Signal Processing 168 (2020) 107337

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# A modified multiple OLS ($m^2$OLS) algorithm for signal recovery in compressive sensing

Samrat Mukhopadhyay[a], Siddhartha Satpathi[b], Mrityunjoy Chakraborty[a,*]

[a] Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India
[b] Department of Electrical and Computer Engg., University of Illinois at Urbana Champaign, USA

ARTICLE INFO

*Article history:*
Received 13 March 2019
Revised 29 September 2019
Accepted 9 October 2019
Available online 12 October 2019

*Keywords:*
Compressive sensing
mOLS
Restricted isometry property (RIP)

ABSTRACT

Orthogonal least square (OLS) is an important sparse signal recovery algorithm in compressive sensing, which enjoys superior probability of success over other well known recovery algorithms under conditions of correlated measurement matrices. Multiple OLS (mOLS) is a recently proposed improved version of OLS which selects multiple candidates per iteration by generalizing the greedy selection principle used in OLS and enjoys faster convergence than OLS. In this paper, we present a refined version of the mOLS algorithm where at each step of iteration, we first preselect a submatrix of the measurement matrix suitably and then apply the mOLS computations to the chosen submatrix. Since mOLS now works only on a submatrix and not on the overall matrix, computations reduce drastically. Convergence of the algorithm, however, requires to ensure passage of true candidates through the two stages of preselection and mOLS based identification successively. This paper presents convergence conditions for both noisy and noise free signal models. The proposed algorithm enjoys faster convergence properties similar to mOLS, at a much reduced computational complexity.



## 1. Introduction

Signal recovery in compressive sensing (CS) requires evaluation of the sparsest solution to an underdetermined set of equations $\boldsymbol{y} = \boldsymbol{\Phi x}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$ $(m << n)$ is the so-called measurement matrix and $\boldsymbol{y}$ is the $m \times 1$ observation vector. It is usually presumed that the sparsest solution is $K$-sparse, i.e., not more than $K$ elements of $\boldsymbol{x}$ are non-zero, and also that the sparsest solution is unique which can be ensured by maintaining every $2K$ columns of $\boldsymbol{\Phi}$ as linearly independent. There exist a popular class of algorithms in literature called greedy algorithms, which obtain the sparsest $\boldsymbol{x}$ by iteratively constructing the support set of $\boldsymbol{x}$ (i.e., the set of indices of non-zero elements in $\boldsymbol{x}$) via some greedy principles. Orthogonal Matching Pursuit(OMP) [1] is a prominent algorithm in this category, which, at each step of iteration, enlarges a partially constructed support set by appending a column of $\boldsymbol{\Phi}$ that is most strongly correlated with a residual vector, and updates the residual vector by projecting $\boldsymbol{y}$ on the column space of the submatrix of $\boldsymbol{\Phi}$ indexed by the updated support set, and then taking the projection error. Tropp and Gilbert [1] have shown that OMP can recover the original sparse vector from a few measurements with exceedingly high probability when the measurement matrix has i.i.d Gaussian entries. OMP was extended by Wang et al. [2] to the generalized orthogonal matching pursuit (gOMP)where at the indentification stage, multiple columns are selected based on the correlation of the columns of matrix $\boldsymbol{\Phi}$ with the residual vector, which allows gOMP to enjoy faster convergence compared to OMP. It has, however, been shown recently by Soussen et al. [3] that the probability of success in OMP reduces sharply as the correlation between the columns of $\boldsymbol{\Phi}$ increases, and for measurement matrices with correlated entries, another greedy algorithm, namely, the Orthogonal Least Squares (OLS) [4] enjoys much higher probability of recovery of the sparse signal than OMP. OLS is computationally similar to OMP except for a more expensive greedy selection step. Here, at each step of iteration, the partial support set already evaluated is augmented by an index $i$ which minimizes the energy (i.e., the $l_2$ norm) of the resulting residual vector.

An improved version of OLS called multiple OLS (mOLS) has been proposed recently by Wang et al. [5], where unlike OLS, a total of $L$ $(L > 1)$ indices are appended to the existing partial support set by suitably generalizing the greedy principle used in OLS. As $L$ indices are chosen each time, possibility of selection of multiple "true" candidates in each iteration increases and thus, the probability of convergence in much fewer iterations than OLS becomes significantly high.

* Corresponding author.
*E-mail addresses:* sidd.piku@gmail.com (S. Satpathi), mrityun@ece.iitkgp.ernet.in (M. Chakraborty).

https://doi.org/10.1016/j.sigpro.2019.107337
0165-1684/

In this paper, we present a refinement of the mOLS algorithm, named as modified mOLS (m$^2$OLS), where, at each step of iteration, we first *pre-select* a total of, say, $N$ columns of $\mathbf{\Phi}$ by evaluating the correlation between the columns of $\mathbf{\Phi}$ with the current residual vector and choosing the $N$ largest (in magnitude) of them. The steps of mOLS are then applied to this pre-selected set of columns. Here the preselection strategy is identical to the identification strategy of gOMP so that chances of selection of multiple "true" candidates in the pre-selected set is expected to be high. Furthermore, as the mOLS subsequently works on this preselected set of columns and not on the entire matrix $\mathbf{\Phi}$, to determine a subset of $L$ columns ($L<N$), computational costs reduce drastically compared to conventional mOLS. This is also confirmed by our simulation studies. Derivation of conditions of convergence for the proposed algorithm is, however, tricky, as it requires to ensure simultaneous passage of at least one true candidate from $\mathbf{\Phi}$ to the pre-selected set and then, from the pre-selected set to the mOLS determined subset at every iteration step. This paper presents convergence conditions of the proposed algorithm for the cases of both noise free and noisy observations. It also presents the computational steps of an efficient implementation of both mOLS and m$^2$OLS, and brings out the computational superiority of m$^2$OLS over mOLS analytically. Detailed simulation results in support of the claims made are also presented.[1]

## 2. Preliminaries

The following notations have been used throughout the paper : '$H$' in superscript indicates matrix / vector Hermitian conjugate, $\mathcal{H}$ denotes the set of indices $\{1, 2, \cdots, n\}$, $T$ denotes the true support set of $\boldsymbol{x}$, i.e., $T=\{i\in\mathcal{H}|[\boldsymbol{x}]_i\neq 0\}$ and $\boldsymbol{\phi}_i$ denotes the $i$th column of $\mathbf{\Phi}$, $i\in\mathcal{H}$. For any two vectors $\boldsymbol{u},\boldsymbol{v}\in\mathbb{C}^n$, $\langle\boldsymbol{u},\boldsymbol{v}\rangle=\boldsymbol{u}^H\boldsymbol{v}$. All the columns of $\mathbf{\Phi}$ are assumed to have unit $l_2$ norm, i.e., $\|\boldsymbol{\phi}_i\|_2=1$, which is a common assumption in literature [1,5]. For any $S\subseteq\mathcal{H}$, $\boldsymbol{x}_S$ denotes a vector comprising those entries of $\boldsymbol{x}$ that are indexed by numbers belonging to $S$. Similarly, $\mathbf{\Phi}_S$ denotes the submatrix of $\mathbf{\Phi}$ formed with the columns of $\mathbf{\Phi}$ having column numbers given by the index set $S$. If $\mathbf{\Phi}_S$ has full column rank of $|S|$ ($|S|<m$), then the Moore-Penrose pseudo-inverse of $\mathbf{\Phi}_S$ is given by $\mathbf{\Phi}_S^\dagger=(\mathbf{\Phi}_S^H\mathbf{\Phi}_S)^{-1}\mathbf{\Phi}_S^H$. The matrices $\mathbf{P}_S=\mathbf{\Phi}_S\mathbf{\Phi}_S^\dagger$ and $\mathbf{P}_S^\perp=\boldsymbol{I}-\mathbf{P}_S$ respectively denote the orthogonal projection operators associated with $span(\mathbf{\Phi}_S)$ and the orthogonal complement of $span(\mathbf{\Phi}_S)$. For any set $S\subseteq\mathcal{H}$, the matrix $\mathbf{P}_S^\perp\mathbf{\Phi}$ is denoted by $\boldsymbol{A}_S$. For a given sparsity order $K$ and a given matrix $\mathbf{\Phi}$, it can be shown that there exists a real, positive constant $\delta_K$ such that $\mathbf{\Phi}$ satisfies the following "Restricted Isometry Property (RIP)" for all $K$-sparse $\boldsymbol{x}$ :

$$(1-\delta_K)\|\boldsymbol{x}\|_2^2\leq\|\mathbf{\Phi}\boldsymbol{x}\|_2^2\leq(1+\delta_K)\|\boldsymbol{x}\|_2^2.$$

The constant $\delta_K$ is called the restricted isometry constant (RIC) [7] of the matrix $\mathbf{\Phi}$ for order $K$. Clearly, it is the minimum such constant for which the RIP is satisfied. Convergence conditions of recovery algorithms, based on either greedy methods [8–10], or $l_p$ minimization methods ($p\in(0, 1]$) [11–13] in CS, are usually given in terms of upper bounds on the RIC.

## 3. Proposed algorithm

The proposed m$^2$OLS algorithm is described in Table 1. At any $k$th step of iteration ($k\geq 1$), assume a residual signal vector $\boldsymbol{r}^{k-1}$ and a partially constructed support set $T^{k-1}$ have already been computed ($\boldsymbol{r}^0=\boldsymbol{y}$ and $T^0=\emptyset$). In the *preselection* stage, $N$ columns of $\mathbf{\Phi}$ are identified that have largest (in magnitude) correlations with $\boldsymbol{r}^{k-1}$ by picking up the $N$ largest absolute entries of $\mathbf{\Phi}^H\boldsymbol{r}^{k-1}$, and the set $S^k$ containing the corresponding indices is selected. This is followed by the *identification* stage, where $\sum_{i\in\Lambda}\|\mathbf{P}^\perp_{T^{k-1}\cup\{i\}}\boldsymbol{y}\|_2^2$ is evaluated for all subsets $\Lambda$ of $S^k$ having $L$ elements, and selecting the subset $h^k$ for which this is minimum. This is the greedy selection stage, which is carried out in practice [3,5,14] by computing $\frac{|\boldsymbol{\phi}_i^H\boldsymbol{r}^{k-1}|}{\|\mathbf{P}^\perp_{T^{k-1}}\boldsymbol{\phi}_i\|_2}$ for all $i\in S^k$ and selecting the indices corresponding to the $L$ largest of them.[2] The partial support set is then updated to $T^k$ by taking set union of $T^{k-1}$ and $h^k$, and the residual vector is updated to $\boldsymbol{r}^k$ by computing $\mathbf{P}^\perp_{T^k}\boldsymbol{y}$. The sets $T^{k-1}, S^k, h^k$ and the true support $T$ are illustrated in the set diagram in Fig. 1.

