Sparse Signal Recovery Using gOMP Assisted mOLS

Samrat Mukhopadhyay, Siddhartha Satpathi and Mrityunjoy Chakraborty
Department of Electronics and Electrical Communication Engineering,
Indian Institute of Technology, Kharagpur, INDIA
Email: 1samratphysics@gmail.com, 2sidd.piku@gmail.com, 3mrityun@ece.iitkgp.ernet.in

Abstract—Because of fast convergence in finite number of steps and low computational complexity, signal recovery from compressed measurements using greedy algorithms have generated a large amount of interest in recent years. Among these greedy algorithms OMP is well studied and recently its generalization, gOMP, have also drawn attention. On the other hand OLS and its generalization mOLS have been studied in recent years because of their potential advantages in signal recovery guarantee compared to OMP or gOMP. But OLS and mOLS have the shortcomings of high computational complexity. In this paper we propose a new algorithm which uses gOMP to preselect a N length set out of which mOLS selects its L best coefficients. We have shown that our new algorithm, named mOLS, is guaranteed to reconstruct a K sparse signals perfectly from compressed measurements if the restricted isometry constant satisfies $\delta_{LK+N} < \frac{\sqrt{L}}{\sqrt{L} + \sqrt{K + L}}$. Moreover experimental results indicate that mOLS selects its potential advantages in signal recovery guarantee compared to OLS and is much faster than OLS.

Index Terms—Compressive Sensing, gOMP, mOLS, restricted isometry property

I. INTRODUCTION

There has been considerable interest for some time to recover signals from their partial measurements. Mathematically the problem can be expressed as finding solution for the following set of linear equations

$$y = \Phi x$$

where $\Phi \in \mathbb{R}^{m \times n}$ and $m < n$. Since this is an underdetermined set of linear equations, the number of solutions is infinite, without further information about structure of $x$. A natural problem that then follows is to find solutions for these set of equations which has the lowest sparsity i.e. lowest number of nonzero elements, i.e the constrained $l_0$ norm minimization problem

$$\min_{x \in \mathbb{R}^n} \|x\|_0$$

s.t. $y = \Phi x$

However, this problem requires combinatorial search which is computationally intractable. The way out to this problem was found by Candes and Tao [1] in a seminal paper where it was shown that the $l_0$ norm can be solved exactly using $l_1$ norm optimization if the number of measurements is $O(K\ln n)$ where $K$ is the sparsity of the vector $x$. The problem that is required to solve under this condition is

$$\min_{x \in \mathbb{R}^n} \|x\|_1$$

s.t. $y = \Phi x$

This is called the basis pursuit(BP) [2]. This can be solved by standard Linear programming approach.

A. Greedy approaches

Linear programming approach has cubic complexity on the number of columns $n$, which can be too slow for large $n$ ($\sim 1000$ is common in many applications). In that case greedy approaches play the role of useful alternatives. A greedy algorithm utilizes some greedy rule to reconstruct the vector at each step of the algorithm. They are often faster than LP. Orthogonal Matching Pursuit(OMP) [3][4], Orthogonal Least Squares(OLS) [5][6] are among the earliest of these greedy algorithms. These algorithms greedily select indices at each step of the algorithm and append them to the already constructed support create a successively increasing support over which they take projections to get reconstructed signal. The only difference between these two algorithms is in the method of selecting an index in the identification step. At an iteration $k$ OMP chooses an index $i$ which maximizes the absolute correlation $\langle r^{k-1}, \phi_i \rangle$ where $r^{k-1}$ is the residual error vector at step $k-1$ and $\phi_i$ is a column of the measurement matrix $\Phi$. In OLS the identification is performed by choosing an index $i$ which would minimize the resulting residual vector norm. Wang et al introduced gOMP algorithm [7] and mOLS algorithm [8] as generalizations of OMP and OLS algorithms respectively. In gOMP and mOLS in the identification step, the algorithms choose multiple instead of single index. See Table. I for details.

The motivation for introducing both gOMP as well as mOLS is that the time complexity of the algorithms can be markedly improved if at each iteration the algorithms choose multiple indices instead of just one. This increases the probability of choosing a correct index at an iteration and hence, for a given sparse signal with sparsity $K$, the algorithms converge faster than OMP or OLS. Wang et al [7] has shown that the computational complexity for OMP is about $2Kmn + 3K^2m$ floating point operations per second(flops) where $m \times n$ is the size of the measurement matrix and the computational complexity of gOMP is about $2snm + (2N^2 + N)s^2m$ flops where $N$ is the number of indices that gOMP chooses at a time at each iteration and $s$ is the number of iterations required for the algorithm to stop. By extensive simulation they have shown that gOMP indeed takes a small number of simulation $s$ to recover the $K$ sparse signal which effectively reduces its computational complexity. Similar conclusions are drawn for mOLS compared to OLS by Wang et al [8] through extensive simulations. Other algorithms of the same flavour...
TABLE I: gOMP AND mOLS ALGORITHM

