the number of snapshots is 10. To model the temporal correlation of MMV problem, we employ an autoregressive process of order 1, AR(1). As a result, the j -th snapshot X_j is generated according to the model

$$X_j = \beta X_{(j-1)} + (1 - \beta)\epsilon_j,$$

where β is the AR model parameter controlling the temporal correlation and ϵ_j is the level of white Gaussian perturbation. The support of a sparse signal is also chosen randomly and the nonzero entries of Gaussian sparse signals are drawn independently from the Gaussian distribution with zero mean and unit variance. A successful recovery is recorded when $\|X - \hat{X}\|_F / \|X\|_F \leq 10^{-4}$, where X is the exact signal matrix and \hat{X} denotes the recovered signal. Each experiment is tested for 100 (random) trials. A matlab implementation of the proposed algorithm is also available at

<https://www.dropbox.com/s/2avudk770m4c6rz/OSNST%2BHT%2Bf-FB.zip?dl=0>.

We first study the mechanisms of f -feedback by introducing six particular index selection functions: $f(x) = x$, $f(x) = 3x$, $f(x) = 6x$, $f(x) = 9x$, $f(x) = 12x$ and $f(x) = x^2$. As discussed, higher critical sparsity represents better empirical recovery performance. Figure 1 shows the frequency of exact recovery and the running time as functions of the sparsity levels s . As shown, linear functions with modest gradients present similar performance, which is better than the quadratic function $f(x) = x^2$. In addition, one can accelerate the convergence of the class of OSNST+HT+ f -FB algorithms by adjusting the cardinality of indices per iteration.

Also presented are comparisons among our OSNST+HT+ f -FB and state-of-the-art techniques such as SOMP [2], $\ell_{2,1}$ norm [8], SHTP [15], [16], RA-ORMP [10], TMSBL [17], SA-MUSIC+OSMP [24], SeqCS-MUSIC [22], [23] in terms of frequency of exact recovery and running time. In this experiment, we adopt a modest setting $f(x) = 6x$, which can be applied to other applications. In Figure 2, experimental results show that OSNST+HT+ f -FB still delivers reasonable performance better than that of SOMP, $\ell_{2,1}$ norm, SHTP, TMSBL, SA-MUSIC+OSMP, and SeqCS-MUSIC, though slightly under-performs that of RA-ORMP. For the execution-time comparison, our algorithm achieves the best performance. Numerical experiments show that our algorithm has a clearly advantageous balance of efficiency, adaptivity and accuracy compared with other state-of-the-art algorithms.

REFERENCES

- [1] 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, Dec. 2006.
- [2] 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.
- [3] J. F. Determeand, L. Louveaux, L. Jacques and F. Horlin, "On the exact recovery condition of simultaneous orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 23, no.1, pp. 164-168, Jan. 2016.
- [4] J. F. Determeand, L. Louveaux, L. Jacques and F. Horlin, "Improving the correlation lower bound for simultaneous orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1642-1646, Nov. 2018.
- [5] J. Determe, J. Louveaux, L. Jacques, and F. Horlin, "On the noise robustness of simultaneous orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 864-875, Feb. 2017.
- [6] S. F. Cotter, B. D. Rao, K. Engang, 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.
- [7] J. A. Tropp, "Algorithms for simultaneous sparse approximation-Part II: Convex relaxation," *Signal Process.*, vol. 86, pp. 589-602, 2006.
- [8] E. V. Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2516-2527, May 2010.
- [9] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, pp. 4692-4702, Oct. 2008.
- [10] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 505-519, Jan. 2010.
- [11] M. M. Hyder and K. Mahata, "A robust algorithm for joint-sparse recovery," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1091-1094, Dec. 2009.
- [12] D. Wipf and S. Nagarajan, "Iterative reweighted and methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317-329, Apr. 2010.

- [13] X. Du, L. Cheng, and L. Liu, "A swarm intelligence algorithm for joint sparse recovery," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 611-614, Jun. 2013.
- [14] S. Khanna and C.R. Murthy, "Sparse recovery from multiple measurement vectors using exponentiated gradient updates," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1485-1489, Oct. 2018.
- [15] S. Foucart, "Recovering jointly sparse vectors via hard thresholding pursuit," in *Proc. SAMPTA*, 2011.
- [16] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jing, "Greedy algorithms for joint sparse recovery," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1694-1704, Apr. 2014.
- [17] Z. L. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 5, no. 5, pp. 912-926, Sep. 2011.
- [18] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340-354, Jan. 2013.
- [19] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Multi-task Bayesian compressive sensing exploiting intra-task dependency," *IEEE Trans. Signal Process.*, vol. 22, no. 4, pp. 430-434, Apr. 2015.
- [20] G. Joseph and C. R. Murthy, "A noniterative online Bayesian algorithm for the recovery of temporally correlated sparse vectors," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5510-5525, Oct. 2017.
- [21] J. Shang, Z. Wang, and Q. Huang, "A robust algorithm for joint sparse recovery in presence of impulsive noise," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1166-1170, Aug. 2015.
- [22] 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.
- [23] J. M. Kim, O. K. Lee, and J. C. Ye, "Improving noise robustness in subspace-based joint sparse recovery," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5799-5809, Nov. 2012.
- [24] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613-3641, Jun. 2012.
- [25] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135-1146, Feb. 2012.
- [26] J. Blanchard and M. Davies, "Recovery guarantees for rank aware pursuits," *IEEE Signal Process. Lett.*, vol. 19, no. 7, pp. 427-430, Jul. 2012.
- [27] J. Kim, J. Wang, and B. Shim, "Nearly Optimal Restricted Isometry Condition for Rank Aware Order Recursive Matching Pursuit," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4449-4463, Sep. 2019.
- [28] Z. Wen, B. Hou, and L. Jiao, "Joint sparse recovery with semisupervised MUSIC," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 629-633, May. 2017.
- [29] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230-2249, May 2009.
- [30] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
- [31] S. D. Li, Y. L. Liu, and T. B. Mi, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 69-88, Jul. 2014.