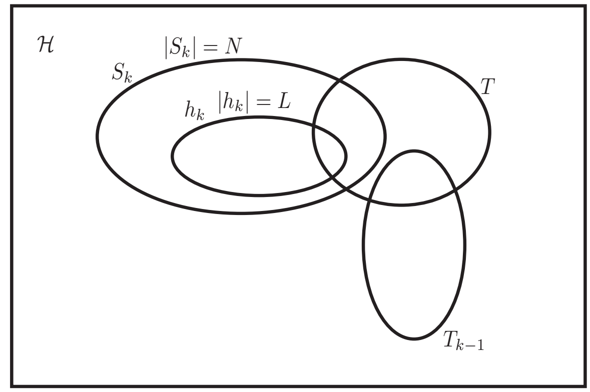



**Fig. 1.** Set diagram illustrating $T^{k-1}, S^k, h^k$ and $T$.

Note that in conventional mOLS algorithm, at a $k$th step of iteration ($k\geq 1$), one has to compute $\frac{|\boldsymbol{\phi}_i^H\boldsymbol{r}^{k-1}|}{\|\mathbf{P}^\perp_{T^{k-1}}\boldsymbol{\phi}_i\|_2}$ for all $i\in\mathcal{H}\setminus T^{k-1}$, involving a total of $n-(k-1)L$ columns, i.e., $\boldsymbol{\phi}_i$'s. In contrast, in the proposed m$^2$OLS algorithm, the above computation is restricted only to the preselected set of $N$ elements, which results in significant reduction of computational complexity.

### 3.1. Lemmas (Existing)

The following lemmas will be useful for the analysis of the proposed algorithm.

**Lemma 3.1** (Monotonicity, Lemma 1 of Dai and Milenkovic [8]). *If a measurement matrix satisfies RIP of orders $K_1$, $K_2$ and $K_1\leq K_2$, then $\delta_{K_1}\leq\delta_{K_2}$.*

**Lemma 3.2** (Consequence of RIP [9]). *For any subset $\Lambda\subseteq\mathcal{H}$, and for any vector $\boldsymbol{u}\in\mathbb{C}^n$,*

$$(1-\delta_{|\Lambda|})\|\boldsymbol{u}_\Lambda\|_2\leq\|\mathbf{\Phi}_\Lambda^H\mathbf{\Phi}_\Lambda\boldsymbol{u}_\Lambda\|_2\leq(1+\delta_{|\Lambda|})\|\boldsymbol{u}_\Lambda\|_2.$$

**Lemma 3.3** (Proposition 3.1 in [9]). *For any $\Lambda\subseteq\mathcal{H}$, and for any vector $\boldsymbol{u}\in\mathbb{C}^m$*

$$\|\mathbf{\Phi}_\Lambda^H\boldsymbol{u}\|_2\leq\sqrt{1+\delta_{|\Lambda|}}\|\boldsymbol{u}\|_2.$$

**Lemma 3.4** (Lemma 1 of Dai and Milenkovic [8]). *If $\boldsymbol{x}\in\mathbb{C}^n$ is a vector with support $S_1$, and $S_1\cap S_2=\emptyset$, then,*

$$\|\mathbf{\Phi}_{S_2}^H\mathbf{\Phi}\boldsymbol{x}\|_2\leq\delta_{|S_1|+|S_2|}\|\boldsymbol{x}\|_2.$$

[1] Some early results of this work were presented at the SPCOM, International Conference on Signal Processing and Communications, July, 2018 [6].

[2] For the first iteration, i.e., for $k=1$, since $T^0=\emptyset$, $\|\mathbf{P}^\perp_{T^{k-1}}\boldsymbol{\phi}_i\|_2=\|\boldsymbol{\phi}_i\|_2=1$, the two steps of the proposed m$^2$OLS algorithm, namely, *preselection* and *identification* amount to finding the $L$ largest absolute entries of $\mathbf{\Phi}^H\boldsymbol{y}$ and forming the set $S^1$ using the corresponding indices, which is identical to the identification step in the first iteration of the mOLS algorithm.

**Table 1**
Proposed m$^2$OLS Algorithm.

**Input:** measurement vector $\boldsymbol{y} \in \mathbb{C}^m$, sensing matrix $\boldsymbol{\Phi} \in \mathbb{C}^{m\times n}$; sparsity level $K$; number of indices preselected $N$; number of indices chosen in identification step, $L (L \le N, L \le K)$, prespecified residual threshold $\epsilon$;
**Initialize:** counter $k = 0$, residue $\boldsymbol{r}^0 = \boldsymbol{y}$, estimated support set, $T^0 = \emptyset$, set selected by preselection step $S^0 = \emptyset$,
**While**($\|\boldsymbol{r}^k\|_2 \ge \epsilon$ and $k < K$)
$k = k + 1$
*Preselect:* $S^k$ is the set containing indices corresponding to the $N$ largest absolute entries of $\boldsymbol{\Phi}^H \boldsymbol{r}^{k-1}$
*Identify:* $h^k = \arg\min_{\Lambda\subset S^k:|\Lambda|=L} \sum_{i\in\Lambda} \|\mathbf{P}^{\perp}_{T^{k-1}\cup\{i\}} \boldsymbol{y}\|_2^2$
*Augment:* $T^k = T^{k-1} \cup h^k$
*Estimate:* $\boldsymbol{x}^k = \arg\min_{\boldsymbol{u}:\boldsymbol{u}\in\mathbb{C}^n,\ supp(\boldsymbol{u})=T^k} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{u}\|_2$
*Update:* $\boldsymbol{r}^k = \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^k$
(Note :Computation of $\boldsymbol{x}^k$ for $1 \le k \le K$ requires every $LK$ columns of $\boldsymbol{\Phi}$ to be linearly independent which is guaranteed by the proposed RIC bound)
**End While**
**Output:** estimated support set $\hat{T} = \arg\max_{\Lambda:|\Lambda|=K} \|\boldsymbol{x}^k_\Lambda\|_2$ and $K$-sparse signal $\hat{\boldsymbol{x}}$ satisfying $\hat{\boldsymbol{x}}_{\hat{T}} = \boldsymbol{\Phi}^{\dagger}_{\hat{T}}\boldsymbol{y}, \hat{\boldsymbol{x}}_{\mathcal{H}\setminus\hat{T}} = \mathbf{0}$

**Lemma 3.5** (Lemma 3 of Satpathi et al.[15]). *If $I_1, I_2 \subset \mathcal{H}$ such that $I_1 \cap I_2 = \emptyset$ and $\delta_{|I_2|} < 1$, then, $\forall \boldsymbol{u} \in \mathbb{C}^n$ such that $supp(\boldsymbol{u}) \subseteq I_2$,*

$$\left(1 - \left(\frac{\delta_{|I_1|+|I_2|}}{1-\delta_{|I_1|+|I_2|}}\right)^2\right)\|\boldsymbol{\Phi u}\|_2^2 \le \|\boldsymbol{A}_{I_1}\boldsymbol{u}\|_2^2 \le \left(1+\delta_{|I_1|+|I_2|}\right)\|\boldsymbol{\Phi u}\|_2^2,$$

*and,*

$$\left(1 - \frac{\delta_{|I_1|+|I_2|}}{1-\delta_{|I_1|+|I_2|}}\right)\|\boldsymbol{u}\|_2^2 \le \|\boldsymbol{A}_{I_1}\boldsymbol{u}\|_2^2 \le \left(1+\delta_{|I_1|+|I_2|}\right)\|\boldsymbol{u}\|_2^2.$$

## 4. Signal recovery using m$^2$OLS algorithm

In this section, we obtain convergence conditions for the proposed m$^2$OLS algorithm. In particular, we derive conditions for selection of at least one correct index at each iteration, which guarantees recovery of a $K$-sparse signal by the m$^2$OLS algorithm in a maximum of $K$ iterations.

Unlike mOLS, proving convergence is, however, trickier in the proposed m$^2$OLS algorithm because of the presence of two selection stages at every iteration, namely, preselection and identification. In order that the proposed algorithm converges in $K$ steps or less, it is essential to ensure that at each step of iteration, at least one true support index $i$ first gets selected in $S^k$ and then, gets passed on from $S^k$ to $h^k$. In the following, we present the convergence conditions for m$^2$OLS where the measurement vector is assumed to be contaminated by an additive noise vector, i.e., $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$. The convergence condition is given by the Theorem 4.1 below. We further use the following performance measures [5] for describing the recovery condition:

- $snr := \frac{\|\boldsymbol{\Phi x}\|_2^2}{\|\boldsymbol{e}\|_2^2}$,
- minimum-to-average-ratio (MAR) [16], $\kappa = \frac{\min_{j\in T}|x_j|}{\|\boldsymbol{x}\|_2/\sqrt{K}}$.

Our main results regarding the signal recovery performance of the m$^2$OLS algorithm is stated in the following theorem.

**Theorem 4.1.** *Given the measurement vector $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$, the m$^2$OLS is guaranteed to collect all the indices of the true support set $T$ within $K$ iterations, if the sensing matrix $\boldsymbol{\Phi}$ satisfies following conditions:*

$$\delta_R < \frac{\sqrt{L}}{\sqrt{K+L}+\sqrt{L}}, \quad (1)$$

$$\sqrt{snr} > \frac{(1+\delta_R)(\sqrt{L}+\sqrt{K})\sqrt{K}}{\kappa\left(\sqrt{L(1-2\delta_R)} - \delta_R\sqrt{K}\right)}, \quad (2)$$

*where $R = LK + N - L + 1$.*

**Proof.** Given in Appendix A. □

Clearly, for the noiseless scenario, Theorem 4.1 implies that the condition (1) is enough to guarantee the perfect recovery of the $K$-sparse vector $\boldsymbol{x}$ in $K$-iterations using m$^2$OLS.