(a) gOMP Algorithm

| Input: measurement vector \( y \in \mathbb{R}^m \), sensing matrix \( \Phi \in \mathbb{R}^{m \times n} \), sparsity level \( K \), total set \( \mathcal{H} = \{1, 2, \ldots, n\} \) |
|---|
| Initialize: counter \( k = 0 \), residue \( r^0 = y \), estimated support set, \( T^0 = \emptyset \) |
| While \( (\|r_k\|_2 \geq \epsilon \text{ and } k < K) \) |
| \( k = k + 1 \) |
| Identify: \( h^k \) is the set containing indices corresponding to \( N \) largest absolute entries of \( \Phi \) \( y \)
| Augment: \( T^k = T^{k-1} \cup h^k \)
| Estimate: \( x^k = \arg\min_{u \in \mathbb{R}^m, \supp(u) = T^k} \|y - \Phi u\|_2 \)
| Update: \( r^k = y - \Phi x^k \)
| End While |
| Output: estimated support set \( T = \arg\min_{S: |S| = K} \|x^k - x^k_S\|_2 \) and \( K \)-sparse signal \( \hat{x}_T = x^k_T \), \( x_{\mathcal{H} \setminus T} = 0 \) |

(b) mOLS Algorithm

| Input: measurement vector \( y \in \mathbb{R}^m \), sensing matrix \( \Phi \in \mathbb{R}^{m \times n} \), sparsity level \( K \), total set \( \mathcal{H} = \{1, 2, \ldots, n\} \) |
|---|
| Initialize: counter \( k = 0 \), residue \( r^0 = y \), estimated support set, \( T^0 = \emptyset \) |
| While \( (\|r_k\|_2 \geq \epsilon \text{ and } k < K) \) |
| \( k = k + 1 \) |
| Identify: \( h^k \) = \( \arg\min_{S \subseteq \mathcal{H}, |S| = L} \sum_{i \in S} \|P_{T^k \setminus \{i\}}^{\perp} y\|_2^2 \)
| Augment: \( T^k = T^{k-1} \cup h^k \)
| Estimate: \( x^k = \arg\min_{u \in \mathbb{R}^m, \supp(u) = T^k} \|y - \Phi u\|_2 \)
| Update: \( r^k = y - \Phi x^k \)
| End While |
| Output: estimated support set \( T = \arg\min_{S: |S| = K} \|x^k - x^k_S\|_2 \) and \( K \)-sparse signal \( \hat{x}_T = x^k_T \), \( x_{\mathcal{H} \setminus T} = 0 \) |

as gOMP or mOLS are stagewise OMP(StOMP) [9], regularized OMP(ROMP) [10], Compressed Sampling Matching Pursuit(CoSaMP) [11] and Subspace Pursuit (SP) [12]. OMP has been the representative of these algorithms because of its simplicity and recovery performance. Tropp and Gilbert [4] have shown that OMP can recover the original sparse vector from a few measurements with exceedingly high probability when the measurement matrix has entries i.i.d Gaussian. But recently Soussen et al [13] have shown that OLS performs considerably better than OMP in recovering sparse signals from few measurements when the measurement matrix has correlated entries. They have also shown that there exists measurement dictionaries for which OMP fails to recover the original signal whereas OLS is unaffected.

B. Our contributions in this paper

- In this paper we propose a new algorithm, referred to as m²OLS (short for mOMP assisted mOLS), that exploits the capability of gOMP to select “good” indices at a step as well as the robustness of mOLS in maintaining high recovery performance in the presence measurement matrices with correlated dictionaries.
- We find a criteria under which our proposed algorithm can exactly recover the unknown \( K \)-sparse signal which is based upon the restricted isometry property of the measurement matrix. A matrix \( \Phi \in \mathbb{R}^{m \times n} \) is said to have restricted isometry property (RIP) of order \( K \) if \( \exists \delta > 0 \) such that \( \forall x \in \mathbb{R}^n \) with sparsity \( K \), the matrix \( \Phi \) satisfies the following inequalities

\[
(1 - \delta) \|x\|^2 \leq \|\Phi x\|^2 \leq (1 + \delta) \|x\|^2
\]

The smallest \( \delta \) for which this property is satisfied is called the restricted isometry constant of order \( K \) of the matrix \( \Phi \) and is denoted by \( \delta_K \). We show that m²OLS with gOMP selecting \( N \) indices and mOLS selecting \( L \leq N \) indices (See Table. [I] for details of the algorithm) is guaranteed to recover a \( K \) sparse vector exactly from compressed measurements if the matrix \( \Phi \) obeys RIP with constant

\[
\delta_{LK+N} < \frac{\sqrt{L}}{\sqrt{K} + L + \sqrt{L}}
\]

- We compare simulation results for recovery performances of OMP, OLS and m²OLS and show that m²OLS performs much better than OMP in recovering sparse vector for measurement matrices with correlated dictionary. Also, gOMPmaOLS is much faster than OLS and thus becomes a suitable upgradation of both OMP and OLS.