Note that the m$^2$OLS algorithm reduces to the gOMP algorithm when $N = L$. Theorem 4.1 suggests that for $N = L$, the m$^2$OLS algorithm can recover the true support of any $K$-sparse signal from noiseless measurements within $K$ iterations if the sensing matrix satisfies $\delta_{NK+1} < \frac{1}{\sqrt{K/N+1}+1}$, where $N \ge 1$. Recently Wen et al. [17] have established that, with $N \ge 1$, $\delta_{NK+1} < \frac{1}{\sqrt{K/N+1}}$ is a sharp sufficient condition for gOMP to exactly recover $K$-sparse signals from noiseless measurements within $K$ iterations. We see that for large $K/N$ ratio, $\sqrt{K/N+1}+1 \approx \sqrt{K/N+1}$, which shows that the bound provided in Theorem 4.1 is nearly sharp when $N = L$, and $K >> L$. Moreover, in our analysis when $N > 1$, and $L = 1$, the bound in Theorem 4.1 reduces to $\delta_{K+N} < \frac{1}{\sqrt{K+1}+1}$, whereas a recent paper [18] suggests the sufficient condition $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$ for the OLS to recover a $K$-sparse signal perfectly from noiseless measurements within $K$ iterations. Again, the bound suggested in Theorem 4.1 is very close to the one proposed in [18] and quoted above for large $K$. However, in Theorem 4.1, the bound applies on the RIC $\delta_{K+N}$, which is larger than $\delta_{K+1}$. This may make the proposed condition stricter than as presented in [18]. However, since in m$^2$OLS, the operations of mOLS are performed on a smaller preselected set of indices to reduce computational cost, intuitively it is expected that the sensing matrix must satisfy some stricter RIP condition in order to yield recovery performance competitive to that of OLS.

The proposed proof uses mathematical induction, where we first find out a sufficient condition for success by m$^2$OLS in the first iteration, and then assuming that m$^2$OLS is successful in each of the previous $k$ ($1 \le k \le K-1$) iterations, find out conditions sufficient for success at the $(k+1)$th iteration. Success of m$^2$OLS at any iteration, however, requires ensuring that first at least one true index is captured in the preselection step, which is identical to the gOMP identification step, and then further ensuring that at least one of these captured true indices is recaptured by the identification step, which is identical to the mOLS identification step (with the selection set restricted to $S^k$). Thus, a sufficient condition for success of the m$^2$OLS is obtained by simultaneous satisfaction of the sufficient conditions for both these steps. Finally, the sufficient conditions for any general iteration $k (2 \le k \le K)$, and the condition for iteration 1 are combined to obtain the final condition for successful recovery of support within $K$ iterations.

The proposed proof follows in part from the framework of the proof of Theorem 3 in Wang et al. [5], and also uses certain steps from the proof of Theorem 1 of the paper of Li et al. [19], more specifically [19], Eqs. (25),(26) and Eq. (9) of Satpathi et al. [15]. However, in our analysis of the identification step, we have used Lemma 3.5 which gives rise to a bound on $\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2^2$. This makes combination of the guarantees of the two steps of the m$^2$OLS algorithm into one quite tractable, something that seems to be extremely difficult if one follows [5, Eq. (E.7)] instead of the above bound.

## 5. Comparative analysis of computational complexities of mOLS and m$^2$OLS

By restricting the steps of mOLS to a pre-selected subset of columns of $\boldsymbol{\Phi}$, the proposed m$^2$OLS algorithm achieves considerable computational simplicity over mOLS. In this section, we analyze the computational steps involved in both mOLS and m$^2$OLS at the $(k+1)^{\text{th}}$ iteration (i.e., assuming that $k$ iterations of either algorithm have been completed), and compare their computational costs in terms of number of floating point operations (flops) required.

### 5.1. Computational steps of mOLS (in iteration $k+1$)

**Step 1** (**Absolute correlation calculation**) : Here $|\langle\boldsymbol{\phi}_i,\ \boldsymbol{r}^k\rangle|$ is calculated $\forall i \in \mathcal{H}\setminus T^k$, where the vector $\boldsymbol{r}^k$ was precomputed at the end of the $k^{\text{th}}$ step. We initialize $\boldsymbol{r}^0 = \boldsymbol{y}$.

**Step 2** (**Identification**) : In this step, mOLS first calculates the ratios $\frac{|\langle\boldsymbol{\phi}_i,\boldsymbol{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2}$, $\forall i \in \mathcal{H}\setminus T^k$. Since $\forall i \in \mathcal{H}\setminus T^k$, the numerator was calculated in Step 1, only the denominator needs to be calculated. However, as will be discussed later, at the end of each $k^{\text{th}}$ step, for each $i \in \mathcal{H}\setminus T^k$, the norm $\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2$ is calculated and stored, which provides the denominators in the above ratios. This means, the above computation requires simply a division operation per ratio and a total of $(n-Lk)$ divisions. This step is followed by finding the $L$ largest of the above ratios, and appending the corresponding columns to the previously estimated subset of columns, $\boldsymbol{\Phi}_{T^k}$, thereby generating $\boldsymbol{\Phi}_{T^{k+1}}$ (for $k=0$, $T^k=\emptyset$ and thus, $\boldsymbol{\Phi}_{T^k}=\emptyset$).

**Step 3** (**Modified Gram Schmidt**) : This step finds an orthonormal basis for $\mathtt{span}\left(\boldsymbol{\Phi}_{T^{k+1}}\right)$. Assuming that an orthonormal basis $\{\boldsymbol{u}_1,\ \cdots,\ \boldsymbol{u}_{|T^k|}\}$ for $\mathtt{span}\left(\boldsymbol{\Phi}_{T^k}\right)$ has already been computed at the $k^{\text{th}}$ step, an efficient way to realize this will be to employ the well known Modified Gram Schmidt (MGS) procedure [20], which first computes $\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i,\ i \in h^{k+1}$ using the above precomputed orthonormal basis and then, orthonormalizes them recursively, generating the orthonormal set $\{\boldsymbol{u}_{|T|^k+1}, \cdots,\ \boldsymbol{u}_{|T^{k+1}|}\}$.

**Step 4** (**Precomputation of orthogonal projection error norm**) : At the $(k+1)^{\text{th}}$ step, after MGS is used to construct an orthonormal basis for $\mathtt{span}\left(\boldsymbol{\Phi}_{T^{k+1}}\right)$, the norms $\|\mathbf{P}^{\perp}_{T^{k+1}}\boldsymbol{\phi}_i\|_2,\ i \in \mathcal{H}\setminus T^{k+1}$, are computed using the following recursive relation, for use in the identification step of $(k+2)^{\text{th}}$ step:

$$\|\mathbf{P}^{\perp}_{T^{k+1}}\boldsymbol{\phi}_i\|_2^2 = \|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2^2 - \sum_{j=|T^k|+1}^{|T^{k+1}|}\left|\langle\boldsymbol{\phi}_i, \boldsymbol{u_j}\rangle\right|^2. \tag{3}$$

**Step 5** (**Calculation of $\boldsymbol{r}^{k+1}$**) : Finally mOLS calculates the residual vector $\boldsymbol{r}^{k+1}$ as follows:

$$\boldsymbol{r}^{k+1} = \boldsymbol{r}^k - \sum_{j=|T^k|+1}^{|T^{k+1}|}\langle\boldsymbol{y}, \boldsymbol{u}_j\rangle\boldsymbol{u}_j. \tag{4}$$

**Table 2**
Computational costs (in flops) of the identification step of mOLS and m$^2$OLS in $(k+1)$th iteration.

| | mOLS | m$^2$OLS |
|---|---|---|
| Identification step | $(n-Lk)(Lr+2)-L(Lr+1)$ | $N(Lkr+2)$ |

### 5.2. Computational steps of m$^2$OLS (in iteration $k+1$)

**Step 1** (**Preselection**): In this step, similar to mOLS, the absolute correlations $|\langle\boldsymbol{\phi}_i,\ \boldsymbol{r}^k\rangle|$ are calculated using the vector $\boldsymbol{r}^k$ which is precomputed at the end of the $k^{\text{th}}$ step. Then the indices corresponding to the $N$ largest absolute correlations are selected to form the set $S^{k+1}$.

**Step 2** (**Identification**): The identification step calculates the ratios $\frac{|\langle\boldsymbol{\phi}_i,\boldsymbol{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2}$, $\forall i \in S^{k+1}$, for which the numerators are already known from Step 1 and the denominators are calculated as per the following:

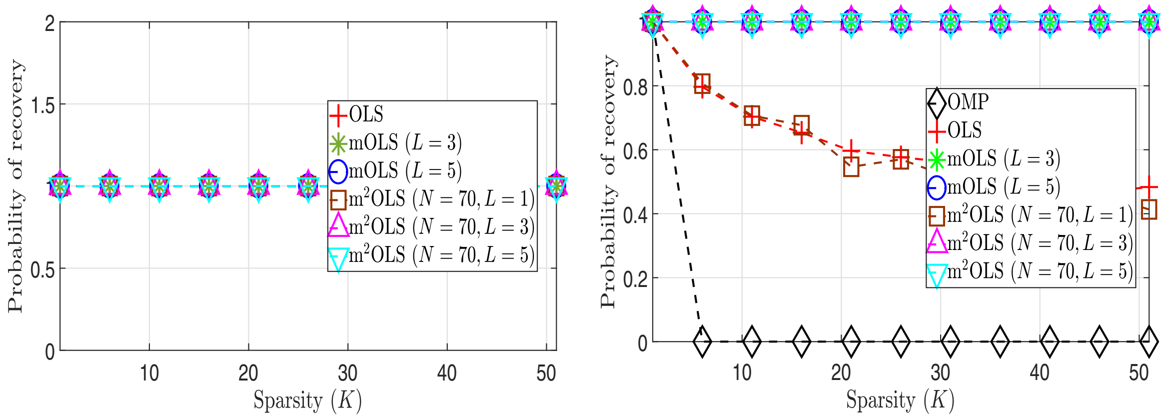
$$\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2^2 \ = \ \|\boldsymbol{\phi}_i\|_2^2 - \sum_{j=1}^{|T^k|}\left|\langle\boldsymbol{\phi}_i, \boldsymbol{u_j}\rangle\right|^2, \tag{5}$$

where, as in mOLS, $\{\boldsymbol{u}_1, \cdots,\ \boldsymbol{u}_{|T|^k}\}$ is the orthonormal basis formed for $\mathtt{span}\left(\boldsymbol{\Phi}_{T^k}\right)$ using MGS at step $k$. For the first step $k=0$, $T^k=\emptyset$, and thus $\|\mathbf{P}^{\perp}_{T^0}\boldsymbol{\phi}_i\|_2 = \|\boldsymbol{\phi}_i\|_2,\ i \in \mathcal{H}$. It is assumed that the norms $\|\boldsymbol{\phi}_i\|_2$ are all precomputed (taken to be unity in this paper). This computation is followed by $N$ divisions as required to form the above ratios. Following this, the indices corresponding to the largest $L$ of the $N$ ratios are determined and the corresponding columns are appended to the previously estimated set of columns $\boldsymbol{\Phi}_{T^k}$ to obtain $\boldsymbol{\Phi}_{T^{k+1}}$.