II. PRELIMINARIES

A. Notation

The following notation will be used throughout the paper.

- \( \| \cdot \|_p \) denotes the \( l^p \) norm of a vector \( v \), i.e., \( \|v\|_p = \left( \sum_{i=1}^{n} |v_i|^p \right)^{1/p} \)
- \( \Phi \in \mathbb{R}^{m \times n} \) will denote the measurement matrix with \( m \) rows and \( n \) columns with \( m < n \)
- The \( i \) th column of \( \Phi \) will be denoted by \( \phi_i; \ i = 1, 2, \ldots, n \). All the columns of \( \Phi \) are assumed to have unit \( l_2 \) norm.
- \( K \) will denote the sparsity level estimated beforehand
- \( x_S \) will denote the vector \( x \) restricted to a subset \( S \subset \{1, 2, \ldots, n\} \). Similarly, \( \Phi_S \) will denote the submatrix of \( \Phi \) formed with the columns restricted to index set \( S \)
- \( \mathcal{H} \) will denote the set of all the indices \( \{1, 2, \ldots, n\} \)
- \( T \) will refer to the support set of the original vector \( x \)
- If \( \Phi_S \) has column rank \( |S| \) \( (|S| < m) \), then \( \Phi_S^\perp = (\Phi_S^\perp \Phi_S)^{-1} \Phi_S^\perp \) will denote the Moore-Penrose pseudo-inverse of \( \Phi_S \)
- \( P_S = \Phi_S \Phi_S^\perp \) will denote the projection operator on \( \text{span}(\Phi_S) \) and \( P^\perp_S = I - P_S \) will denote the projection operator on the orthogonal complement of \( \text{span}(\Phi_S) \)
- \( \mathbf{A}_\Lambda \) is defined, for any \( \Lambda \subset \mathcal{H} \), \( \mathbf{A}_\Lambda = P_\Lambda \mathbf{P}_\Lambda \)

B. Lemmas

The following lemmas will be useful for analysis of our algorithm.
Lemma 2.1 (monotonicity of RIC, Lemma 1 of [12]). 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 2.2 (Lemma 1 of [12]). If $x \in \mathbb{R}^n$ is a vector with support $S_1$, and $S_1 \cap S_2 = \emptyset$, then,

$$\|\Phi_{S_1} x\|_2 \leq \delta_{|S_1|+|S_2|} \|x\|_2$$

Lemma 2.3 (Lemma 3 of [14]). If $I_1, I_2 \subset \mathcal{H}$ such that $I_1 \cap I_2 = \emptyset$, then, $\forall u \in \mathbb{R}^n$ such that $supp(u) \subset I_2$,

$$\left(1 - \left(\frac{\delta_{I_1} |I_1| + |I_2|}{1 - \delta_{I_1} |I_1| + |I_2|}\right)^2\right) \|\Phi_{I_1} u\|_2^2 \leq \|A_{I_1} u\|_2^2 \leq (1 + \delta_{|I_1|+|I_2|}) \|\Phi_{I_1} u\|_2^2$$

and,

$$\left(1 - \frac{\delta_{I_1} |I_1| + |I_2|}{1 - \delta_{I_1} |I_1| + |I_2|}\right) \|u\|_2^2 \leq \|A_{I_1} u\|_2^2 \leq (1 + \delta_{|I_1|+|I_2|}) \|u\|_2^2$$

III. PROPOSED ALGORITHM

| TABLE II: m²OLS Algorithm |

| Input: | measurement vector $y \in \mathbb{R}^m$, sensing matrix $\Phi \in \mathbb{R}^{m \times n}$, sparsity level $K$, true support set $T = supp(x)$, total set $\mathcal{H} = \{1, 2, \ldots, n\}$, $1 \leq N, L \leq K$ |
| Initialize: | counter $k = 0$, residue $r^0 = y$, estimated support set, $T^0 = \emptyset$, set selected by gOMP $S^0 = \emptyset$ |
| While ($|x_k|_2 \geq 0$ and $k < K$) |
| Preselect: | $S^k$ is the set containing indices corresponding to the $N$ largest absolute entries of $\Phi r^{k-1}$ |
| Identify: | $h^k = \arg \max_{\Lambda \subset S^k \setminus \{i\}} \|P_{T^{k-1} \cup \{i\}} y\|_2^2$ |
| Augment: | $T^k = T^{k-1} \cup h^k$ |
| Estimate: | $x^k = \arg \min_{u \in \mathbb{R}^n, supp(u) = T^k} \|y - \Phi x\|_2$ |
| Update: | $r^k = y - \Phi x^k$ |
| End While |

Output: estimated support set $\hat{T} = \arg \min_{S^k \subset S^k \setminus \emptyset} \|x^k - x^k_{S^k}\|_2$ and $K$-sparse signal $\hat{x} \in \mathbb{R}^n$, $supp(\hat{x}) = T^k$, $\hat{x}_{K^c \setminus T} = 0$.

IV. SIGNAL RECOVERY USING m²OLS ALGORITHM WITHOUT NOISE

We seek the conditions under which m²OLS algorithm is guaranteed to recover a $K$-sparse signal within $K$ iterations. We say that m²OLS algorithm makes a success at an iteration if it selects at least one correct index at that iteration. Now, the identification step in the m²OLS algorithm can be expressed in a way convenient to our analysis by the virtue of the following lemma

Lemma 4.1. At the $(k+1)$th iteration, the identification step chooses the set

$$h^{k+1} = \arg \max_{S: S^k \subset S^k \setminus \{i\}} \|P_T^1 \Phi_i \|_2^2$$

Moreover, if

$$g^{k+1} = \arg \max_{S: S^k \subset S^k \setminus \{i\}} \|P_T^1 \Phi_i \|_2^2$$

then,

$$\sum_{i \in h^{k+1}} \|\Phi_i \|_2 = \sum_{i \in g^{k+1}} \|\Phi_i \|_2$$

Proof: This lemma is a direct consequence of Proposition 1 of [18]. Nevertheless we provide a different proof in Appendix A that utilizes properties of a vector space. 

1) Success at the first iteration: The preselection step gives the set $S^1$ which consists of the indices corresponding to the $N$ largest coordinates of $\Phi y$. Now, $k = 0 \implies \|P_T^1 \Phi_i \|_2 = \|\Phi_i \|_2 = 1$. Hence

$$T^1 = h^1 = \arg \max_{S^1 \subset S^1 \setminus \emptyset} \|P_T^1 \Phi_i \|_2$$

Now we make the following observations and state them in the form of the following lemma:

Lemma 4.2.

$$\frac{1}{\sqrt{N}} \|\Phi_{T^1} y\|_2 \geq \frac{1}{\sqrt{K}} \|\Phi_{T^1} y\|_2, \quad (N \leq K)$$

$$\|\Phi_{T^1} y\|_2 \geq \|\Phi_{T^1} y\|_2, \quad (N > K)$$

Now, we find the recovery conditions separately for the cases $N \leq K$ and $N > K$.

Proof: The proof is postponed to Appendix A. 

Our proposed algorithm is described in Table III. It is different from the conventional mOLS algorithm (described in Table IIA) in the identification step as well as a new step, that we call preselection. A conventional mOLS algorithm searches for the new index in $\mathcal{H}$. This search procedure is conducted by going to each index and measuring the norm of residue if this index were used in the augmentation step. The $L$ indices which gives the minimum residue norms arranged in ascending order, are chosen. For large $n$, this step itself is computationally expensive. We modify this step by breaking it into a preselection step and an identification step. The preselection step uses gOMP (described in Table IIA) to contract the set of potential candidates to a set of cardinality $N(<n)$, choosing the indices corresponding to the $N$ largest coordinates of $\Phi r^{k-1}$, norm wise. The resulting set is then considered for the mOLS identification step.
1) When $N \leq K$. In this case, we have, from Eq. (5)
\[
\|\Phi_S^i y\|_2 \geq \sqrt{\frac{N}{K}} \|\Phi_T^i y\|_2 \\
= \sqrt{\frac{N}{K}} \|\Phi_T^i \Phi_T x_T\|_2 \\
\geq \sqrt{\frac{N}{K}} (1 - \delta_K) \|x\|_2 \tag{6}
\]
Now if no correct index were selected in $S^1$, then the first iteration is bound to fail. Then
\[
\|\Phi_S^i y\|_2 = \|\Phi_S^i \Phi_T x_T\|_2 \leq \delta_{K+N} \|x\|_2 \tag{7}
\]
So, $S^1$ is guaranteed to contain at least one correct index if
\[
\delta_{K+N} < \sqrt{\frac{N}{K}} (1 - \delta_K) \\
\implies \sqrt{K} \delta_{K+N} + \sqrt{K} \delta_K < \sqrt{N} \tag{8}
\]
2) When $N > K$, the observations in Eq. (4) is valid. Hence, after performing a series of deductions similar to the case $N \leq K$, we get the following sufficient condition for $S^1$ to contain at least one correct index:
\[
\delta_{K+N} + \delta_K < 1 \tag{9}
\]
In both the cases, Eq. (5) is valid, from which we get
\[
\|\Phi_T^i y\|_2 \geq \sqrt{\frac{L}{K}} \|\Phi_T^i x_T\|_2 \\
= \sqrt{\frac{L}{K}} \|\Phi_T^i \Phi_T x_T\|_2 \\
\geq \sqrt{\frac{L}{K}} (1 - \delta_K) \|x\|_2 \tag{10}
\]
Even if $S^1$ consists of at least one correct index, a correct index may not be selected at the first iteration. Then
\[
\|\Phi_T^i y\|_2 = \|\Phi_T^i \Phi_T x_T\|_2 \leq \delta_{K+L} \|x\|_2 \tag{11}
\]
Thus sufficient condition for success in the first iteration, given the condition in Eq. (8) or Eq. (9) is guaranteed by
\[
\delta_{K+L} < \sqrt{\frac{L}{K}} (1 - \delta_K) \\
\implies \sqrt{K} \delta_{K+L} + \sqrt{L} \delta_K < \sqrt{L} \tag{12}
\]
Now, if we have
\[
\delta_{N+K} < \frac{\sqrt{L}}{\sqrt{L} + \sqrt{K}} \tag{13}
\]
then, since $L \leq N$, by monotonicity $\delta_{L+K} \leq \delta_{N+K} < \frac{\sqrt{L}}{\sqrt{L} + \sqrt{K}}$ and thus the inequality in Eq. (12) is satisfied. Also, since $L \leq N$, under the condition in Eq. (13), $\delta_{N+K} < \frac{\sqrt{L}}{\sqrt{L} + \sqrt{K}} \leq \frac{\sqrt{N}}{\sqrt{N} + \sqrt{K}}$. So, if $N \leq K$, inequality in Eq. (8) is satisfied. Further, if $N > K$, it is easy to see that the inequality in Eq. (9) is satisfied under the condition in Eq. (13), since $L \leq K$. Thus under the condition in Eq. (13) the first iteration is guaranteed to succeed.