**Step 3** (**Modified Gram Schmidt**): This step is identical to the MGS step in mOLS, which generates an orthonormal basis for $\mathtt{span}\left(\boldsymbol{\Phi}_{T^{k+1}}\right)$.

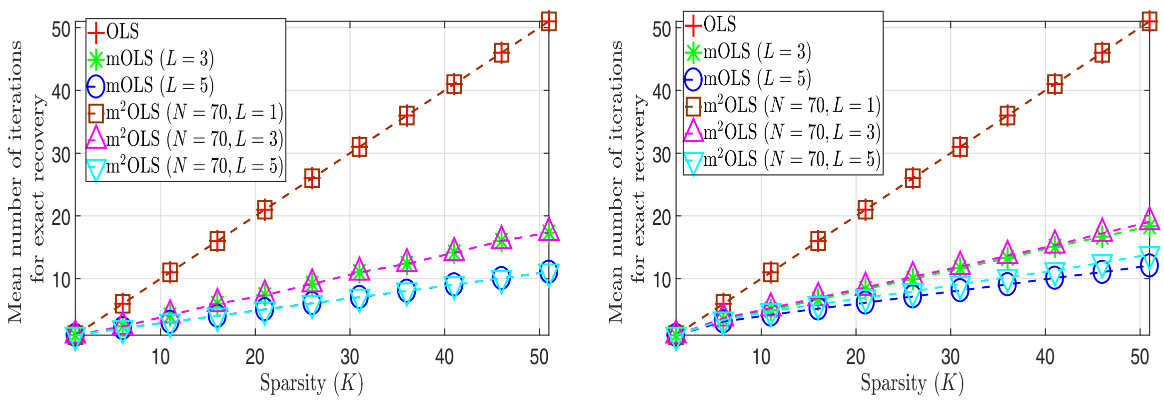
**Step 4** (**Computation of $\boldsymbol{r}^{k+1}$**): As in mOLS, the residual $\boldsymbol{r}^{k+1}$ is updated using Eq. (4).

**Comparison between computational complexities of mOLS and m$^2$OLS:** While certain operations like MGS, computation of absolute correlation and the residual $\boldsymbol{r}^k$ are same in both mOLS and m$^2$OLS, the major computational difference between them lies in the following : at the end of every $(k+1)^{\text{th}}$ step, the mOLS computes $\|\mathbf{P}^{\perp}_{T^{k+1}}\boldsymbol{\phi}_i\|_2^2\ \forall i \in \mathcal{H}\setminus T^{k+1}$ using recursion of the form (3). If computation of $|\langle\boldsymbol{\phi}_i,\ \boldsymbol{u_j}\rangle|^2$ has a complexity of $r$ flops, this requires a total of $(n-L(k+1))(Lr+1)$ flops. Additionally, mOLS requires $(n-Lk)$ divisions to compute the ratios $\frac{|\langle\boldsymbol{\phi}_i,\boldsymbol{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2}$, $\forall i \in \mathcal{H}\setminus T^k$. The m$^2$OLS, on the other hand, calculates $|\langle\boldsymbol{\phi}_i,\ \boldsymbol{u_j}\rangle|^2$ only for $i \in S^{k+1}$, following (5), involving at the most just $N$ and not $(n-Lk)$ columns. The summation on the RHS (Right Hand Side) of (5), however, has $Lk$ terms, meaning this step requires a total of $N(Lkr+2)$ flops (including the $N$ divisions to compute $\frac{|\langle\boldsymbol{\phi}_i,\boldsymbol{r}^k\rangle|}{\|\mathbf{P}^{\perp}_{T^k}\boldsymbol{\phi}_i\|_2}$, $\forall i \in S^{k+1}$). Table 2 compares the computational costs of the $(k+1)^{\text{th}}$ iteration of mOLS and m$^2$OLS in terms of the computational costs of the identification steps (the computational costs of the other steps are at par for both the algorithms). Clearly, mOLS will require more computations than m$^2$OLS as long as $1 < \frac{(n-Lk)(Lr+2)-L(Lr+1)}{N(Lkr+2)} \approx \frac{n-Lk-L}{Nk}$, or, equivalently, for $k < \frac{n-L}{N+L}$. Thus, for large $n$ and / or small $K$, as $k \le K$, the mOLS will have significantly higher computational overhead as compared to m$^2$OLS at each iteration $k$ and the difference in cumulative computational cost over all iterations put together will be huge. Even when the sparsity $K$ is larger $(2K < m)$, the actual number of iterations, say, $J$ required



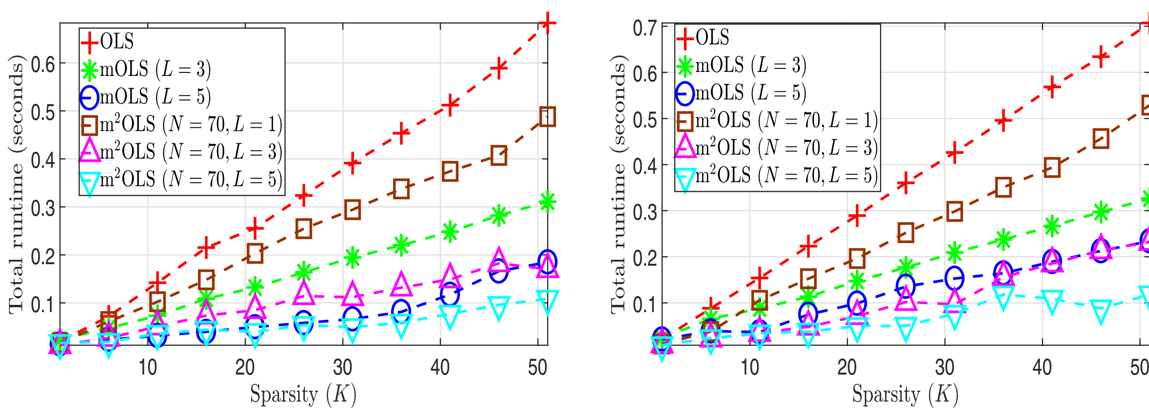

(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 2.** Recovery probability vs sparsity.




(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 3.** No. of iterations for exact recovery vs sparsity.



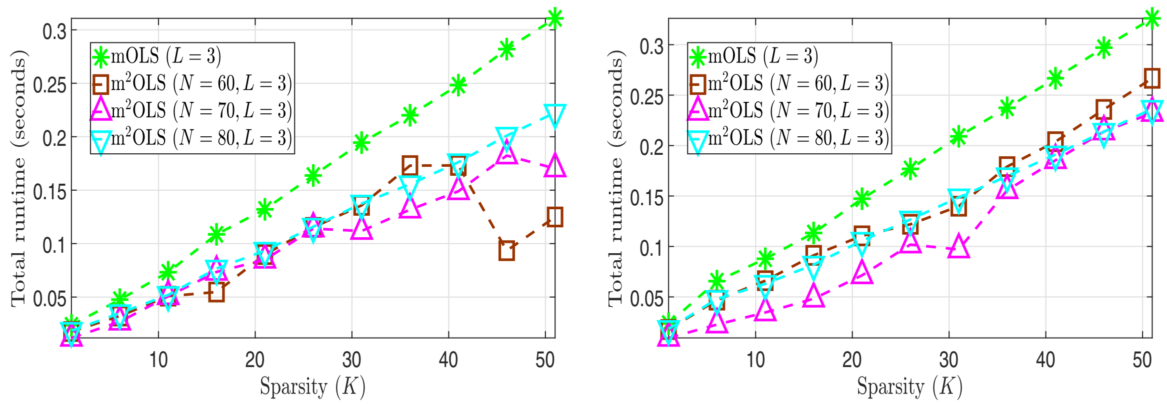

(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 4.** Mean runtime vs sparsity (mOLS and m$^2$OLS) for different $L$.

for convergence by both mOLS and m$^2$OLS is usually much less than $K$ and the above difference continues. In case of large $K$ and $J$ close to $K$, the mOLS will require more computations than m$^2$OLS for $k$ upto certain value, beyond which m$^2$OLS will start having more computations and thus, the difference in cumulative computational cost between mOLS and m$^2$OLS will start reducing with $k$. This means, for large $K$, we have a reasonably large range of $J$ for which the overall computational cost of mOLS remains substantially higher than that of m$^2$OLS. The above comparative assessment of mOLS and m$^2$OLS in terms of computations required is also validated by simulation studies as presented in the next section.