\[
\begin{align*}
\beta^k_1 & = \max_{i \in T} |\langle \phi_i, r^k \rangle|^2 \tag{16} \\
\alpha^k_N & = \min_{i \in W^{k+1}} |\langle \phi_i, r^k \rangle|^2 \tag{17} \text{ where } W^{k+1} := \arg \max_{S \in \{T:|S|=N\}} \|\Phi_S^{k+1}\|_2
\end{align*}
\]
Then at $(k+1)$th iteration, we can see that the set $S^{k+1}$ is guaranteed to consist of at least one index from $T$ if we have $\beta^k_1 > \alpha^k_N$. We can find a condition for this along the lines of [14].

As noted in [14], it can be seen that $r^k \in \text{Span}(\Phi_{T \cap T^k})$. This is because, looking at the relationship between $T, T^k$ in Fig. 1 and keeping in mind the property of the projection operator, we find
\[
\begin{align*}
r^k & = y - \Phi_T^k \Phi_T^k y \\
& = \Phi_T^k x_T - P_{T^k} \Phi_T^k x_T \\
& = (\Phi_{T \cap T^k} x_{T \cap T^k} + \Phi_{T \setminus T^k} x_{T \setminus T^k}) - P_{T^k} (\Phi_{T \cap T^k} x_{T \cap T^k} + \Phi_{T \setminus T^k} x_{T \setminus T^k}) \\
& = (\Phi_{T \cap T^k} x_{T \cap T^k} + \Phi_{T \setminus T^k} x_{T \setminus T^k}) - \Phi_{T \setminus T^k} x_{T \setminus T^k} - P_{T^k} \Phi_{T \setminus T^k} x_{T \setminus T^k} \\
& = \Phi_{T \cap T^k} x_{T \cap T^k} - \Phi_{T^k} u_{T^k} \\
& = \Phi_{T \cup T^k} x_{T \cup T} - \Phi_{T^k} u_{T^k} \\
& = \Phi_{T \cup T^k} x_{T \cup T} - \Phi_{T^k} u_{T^k} \tag{18}
\end{align*}
\]
where
\[
x_{T \cup T} = \begin{bmatrix} x_{T \cap T^k} \\ x_{T \setminus T^k} \end{bmatrix} \tag{19}
\]
Then
\[
\alpha_N^k = \min_{i \in W^{k+1}} |\langle \phi_i, r^k \rangle|^2 \\
\leq \frac{||\Phi_{W^{k+1}} r^k||^2}{N} \\
= \frac{||\Phi_{W^{k+1}\setminus T} r^k||^2}{N} (\because \Phi_T r^k = 0) \\
= \frac{||\Phi_{W^{k+1}\setminus T} \Phi_{T \cup T} x'_{T \cup T}||^2}{N} \\
\leq \frac{1}{N} \delta_{Lk+N} ||x'_{T \cup T}||^2 \\
\leq \frac{1}{N} \delta_{Lk+N} ||x'_{T \cup T}||^2
\] (20)

The last two steps use the facts that \(|W^{k+1}\setminus T \cup (T \cup T^k)| = |W^{k+1}\setminus T^k| + |T \cup T^k| \leq N + Lk + K - c_k\) and that \(c_k \geq k, k \leq K\) along with the monotonic increasing property of RIC.