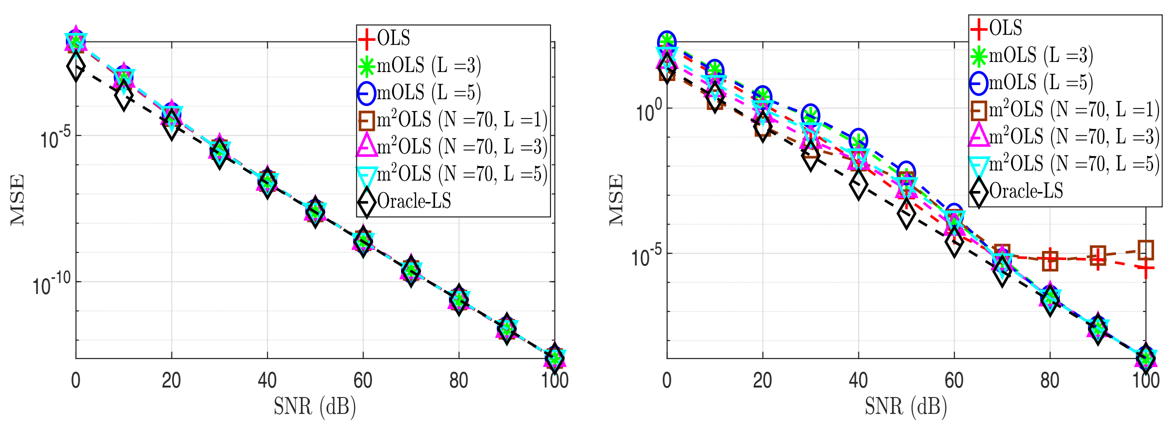
## 6. Simulation results

For simulation, we constructed measurement matrices with correlated entries, as used by Soussen et al. [3]. For this, first a matrix $\boldsymbol{A}$ is formed such that $a_{ij} = [\boldsymbol{A}]_{ij}$ is given by $a_{ij} = n_{ij} + t_j$ where




(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 5.** Mean runtime vs sparsity (mOLS and m$^2$OLS) for different $N$.

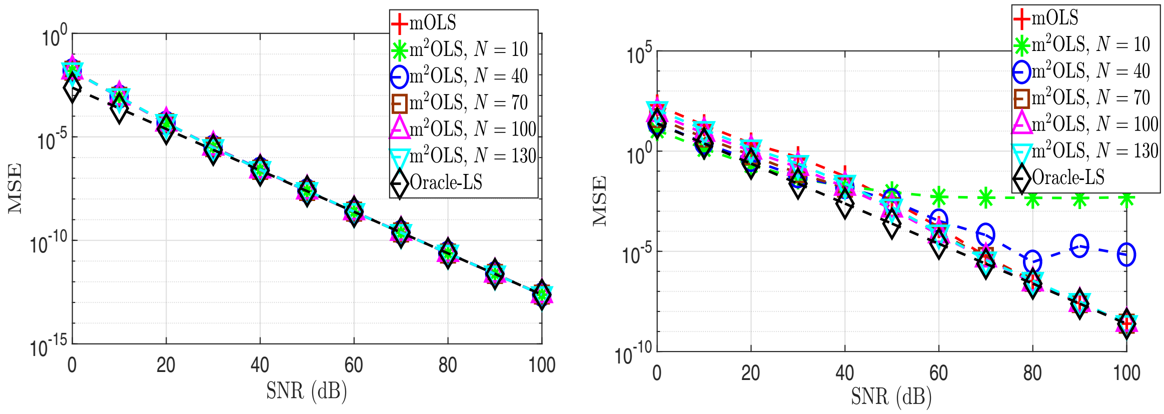



(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 6.** Mean Square Error (MSE) vs SNR ($K = 30$).

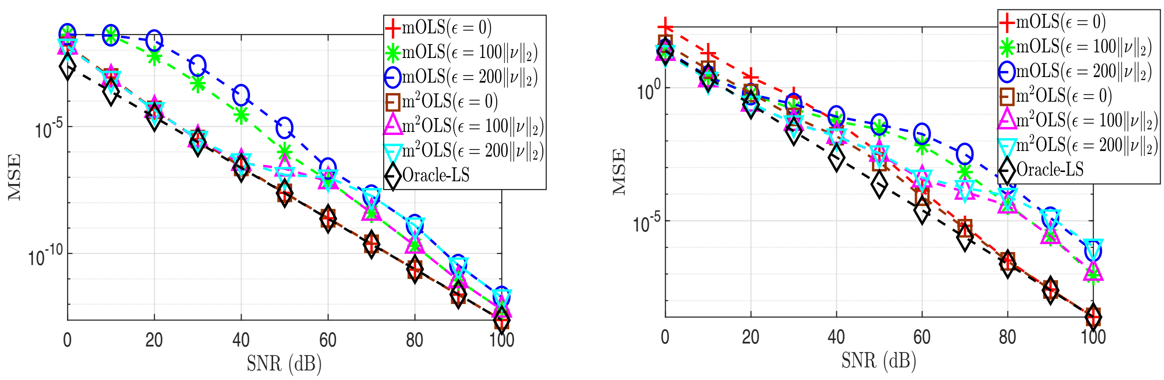
$n_{ij} \sim \mathcal{N}(0, 1/m)$ i.i.d. $\forall i, j$, $t_j \sim \mathcal{U}[0, \tau]\forall j$, and $\{n_{ij}\}$ is statistically independent of $\{t_k\}$, $\forall i, j, k$. The measurement matrix $\boldsymbol{\Phi}$ is then constructed from $\boldsymbol{A}$ as $\phi_{ij} = a_{ij}/\|\boldsymbol{a}_j\|_2$, where $\phi_{ij} = [\boldsymbol{\Phi}]_{ij}$ and $\boldsymbol{a}_i$ denotes the $i$th column of $\boldsymbol{A}$. Note that in the construction process for $\boldsymbol{\Phi}$, the random variables $n_{ij}$ play the role of additive i.i.d. noise process, added to the elements of a rank 1 matrix, with columns $\{t_i\boldsymbol{1}\}_{i=1}^n$, where $\boldsymbol{1}$ denotes a $m \times 1$ vector with all entries equal to one. If the value of $\tau$ becomes large as compared to the variance $1/m$ of $n_{ij}$, then the matrix $\boldsymbol{\Phi}$ resembles a rank 1 matrix with normalized columns. For all the simulations, the values of $m, n$ were fixed at 500, 800 respectively while the sparsity $K$ was varied. The nonzero elements of $\boldsymbol{x}$ were drawn randomly from i.i.d Gaussian distribution and $\tau$ was chosen to have values either 0 or 8. Note that higher the value of $\tau$, more will be the correlation (taken as the absolute value of the inner product, which is a measure of *coherence*) between the columns of $\boldsymbol{\Phi}$. Thus, $\tau = 0$ produces a matrix with uncorrelated columns while $\tau = 8$ produces a matrix with reasonably correlated columns. Furthermore, different values for the window sizes for the preselection stage ($N$) of m$^2$OLS and the identification stages ($L$) of both mOLS and m$^2$OLS were used for the simulation. Particularly, the values $\{60, 70, 80\}$ were used for $N$, and $\{1, 3, 5\}$ were used for $L$.

For each value of $K$, the mOLS and m$^2$OLS were run till they converge or upto the $K$th step of iteration, whichever is earlier, and the experiment was conducted 500 times. To evaluate the performance of the algorithms, three performance metrics were considered, namely, recovery probability, mean number of iterations ($\leq K$) for convergence and mean runtime. Of these the recovery probability was obtained by counting the number of times out of the 500 trials each algorithm converges, while for the other two, averaging was done over the 500 trials [1]. In the **first** simulation exercise, the recovery probabilities are plotted against $K$. The plots, shown in Figs. 2(a) and (b) for $\tau = 0$ and $\tau = 8$ respectively, suggest that even for highly correlated dictionaries ($\tau = 8$), the probability of recovery exhibited by m$^2$OLS is identical to mOLS over the entire sparsity range considered. The **second** simulation exercise evaluates the average no. of iterations required by the two algorithms for exact recovery for each value of $K$. The corresponding results, shown in Figs. 3(a) and (b) for $\tau = 0$ and $\tau = 8$ respectively, reveal that for the uncorrelated case ($\tau = 0$), both mOLS and m$^2$OLS algorithms require the same average number of iterations for successful recovery, and it is only under $\tau = 8$ that as $K$ increases beyond a point, there is a marginal increase in the average number of iterations in m$^2$OLS over mOLS. In our **third** exercise, we evaluated the average of total runtime for the algorithms, against $K$. The corresponding results are shown in Fig. 4(a) and (b), for $\tau = 0$ and $\tau = 8$ respectively. The plots demonstrate the superiority of the proposed m$^2$OLS algorithm over mOLS as the former is seen to require much less running time than mOLS for both values of $\tau$. The results also illustrate that the runtime of m$^2$OLS decreases with increasing $L$ while maintaining lesser runtime than mOLS. This validates the conjectures made at the end of Section 5 regarding the reduced computational overhead of m$^2$OLS over a large range of sparsity values. We also plot in Fig. 5 runtimes for mOLS and m$^2$OLS against $K$, keeping $L = 3$ and taking $N$ as a parameter ($N = 60, 70, 80$). The results here demonstrate that




(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 7.** Mean Square Error (MSE) vs SNR ($K = 30,\ L = 3$) for different values of $N$.




(a) $\tau = 0$ (b) $\tau = 8$

**Fig. 8.** Mean Square Error (MSE) vs SNR ($K = 30,\ N = 70,\ L = 3$) for different values of $\epsilon$.

the runtime of m$^2$OLS can be controlled by changing the preselection step size $N$, which is intuitively expected. In our **fourth** exercise, we ran the mOLS and m$^2$OLS algorithms with measurements corrupted by additive Gaussian noise with varying SNR (as defined in Section 4). The mean square error (MSE) is computed as $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2$ where $\boldsymbol{x}$ is the original vector and $\hat{\boldsymbol{x}}$ is its estimate as produced by the algorithm. The MSE is plotted against SNR keeping $N = 70$ and taking $L$ as a parameter ($L = 1, 3, 5$). As a benchmark, the MSE of the Oracle estimator is plotted, where the Oracle estimator computes the least squares estimate of the optimal vector in presence of noise. The plots shown in Figs. 6(a) and (b) demonstrate that for uncorrelated dictionaries ($\tau = 0$), the two algorithms exhibit almost the same performance, while for correlated dictionaries ($\tau = 8$), m$^2$OLS has actually a slightly better MSE performance than mOLS for different values of $L$. We also investigate the effects of the preselection parameter $N$ on the performance of m$^2$OLS by varying $N$ from a small value to a relatively high value, and plotting the MSE against the SNR with $N$ as a parameter. The corresponding plots are shown in Figs. 7(a) and (b), for $\tau = 0$, and $\tau = 8$ respectively, where $N$ is chosen from the set {10, 40, 70, 100, 130} and $L$ is fixed at 3. From the two plots, we observe that there is no significant degradation in the MSE performance of the m$^2$OLS for small $N$ when the columns are uncorrelated (i.e. $\tau = 0$). However, when the columns are correlated ($\tau = 8$), from Fig. 7(b), it is seen that the MSE performance becomes very poor for $N = 10$ (even for large SNR), and the performance gradually improves as larger values of $N$ are chosen. For $N \geq 70$, we see that the MSE performance of m$^2$OLS is almost equivalent to that of the Oracle-LS. Of course, this value of $N$ is obtained when $L = 3$, and larger $N$ will be required when larger $L$ is used. This observation, along with the plots in Fig. 5 suggest that, given $L$, the value of $N$ should be carefully chosen, as a small value of $N$ will result in low complexity but poor recovery performance, while larger $N$ might give good recovery performance at the cost of higher computational cost. Numerically we have seen that choosing $N$ of the order of $n/10$, and then choosing $L$ of the order of $N/20$ often yield satisfactory overall performance.