To find a lower bound of \(\beta^k_1\), we proceed as follows

\[
\beta^k_1 = ||\Phi_T r^k||^2 \\
\geq \frac{||\Phi_T r^k||^2}{K} \\
= \frac{1}{K} ||\Phi_T \Phi_{T \setminus T} r^k||^2 \\
= \frac{1}{K} ||\Phi_{T \cap T} \Phi_{T \setminus T} x'_{T \setminus T}||^2 \\
\geq \frac{1}{K} (1 - \delta_{Lk-K-c_k})^2 ||x'_{T \setminus T}||^2 \\
\geq \frac{1}{K} (1 - \delta_{Lk+N})^2 ||x'_{T \setminus T}||^2
\] (21)

where the last step follows from the monotonic increasing property of RIC along with the observation that \(Lk + K - c_k \leq Lk + K - k \leq K - (K - 1)(L - 1) + K = LK - (L - 1) \leq LK + N\), which follows since \(c_k \geq k, k < K\), and \(N \geq L\). Thus from Eq. (20) and Eq. (21) we see that
\[
\delta_{Lk+N} < \frac{\sqrt{N}}{\sqrt{K} + \sqrt{N}}
\] (22)

will assure that atleast one correct index from \(T\) is chosen in \(S^{k+1}\).

**Condition to ensure Eq. (15):** Eq. (22) guarantees \(m_{k+1} = |T \cap S^{k+1}| \geq 1\). Now, along the lines of (8) let us define
\[
u_L = \min_{i \in V^{k+1}} |\langle \phi_i, r^k \rangle| \\
v_L = \min_{i \in V^{k+1}} |\langle \phi_i, r^k \rangle| \\
\sqrt{L} \sum_{i \in V^{k+1}} |\langle \phi_i, r^k \rangle|^{2} \\
\leq \frac{1}{\sqrt{L}} \sum_{i \in V^{k+1}} \frac{\|\phi_i r^k\|}{||\Phi_{T^k} r^k||_2}
\] (23)

Now, \(\forall i \in S^{k+1} \setminus T\), we have
\[
\|\Phi_{T^k} r^k\|_2^2 = \|A_{T^k} e_i\|_2^2 \\
= \left(1 - \frac{\delta_{Lk+1}}{1 - \delta_{Lk+1}}\right) \|\phi_i\|_2^2
\] (28)
and
\[
\frac{1}{\sqrt{L}}\|\Phi_{S^{k+1}\setminus T} e^k\|_2 \\
= \frac{1}{\sqrt{L}}\|\Phi_{S^{k+1}\setminus (T\cup T^k)} e^k\|_2 (\because \Phi_T e^k = 0) \\
\leq \frac{1}{\sqrt{L}}\|\Phi_{S^{k+1}\setminus (T\cup T^k)} \Phi_{T\cup T^k} x'_{T\cup T^k}\|_2 \\
\leq \frac{1}{\sqrt{L}}\delta_{N+L} \|x'_{T\cup T^k}\|_2 \\
\leq \frac{1}{\sqrt{L}}\delta_{N+L} \|x'_{T\cup T^k}\|_2 \tag{29}
\]
The last two steps follow from the observations \(|(S^{k+1}\setminus (T\cup T^k)) \cup (T\cup T^k)| = |S^{k+1}\setminus T| + |T\cup T^k| \leq |S^{k+1}\setminus T| + |T\cup T^k|\) and \(m_k \geq 1, c_k \geq k, k \leq K\). Thus
\[
v_L \leq \frac{1}{L \left(1 - \left(\frac{\delta_{L_{k+1}}}{1 - \delta_{L_{k+1}}}\right)^2\right)} \delta_{N+L} \|x'_{T\cup T^k}\|_2 \\
< \frac{1}{L \left(1 - \left(\frac{\delta_{L_{k+1}}}{1 - \delta_{L_{k+1}}}\right)^2\right)} \delta_{N+L} \|x'_{T\cup T^k}\|_2 \tag{30}
\]
Then \(v_1 > v_L\) will be satisfied if we have
\[
\frac{1}{L \left(1 - \left(\frac{\delta_{L_{k+1}}}{1 - \delta_{L_{k+1}}}\right)^2\right)} \delta_{N+L} \|x'_{T\cup T^k}\|_2 < \frac{1}{\sqrt{K}} (1 - \delta_{L_{k+1}}) \|x'_{T\cup T^k}\|_2
\]
\[
\Rightarrow \frac{\sqrt{K}}{L} < \sqrt{1 - x^2} \quad (x := \frac{\delta_{L_{k+1}}}{1 - \delta_{L_{k+1}}}) \tag{31}
\]
Thus \(31\) is satisfied under
\[
\delta_{L_{k+1}} < \frac{\sqrt{K}}{L + \sqrt{L}} \tag{32}
\]
Thus we have the following theorem

**Theorem 4.1.** \(\text{m}^2\text{OLS}\) can recover a \(K\) sparse vector \(x \in \mathbb{R}^n\) perfectly from the measurement vector \(y = \Phi x, y \in \mathbb{R}^m, m < n,\) if
\[
\delta_{L_{k+1}} < \frac{\sqrt{K}}{L + \sqrt{L}} \tag{33}
\]
is satisfied by matrix \(\Phi\).

**Proof:** If Eq. \((33)\) is satisfied then Eq. \((32)\) is satisfied due to monotonicity of \(\delta\) from Lemma \((2.1)\) and since \(L \leq N\), Eq. \((22)\) is also satisfied.

V. SIGNAL RECOVERY USING \(\text{m}^2\text{OLS}\) ALGORITHM: IN THE PRESENCE OF NOISE

The measurement model considered here is the following:
\[
y = \Phi x + e \tag{34}
\]
where \(e\) is an unwanted noise vector. To characterize recovery performance of \(\text{m}^2\text{OLS}\) in the presence of noise, we use the following performance measures:

- \(\text{snr} := \frac{\|\Phi x\|_2^2}{\|e\|_2^2}\)
- minimum-to-average-ratio (MAR) \(\{15\}\)
\[
\kappa = \min_{j \in T} \frac{|x_j|}{\|x\|_2 / \sqrt{K}}
\]
The following theorem establishes the requirement on \(\text{snr}\) under which \(\text{m}^2\text{OLS}\) is guaranteed to recover the true support set in \(K\) iterations.

**Theorem 5.1.** Under the noisy measurement model described by Eq. \((34)\), \(\text{m}^2\text{OLS}\) is guaranteed to recover the true support set \(T\) in \(K\) iterations, if the sensing matrix \(\Phi\) satisfies equation \((35)\) and the \(\text{snr}\) satisfies following condition:
\[
\sqrt{\text{snr}} > \frac{(\sqrt{N} + \sqrt{K})(1 + \delta_{LK+N})\sqrt{K}}{\kappa \left(\sqrt{L(1 - 2\delta_{LK+N}) - \delta_{LK+N}\sqrt{K}}\right)} \tag{35}
\]

**A. Proof of Theorem 5.1**

1) Success at the first iteration: At the first iteration, the conditions for success are
\[
S^1 \cap T \neq \emptyset, \quad h^1 \cap T \neq \emptyset
\]
Now, note that,
\[
\|\Phi_{S^1} y\|_2 \geq \min \left\{1, \sqrt{\frac{N}{K}} \right\} \|\Phi_{S^1} y\|_2 \\
= \min \left\{1, \sqrt{\frac{N}{K}} \right\} \|\Phi_{S^1} \Phi_T y + \Phi_{S^1} e\|_2 \\
\geq \min \left\{1, \sqrt{\frac{N}{K}} \right\} \left[\|\Phi_{T} x_T + \Phi_{S^1} e\|_2 \right] \\
\geq \min \left\{1, \sqrt{\frac{N}{K}} \right\} \left[1 - \delta_K\|x\|_2 - \sqrt{1 + \delta_N\|e\|_2}\right]
\]
If \(S^1 \cap T = \emptyset\), then,
\[
\|\Phi_{S^1} y\|_2 = \|\Phi_{S^1} \Phi_T y + \Phi_{S^1} e\|_2 \\
\leq \delta_{N+K}\|x\|_2 + \sqrt{1 + \delta_N\|e\|_2}
\]
Hence \(S^1 \cap T \neq \emptyset\) is guaranteed if
\[
\delta_{N+K}\|x\|_2 + \sqrt{1 + \delta_N\|e\|_2} < \min \left\{1, \sqrt{\frac{N}{K}} \right\} \left[1 - \delta_K\|x\|_2 - \sqrt{1 + \delta_N\|e\|_2}\right]
\]
On the other hand,
\[
\|\Phi_{S^1} y\|_2 \geq \sqrt{\frac{L}{K}} \|\Phi_{S^1} y\|_2 \\
\geq \sqrt{\frac{L}{K}} \left[1 - \delta_K\|x\|_2 - \sqrt{1 + \delta_K\|e\|_2}\right]
\]
If $T^1 \cap T = \emptyset$, we have
\begin{equation}
\left\| \Phi^T_{x^2} \right\|_2^2 = \left\| \Phi^T_{x^2} \Phi^T \Phi^T_{x^2} + \Phi^T_{x^2} e \right\|_2^2 \\
\leq \delta_{L+K} \left\| x \right\|_2 + \sqrt{1 + \delta_L} \left\| e \right\|_2 \tag{38}
\end{equation}

Hence, given that $S^1 \cap T \neq \emptyset$, $T^1 \cap T \neq \emptyset$ is guaranteed, if
\[
\delta_{L+K} \left\| x \right\|_2 + \sqrt{1 + \delta_L} \left\| e \right\|_2 < \sqrt{\frac{L}{K} \left[ (1 - \delta_K) \left\| x \right\|_2 - \sqrt{1 + \delta_K} \left\| e \right\|_2 \right]}
\]