We also examined numerically the sensitivity of the MSE performance of the mOLS and m$^2$OLS algorithms on the parameter $\epsilon$ on which termination of mOLS and m$^2$OLS depend. We fixed $N = 70,\ L = 3$, and took the values of $\epsilon$ as $\epsilon = \eta\|\boldsymbol{\nu}\|_2$, where $\boldsymbol{\nu}$ is the measurement noise vector, and $\eta \in \{0,\ 100,\ 200\}$. The corresponding MSE vs SNR plots with $\epsilon$ as parameter are shown in Figs. 8(a) and (b) for the uncorrelated ($\tau = 0$) and correlated ($\tau = 8$) dictionaries, respectively. It is seen from these figures that the MSE of m$^2$OLS is smaller than that of mOLS. Furthermore, early termination (high $\epsilon$) increases the MSE of m$^2$OLS in the high SNR region when $\tau = 0$, whereas the MSE of mOLS increases in an even lower SNR region. Moreover, in the correlated dictionary, early termination reduces MSE of both mOLS and m$^2$OLS in low SNR region, and increases it for high SNR region.

## 7. Conclusion

In this paper we have proposed a greedy algorithm for sparse signal recovery which preselects a few ($N$) possibly “good” indices according to correlation of the respective columns of the measurement matrix with a residual vector, and then uses an mOLS step to identify a subset of these indices (of size $L$) to be included in the estimated support set. We have carried out a theoretical analysis of the algorithm using RIP and have shown that for the noiseless signal model, if the sensing matrix satisfies the RIP condition $\delta_{LK+N-L+1} < \frac{\sqrt{L}}{\sqrt{L}+\sqrt{L+K}}$, then the m$^2$OLS algorithm is guaranteed to exactly recover a $K$ sparse unknown vector, satisfying the measurement model, within $K$ steps. We further extended our analysis to a noisy measurement setup and worked out bounds on the measurement SNR analytically, which guarantees exact recovery of the support of the unknown sparse vector within $K$ iterations. We have also presented the computational steps of both mOLS and m$^2$OLS in a MGS based efficient implementation and carried out a comparative analysis of their computational complexities, which showed that m$^2$OLS enjoys significantly reduced computational overhead compared to mOLS, especially for large $n$ and / or small $K$. Finally, through numerical simulations, we have verified that the introduction of the preselection step indeed leads to less computation time, and that the recovery performance of m$^2$OLS in terms of recovery probability and number of iterations for success is highly competitive with mOLS for a wide range of parameter values.

## Declaration of Competing Interest

The authors declare that they do not have any financial or non-financial conflict of interests.

## Acknowledgment

This research was in part supported by a research grant from the Science and Engineering Research Board, Govt. of India (grant number: EMR/2016/005290) and a chair professorship grant from the Indian National Academy of Engineering.

## Appendix A. Proof of Theorem 4.1

*A.0.1. Success at the first iteration*

Since the first iteration computations of both m$^2$OLS and mOLS are the same, namely that of finding the $L$ largest absolute entries of $\boldsymbol{\Phi}^H\boldsymbol{y}$ and forming the set $S^1$ using the corresponding indices, selection of at least one true index in $T^1$, or, equivalently, satisfaction of $T^1 \cap T \neq \emptyset$ will be guaranteed under the following sufficient condition, which can be developed in exactly the same manner as the inequality (23) in [5] is developed:

$$\delta_{L+K}\|\boldsymbol{x}\|_2 + \sqrt{1+\delta_L}\|\boldsymbol{e}\|_2 < \sqrt{\frac{L}{K}}\left[(1-\delta_K)\|\boldsymbol{x}\|_2 - \sqrt{1+\delta_K}\|\boldsymbol{e}\|_2\right]. \tag{A.1}$$

*A.0.2. Success at $(k+1)$th iteration*

For finding the condition for success in a general $(k+1)$th iteration, we assume that in each of the previous $k (k < K)$ iterations, at least one correct index was selected, meaning, if $|T \cap T^k| = c_k$, then $c_k \geq k$. Let $c_k < K$. Also define $m_k := |S^k \cap T \setminus T^{k-1}|,\ k \geq 1$, meaning, $m_i \geq 1, 1 \leq i \leq k$. For success of the $(k+1)^{th}$ iteration, we require $S^{k+1} \cap T \setminus T^k \neq \emptyset$, and $h^{k+1} \cap T \setminus T^k \neq \emptyset$ simultaneously, as this will ensure selection of at least one new true index at the $(k+1)$th iteration.

**Condition to ensure** $S^{k+1} \cap T \setminus T^k \neq \emptyset$: For this, we follow the approach of Wang et al. [2], and Li et al. [19] and we define the following:

- $W^{k+1} := \arg\max_{S \subset \mathcal{H}\setminus(T\setminus T^k):\ |S|=N} \|\boldsymbol{\Phi}_S^H \boldsymbol{r}^k\|_2$.
- $\alpha_N^k := \min_{i \in W^{k+1}} \left|\langle \boldsymbol{\phi}_i, \boldsymbol{r}^k\rangle\right|$.
- $\beta_1^k := \max_{i \in T\setminus T^k} \left|\langle \boldsymbol{\phi}_i, \boldsymbol{r}^k\rangle\right|$.

Clearly, $S^{k+1} \cap T \setminus T^k \neq \emptyset$, if $\beta_1^k > \alpha_N^k$. In both $\beta_1^k$ and $\alpha_N^k$, we use a form of $\boldsymbol{r}^k$ presented by Satpathi et al. in [15] and given by $\boldsymbol{r}^k = \boldsymbol{\Phi}_{T\cup T^k}\boldsymbol{x}'_{T\cup T^k} + \mathbf{P}_{T^k}^{\perp}\boldsymbol{e}$, where $\boldsymbol{x}'_{T\cup T^k} = \begin{bmatrix} \boldsymbol{x}_{T\setminus T^k} \\ -\boldsymbol{u}_{T^k}\end{bmatrix}$ and $\boldsymbol{u}_{T^k}$ is some vector belonging to $\mathbb{C}^{Lk}$ (obtained by first writing $\boldsymbol{r}^k = \mathbf{P}_{T^k}^{\perp}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \mathbf{P}_{T^k}^{\perp}\boldsymbol{e} = \boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} - \mathbf{P}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k} + \mathbf{P}_{T^k}^{\perp}\boldsymbol{e}$, and then expressing $\mathbf{P}_{T^k}\boldsymbol{\Phi}_{T\setminus T^k}\boldsymbol{x}_{T\setminus T^k}$ as a linear combination of the columns of $\boldsymbol{\Phi}_{T^k}$, i.e., as $\boldsymbol{\Phi}_{T^k}\boldsymbol{u}_{T^k}$ for some $\boldsymbol{u}_{T^k} \in \mathbb{C}^{Lk}$).

Using the above form of $\boldsymbol{r}^k$, followed by an use of triangle inequality, we obtain

$$\alpha_N^k \leq \frac{\|\boldsymbol{\Phi}_{W^{k+1}}^t \boldsymbol{r}^k\|_2}{\sqrt{N}} = \frac{\|\boldsymbol{\Phi}_{W^{k+1}\setminus T^k}^t \boldsymbol{r}^k\|_2}{\sqrt{N}} \leq \frac{\left(\|\boldsymbol{\Phi}_{W^{k+1}\setminus T^k}^H \boldsymbol{\Phi}_{T\cup T^k}\boldsymbol{x}'_{T\cup T^k}\|_2 + \|\boldsymbol{\Phi}_{W^{k+1}\setminus T^k}^H \mathbf{P}_{T^k}^{\perp}\boldsymbol{e}\|_2\right)}{\sqrt{N}}. \tag{A.2}$$

An upper bound on a similarly defined quantity $\alpha_N^k$ was derived by Li et al. [19, Eq. (26)] using Lemmas 3.4 and 3.3 given in Section 2. Using similar steps as well as the fact that $\|\mathbf{P}_{T^k}^{\perp}\boldsymbol{e}\|_2 \leq \|\boldsymbol{e}\|_2$, we obtain,

$$\alpha_N^k \leq \frac{1}{\sqrt{N}}\left(\delta_{N+Lk+K-c_k}\|\boldsymbol{x}'_{T\cup T^k}\|_2 + \sqrt{1+\delta_N}\|\boldsymbol{e}\|_2\right). \tag{A.3}$$

On the other hand,

$$\beta_1^k \geq \frac{1}{\sqrt{K-c_k}}\|\boldsymbol{\Phi}_{T\setminus T^k}^H \boldsymbol{r}^k\|_2 = \frac{1}{\sqrt{K-c_k}}\|\boldsymbol{\Phi}_{T\cup T^k}^H\left(\boldsymbol{\Phi}_{T\cup T^k}\boldsymbol{x}'_{T\cup T^k} + \mathbf{P}_{T^k}^{\perp}\boldsymbol{e}\right)\|_2.$$

Again, a lower bound on a similarly defined quantity $\beta_1^k$ was derived by Li et al. [19, Eq. (25)] using Lemmas 3.2, and 3.3 given in Section 2. Using similar steps, followed by an use of reverse triangle inequality, we obtain

$$\beta_1^k \geq \frac{(1-\delta_{Lk+K-c_k})\|\boldsymbol{x}'_{T\cup T^k}\|_2 - \sqrt{1+\delta_{Lk+K-c_k}}\|\boldsymbol{e}\|_2}{\sqrt{K-c_k}}. \tag{A.4}$$

Hence, $S^{k+1} \cap T \setminus T^k \neq \emptyset$ is ensured if

$$\frac{(1-\delta_{Lk+K-c_k})\|\boldsymbol{x}'_{T\cup T^k}\|_2 - \sqrt{1+\delta_{Lk+K-c_k}}\|\boldsymbol{e}\|_2}{\sqrt{K-c_k}} > \frac{1}{\sqrt{N}}\left(\delta_{N+Lk+K-c_k}\|\boldsymbol{x}'_{T\cup T^k}\|_2 + \sqrt{1+\delta_N}\|\boldsymbol{e}\|_2\right). \tag{A.5}$$

Furthermore, from the monotonicity of RIC, we have $\delta_{Lk+K-c_k} < \delta_{N+Lk+K-c_k}$, and $1 \leq k \leq c_k$ and $k \leq K-1$, imply, $Lk+K-c_k \leq (L-1)k+K \leq (L-1)(K-1)+K = LK-L+1$. Finally, using the monotonicity of the RIC again, a sufficient condition for ensuring $S^{k+1} \cap T \setminus T^k \neq \emptyset$ is given by

$$\frac{\left((1-\delta_R)\|\boldsymbol{x}'_{T\cup T^k}\|_2 - \sqrt{1+\delta_R}\|\boldsymbol{e}\|_2\right)}{\sqrt{K-c_k}} > \frac{\left(\delta_R\|\boldsymbol{x}'_{T\cup T^k}\|_2 + \sqrt{1+\delta_R}\|\boldsymbol{e}\|_2\right)}{\sqrt{N}}, \tag{A.6}$$

where $R = N + LK - L + 1$.

**Condition to ensure** $h^{k+1} \cap T \setminus T^k \neq \emptyset$ : We first consider the set $S^{k+1} \setminus (T \setminus T^k)$. If $|S^{k+1} \setminus (T \setminus T^k)| < L$, then the condition $h^{k+1} \cap T \setminus T^k \neq \emptyset$ is satisfied trivially. Therefore, we consider cases where $|S^{k+1} \setminus (T \setminus T^k)| \geq L$. To this end, we define $a_i = \frac{|\langle \boldsymbol{\phi}_i, \boldsymbol{r}^k\rangle|}{\|\mathbf{P}_{T^k}^{\perp}\boldsymbol{\phi}_i\|_2}$ if $i \in S^{k+1} \setminus \tilde{T}^k$, and $a_i = 0$ for $i \in S^{k+1} \cap \tilde{T}^K$, where $\tilde{T}^k = \{i \in H | \boldsymbol{\phi}_i \in span(\boldsymbol{\Phi}_{T^k})\}$, meaning, $T^k \subseteq \tilde{T}^K$ and for $i \in \tilde{T}^k$, $\|\mathbf{P}_{T^k}^{\perp}\boldsymbol{\phi}_i\|_2 =$

0, $\langle \boldsymbol{\phi_i}, \boldsymbol{r}^k \rangle = 0$. We now use the framework of the proof of success at the $(k+1)^{\text{th}}$ iteration of Theorem 3 of Wang et al. [5]. More specifically, we define two variables $u_1^k$, $v_L^k$ and derive a lower bound on $u_1^k$ and an upper bound on $v_L^k$ which are then compared to find a sufficient condition that ensures $h^{k+1} \cap T \setminus T^k \neq \emptyset$. Define $V^{k+1} = \arg\max_{S \subset S^{k+1} \setminus (T \setminus T^k): |S| = L} \sum_{i \in S} a_i$, $u_1^k := \max_{i \in S^{k+1} \cap T \setminus T^k} a_i \equiv \max_{i \in S^{k+1} \cap T} a_i$, and $v_L^k = \min_{i \in V^{k+1}} a_i$. Clearly, $h^{k+1} \cap T \setminus T^k \neq \emptyset$ if $u_1^k > v_L^k$. Now, following the steps used in arriving at Eq. (F.1) of Wang and Li [5], and recalling that $\left|\langle \boldsymbol{\phi}_i, \boldsymbol{r}^k \rangle\right| = 0 \; \forall i \in T^k$, one can obtain $u_1^k \geq \frac{\|\boldsymbol{\Phi}^H_{T \setminus T^k} \boldsymbol{r}^k\|_2}{\sqrt{K - c_k}} = \frac{\|\boldsymbol{\Phi}^H_{T \cup T^k} \boldsymbol{r}^k\|_2}{\sqrt{K - c_k}}$. Now, recalling that $\boldsymbol{r}^k = \boldsymbol{\Phi}_{T \cup T^k} \boldsymbol{x}'_{T \cup T^k} + \mathbf{P}^{\perp}_{T^k} \boldsymbol{e}$ and using the third result of Proposition 3.1 of Needell and Tropp [9], one obtains,

$$u_1^k \geq \frac{1}{\sqrt{K - c_k}} \left[ (1 - \delta_{LK-L+1}) \|\boldsymbol{x}'_{T \cup T^k}\|_2 - \sqrt{1 + \delta_{LK-L+1}} \|\boldsymbol{e}\|_2 \right]. \tag{A.7}$$

On the other hand, from the steps used to derive Eq. (F.3) in [5], one has $v_L^k \leq \frac{\frac{1}{\sqrt{L}} \|\boldsymbol{\Phi}^H_{S^{k+1} \setminus T} \boldsymbol{r}^k\|_2}{\min_{S^{k+1} \setminus (T \cup \tilde{T}^k)} \|\mathbf{P}^{\perp}_{T^k} \boldsymbol{\phi}_i\|_2}$. Now, note that for any $i \in S^{k+1} \setminus (\tilde{T}^k \cup T)$, $\|\mathbf{P}^{\perp}_{T^k} \boldsymbol{\phi}_i\|_2 = \|\boldsymbol{A}_{T^k} \boldsymbol{v}_i\|_2$, where the vector $\boldsymbol{v}_i \in \mathbb{C}^n$ has the $i$th coordinate as 1, and all other coordinates 0. Using Lemma 1 of Satpathi et al. [15] and the monotonicity of RIC, along with the fact that $Lk < N + LK - L$, it then follows that, $\forall i \in S^{k+1} \setminus (\tilde{T}^k \cup T)$,

$$\|\mathbf{P}^{\perp}_{T^k} \boldsymbol{\phi}_i\|_2 \geq \sqrt{1 - \left( \frac{\delta_{Lk+1}}{1 - \delta_{Lk+1}} \right)^2} \geq \sqrt{1 - \gamma^2}, \tag{A.8}$$

where $\gamma = \left( \frac{\delta_R}{1 - \delta_R} \right)$. Again, using the above expression for $\boldsymbol{r}^k$ along with Lemma 1.2 of Dai and Milenkovic [8], one obtains,

$$\begin{aligned} \|\boldsymbol{\Phi}^H_{S^{k+1} \setminus T} \boldsymbol{r}^k\|_2 &= \|\boldsymbol{\Phi}^H_{S^{k+1} \setminus (T \cup T^k)} \left( \boldsymbol{\Phi}_{T \cup T^k} \boldsymbol{x}'_{T \cup T^k} + \mathbf{P}^{\perp}_{T^k} \boldsymbol{e} \right)\|_2 \\ &\leq \delta_{Lk+K+N-m_{k+1}-c_k} \|\boldsymbol{x}'_{T \cup T^k}\|_2 + \sqrt{1 + \delta_{N-m_{k+1}}} \|\boldsymbol{e}\|_2 \\ &\leq \delta_R \|\boldsymbol{x}'_{T \cup T^k}\|_2 + \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2, \end{aligned} \tag{A.9}$$

where we use monotonicity of the RIC and the relation $Lk + K - c_k \leq LK - L + 1$ obtained earlier. The inequalities (A.8) and (A.9) result in

$$v_L^k \leq \frac{\delta_R \|\boldsymbol{x}'_{\tilde{T}^k \cup T}\|_2 + \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2}{\sqrt{L}} (1 - \gamma^2)^{-1/2}. \tag{A.10}$$

In order to ensure that the denominator of the RHS of above remains real, we need $\delta_R < 1/2$. This is seen to be satisfied trivially by the proposed sufficient condition (1), which is derived below. Then, from (A.7) and (A.10), using $\delta_{LK-L+1} < \delta_{LK-L+N+1}$, $h^{k+1} \cap T \setminus T^k \neq \emptyset$ is ensured if

$$\begin{aligned} &\frac{1}{\sqrt{K - c_k}} \left[ (1 - \delta_R) \|\boldsymbol{x}'_{T \cup T^k}\|_2 - \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2 \right] \\ &> \frac{\delta_R \|\boldsymbol{x}'_{T \cup T^k}\|_2 + \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2}{\sqrt{L}} (1 - \gamma^2)^{-1/2}. \end{aligned} \tag{A.11}$$

Hence, $S^{k+1} \cap T \setminus T^k \neq \emptyset$, and $h^{k+1} \cap T \setminus T^k \neq \emptyset$ is ensured if

$$\begin{aligned} &\frac{1}{\sqrt{K - c_k}} \left[ (1 - \delta_R) \|\boldsymbol{x}'_{T \cup T^k}\|_2 - \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2 \right] \\ &> \max \left\{ \frac{1}{\sqrt{N}}, \frac{1}{\sqrt{L\left(1 - \left(\frac{\delta_R}{1 - \delta_R}\right)^2\right)}} \right\} \\ &\times \left( \delta_R \|\boldsymbol{x}'_{T \cup T^k}\|_2 + \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2 \right) \\ &\overset{N \geq L}{\equiv} \frac{1}{\sqrt{L}} \left( \delta_R \|\boldsymbol{x}'_{T \cup T^k}\|_2 + \sqrt{1 + \delta_R} \|\boldsymbol{e}\|_2 \right) (1 - \gamma^2)^{-1/2} \\ &\Leftrightarrow \|\boldsymbol{x}'_{T \cup T^k}\|_2 \left( \sqrt{L(1 - \gamma^2)} - \gamma \sqrt{K - c_k} \right) \\ &> \left( \sqrt{(1+\gamma)(1+2\gamma)} \left( \sqrt{K - c_k} + \sqrt{L(1 - \gamma^2)} \right) \right) \|\boldsymbol{e}\|_2. \end{aligned} \tag{A.12}$$

Note that since the RHS of (A.12) is positive, we require $\sqrt{L(1 - \gamma^2)} - \gamma \sqrt{K - c_k} > 0 \Leftrightarrow \delta_R < \frac{\sqrt{L}}{\sqrt{L} + \sqrt{L + K - c_k}}$, for (A.12) to hold. It is also seen from (A.12) that for noise-free cases, i.e., when $\|\boldsymbol{e}\|_2 = 0$, the above provides a sufficient condition for success at the $(k+1)^{th}$ iteration. Further, this also satisfies the aforementioned requirement of $\delta_R < 1/2$. A stronger condition on $\delta_R$ which is independent of $c_k$ will then be given by

$$\delta_R < \frac{\sqrt{L}}{\sqrt{L} + \sqrt{L + K}}, \tag{A.13}$$

which proves (1)[3] Subject to satisfaction of (A.13), a sufficient condition for success at the $(k+1)^{th}$ iteration is obtained from (A.12) as,

$$\frac{\|\boldsymbol{x}'_{T \cup T^k}\|_2}{\|\boldsymbol{e}\|_2} > \frac{\sqrt{(1+\gamma)(1+2\gamma)} \left( \sqrt{K - c_k} + \sqrt{L(1 - \gamma^2)} \right)}{\sqrt{L(1 - \gamma^2)} - \gamma \sqrt{K - c_k}} \tag{A.14}$$

The lower bound of (A.14) can be further simplified by noting that the RHS of (A.14) can be upper bounded as below:

$$\begin{aligned} RHS\, of\, (A.1) &< \sqrt{(1+\gamma)(1+2\gamma)} \frac{\sqrt{K} + \sqrt{L(1 - \gamma^2)}}{\sqrt{L(1 - \gamma^2)} - \gamma \sqrt{K}} \\ &= \sqrt{\frac{1}{1 - \delta_R} \cdot \frac{1 + \delta_R}{1 - \delta_R}} \cdot \frac{\frac{\sqrt{K}(1 - \delta_R) + \sqrt{L(1 - 2\delta_R)}}{1 - \delta_R}}{\frac{\sqrt{L(1 - 2\delta_R)} - \delta_R \sqrt{K}}{1 - \delta_R}} \\ &= \frac{\sqrt{1 + \delta_R}}{1 - \delta_R} \frac{\sqrt{K}(1 - \delta_R) + \sqrt{L(1 - 2\delta_R)}}{\sqrt{L(1 - 2\delta_R)} - \delta_R \sqrt{K}} \\ &< \frac{\sqrt{1 + \delta_R}(\sqrt{K} + \sqrt{L})}{\sqrt{L(1 - 2\delta_R)} - \delta_R \sqrt{K}}, \end{aligned}$$

since $\sqrt{L(1 - 2\delta_R)} < \sqrt{L}(1 - \delta_R)$. Thus, a modified condition for success at the $(k+1)^{th}$ iteration which also implies (A.14) is given by

$$\frac{\|\boldsymbol{x}'_{T \cup T^k}\|_2}{\|\boldsymbol{e}\|_2} > \frac{\sqrt{1 + \delta_R}(\sqrt{K} + \sqrt{L})}{\sqrt{L(1 - 2\delta_R)} - \delta_R \sqrt{K}}, \tag{A.15}$$

along with the condition (A.13). Next, from the definition of $\kappa$ (section IV), $\|\boldsymbol{x}'_{T \cup T^k}\|_2 \geq \|\boldsymbol{x}_{T \setminus T^k}\|_2 \geq |T \setminus T^k| \min_{j \in T} |x_j| = \|\boldsymbol{x}\|_2 \cdot \kappa \cdot \sqrt{\frac{K - c_k}{K}} \geq \frac{\|\boldsymbol{x}\|_2 \cdot \kappa}{\sqrt{K}}$, since $\min_{j \in T \setminus T^k} |x_j| \geq \min_{j \in T} |x_j|$ and $c_k \leq K - 1$. Combining with Eq. (A.15), we obtain a sufficient condition for successful recovery at any step $k$, $k \geq 2$ as

$$\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{e}\|_2} > \frac{\sqrt{1 + \delta_R}(\sqrt{K} + \sqrt{L})\sqrt{K}}{\kappa \left( \sqrt{L(1 - 2\delta_R)} - \delta_R \sqrt{K} \right)}, \tag{A.16}$$

along with the condition (A.13).

[3] We may emphasize here that this rather simple bound on $\delta_R$ could be obtained because of our use of Lemma 3.5, which led to a tractable function of $\delta_R$, i.e., $\sqrt{L(1 - \gamma^2)} - \gamma \sqrt{K - c_k}$ where $\gamma = \frac{\delta_R}{1 - \delta_R}$. On the other hand, instead of using Lemma 3.5, if we had used the inequality [5, Eq. (E.7)], it would have resulted in a complicated fourth order polynomial in $\delta_R$, difficult to simplify to produce a bound on $\delta_R$.

*A.0.3. Condition for overall success*

The condition for overall success is obtained by combining the conditions for success for $k = 1$ and for $k \geq 2$. In order to do this, we first note that $\delta_L < \delta_{L+K}, \ \delta_K < \delta_{L+K}$. Using this in (A.17), a stronger condition for success at the first iteration will be given by

$$\frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{e}\|_2} > \frac{(\sqrt{L}+\sqrt{K})\sqrt{1+\delta_{L+K}}}{\sqrt{L}(1-\delta_{L+K})-\sqrt{K}\delta_{L+K}}, \tag{A.17}$$

with the additional requirement

$$\delta_{L+K} < \frac{\sqrt{L}}{\sqrt{L}+\sqrt{K}}, \tag{A.18}$$

which is required to maintain $\sqrt{L}(1-\delta_{L+K})-\sqrt{K}\delta_{L+K} > 0$. Since $R-(L+K) = N-L+(L-1)(K-1) \geq 0$, as both $L, K$ are positive integers, we see that the condition in Eq. (A.13) implies the condition in Eq. (A.18).

On the other hand, it is easy to check that the denominator of the RHS of (A.16) is smaller than that of the RHS of (A.17). First, by definition, $\kappa < 1$ and $\delta_R > \delta_{L+K}$. Furthermore,

$$(1-2\delta_{LK+N-L+1})-(1-\delta_{L+K})^2 = -\delta_{L+K}^2+2(\delta_{L+K}-\delta_{N+LK-L+1}) < 0.$$

Again, it is easily seen that the numerator of the RHS of (A.16) is larger than that of the RHS of (A.17). The overall implication of these is that the condition in (A.16) implies the condition in (A.17). Finally, noting that $\|\boldsymbol{\Phi x}\|_2 \leq \sqrt{1+\delta_K}\|\boldsymbol{x}\|_2 < \sqrt{1+\delta_{LK+N-L+1}}\|\boldsymbol{x}\|_2$, the conditions stated in Theorem 4.1 are sufficient for overall successful recovery. This proves Theorem 4.1.

## References


[1] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.

[2] J. Wang, S. Kwon, B. Shim, Generalized orthogonal matching pursuit, IEEE Trans. Signal Process. 60 (12) (2012) 6202–6216.

[3] C. Soussen, R. Gribonval, J. Idier, C. Herzet, Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares, IEEE Trans. Inf. Theory 59 (5) (2013) 3158–3174.

[4] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, Int. J. Control 50 (5) (1989) 1873–1896.

[5] J. Wang, P. Li, Recovery of sparse signals using multiple orthogonal least squares, IEEE Trans. Signal Process. 65 (8) (2017) 2049–2062, doi:10.1109/TSP.2016.2639467.

[6] S. Mukhopadhyay, S. Satpathi, M. Chakraborty, A low complexity orthogonal least squares algorithm for sparse signal recovery, in: 2018 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2018, pp. 75–79.

[7] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215, doi:10.1109/TIT.2005.858979.

[8] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inf. Theory 55 (5) (2009) 2230–2249.

[9] D. Needell, J.A. Tropp, Cosamp: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.

[10] S.K. Ambat, K. Hari, An iterative framework for sparse signal reconstruction algorithms, Signal Process. 108 (2015) 351–364.

[11] L. Jacques, A short note on compressed sensing with partially known signal support, Signal Process. 90 (12) (2010) 3308–3312.

[12] T. Ince, A. Nacaroglu, N. Watsuji, Nonconvex compressed sensing with partially known signal support, Signal Process. 93 (1) (2013) 338–344.

[13] W. Chen, Y. Li, G. Wu, Recovery of signals under the high order rip condition via prior support information, Signal Process. 153 (2018) 83–94.

[14] L. Rebollo-Neira, D. Lowe, Optimized orthogonal matching pursuit approach, IEEE Signal Process. Lett. 9 (4) (2002) 137–140.

[15] S. Satpathi, R.L. Das, M. Chakraborty, Improving the bound on the rip constant in generalized orthogonal matching pursuit, IEEE Signal Process. Lett. 20 (11) (2013) 1074–1077.

[16] A.K. Fletcher, S. Rangan, Orthogonal matching pursuit: a brownian motion analysis, IEEE Trans. Signal Process. 60 (3) (2012) 1010–1021.

[17] J. Wen, Z. Zhou, D. Li, X. Tang, A novel sufficient condition for generalized orthogonal matching pursuit, IEEE Commun. Lett. 21 (4) (2017) 805–808.

[18] J. Wen, J. Wang, Q. Zhang, Nearly optimal bounds for orthogonal least squares, IEEE Trans. Signal Process. 65 (20) (2017) 5347–5356.

[19] B. Li, Y. Shen, Z. Wu, J. Li, Sufficient conditions for generalized orthogonal matching pursuit in noisy case, Signal Process. 108 (2015) 111–123.

[20] G.H. Golub, C.F. Van Loan, Matrix Computations, 3, JHU Press, 2012.