If $N \leq K$, then from Eq. (36) and Eq. (38), the required conditions can be stated as:
\begin{align*}
\frac{\left\| x \right\|_2}{\left\| e \right\|_2} &> \frac{\sqrt{K(1 + \delta_N) + N(1 + \delta_K)}}{\sqrt{N(1 - \delta_K) - \delta_{N+K} \sqrt{K} + \sqrt{N}}} \\
\frac{\left\| x \right\|_2}{\left\| e \right\|_2} &> \frac{\sqrt{K(1 + \delta_N) + \sqrt{L(1 + \delta_K)}}}{\sqrt{L(1 - \delta_K) - \delta_{N+K} \sqrt{K} + \sqrt{L}}}
\end{align*}

Since, $L \leq N$, and we have here $N \leq K$, due to the monotonic increasing property of $\delta$, both of the conditions above are satisfied if the following holds
\begin{equation}
\frac{\left\| x \right\|_2}{\left\| e \right\|_2} > \frac{\sqrt{K(1 + \delta_N) + \sqrt{N}}}{\sqrt{L(1 - \delta_K) - \delta_{N+K} \sqrt{K} + \sqrt{L}}} \tag{39}
\end{equation}

If $N > K$, then from Eq. (36) and Eq. (38), we have
\begin{align*}
\frac{\left\| x \right\|_2}{\left\| e \right\|_2} &> \frac{\sqrt{1 + \delta_N}}{\sqrt{1 - \delta_K} - \delta_{N+K}} \\
\frac{\left\| x \right\|_2}{\left\| e \right\|_2} &> \frac{\sqrt{K(1 + \delta_N) + \sqrt{L(1 + \delta_K)}}}{\sqrt{L(1 - \delta_K) - \delta_{N+K} \sqrt{K} + \sqrt{L}}}
\end{align*}

Now note that, $\sqrt{K + \sqrt{N}} \geq 2\sqrt{T}$, and $\sqrt{K + \sqrt{L}} \geq 2\sqrt{T}$, thus
\[
\frac{\sqrt{K + \sqrt{N}}}{\sqrt{L - \delta_{N+K} \sqrt{K} + \sqrt{L}}} \geq \frac{2}{1 - 2\delta_{N+K}} \geq \frac{2}{1 - \delta_K - \delta_{N+K}}
\]

which implies that Eq. (39) is also a sufficient condition for success at the first iteration when $N > K$. Thus, from the definition of $SNR$, it is easily seen that the preceding condition holds true if the following holds:
\begin{equation}
\sqrt{SNR} \geq \frac{(1 + \delta_{N+K}) \sqrt{N + \sqrt{K}}}{\sqrt{L - \sqrt{L + \sqrt{K}} \delta_{N+K}}} \tag{40}
\end{equation}

2) **Success at $(k + 1)^{th}$ iteration:** We assume that the previous $k$ iterations have been successful, that is at least one correct index was found in each of the last $k$ iterations. Denote $|T^k| = L^k$, $|T \cap T^k| = c_k$, so that $K > c_k \geq K$. Also, let $m_k := |S_k \cap T|$, $k \geq 1$, so that $m_i \geq 1$, $\forall i : 1 \leq i \leq K$. For the success of the $(k + 1)^{th}$ iteration, we require
\[
S^{k+1} \cap T \neq \emptyset \\
h^{k+1} \cap T \neq \emptyset
\]

**Condition to ensure $S^{k+1} \cap T \neq \emptyset$:** To find the condition under which $S^{k+1} \cap T \neq \emptyset$ is satisfied, we use the following notation:
\[
\begin{align*}
W^{k+1} := \arg \max_{S \subset \mathcal{C}} \| \Phi^T_{x^2} \Phi^T x \|_2^2 \\
\alpha_{max}^k := \min_{i \in W^{k+1}} (| \Phi_{i^k} r^k |)^2 \\
\beta_{max}^k := \max_{i \in W^{k+1}} (| \Phi_{i^k} r^k |)^2
\end{align*}
\]

Clearly, $S^{k+1} \cap T \neq \emptyset$, if $\beta_{max}^k > \alpha_{max}^k$. It is easy to see that
\[
\alpha_{max}^k \leq \frac{\| \Phi^T_{x^2} \Phi^T x \|_2^2}{N}
\]

Now, observe that
\[
\begin{align*}
r^k &= P_{T^k}^\perp y \\
&= P_{T^k}^\perp \Phi x + P_{T^k}^\perp e \\
&= \Phi_{T \cup T^k} x_{T \cup T^k}^\perp + P_{T^k}^\perp e
\end{align*}
\]

which follows from equations Eq. (18) and Eq. (19). Then, it follows that
\[
\begin{align*}
\alpha_{max}^k &\leq \frac{1}{N} \left( \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2 + \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T^k} \|_2 \right)^2 \\
\alpha_{max}^k &\leq \frac{1}{N} \left( \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2 + \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T^k} \|_2 \right)^2 \\
\alpha_{max}^k &\leq \frac{1}{N} \left( \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2 + \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T^k} \|_2 \right)^2 \\
\end{align*}
\]

Thus,\[
\alpha_{max}^k \leq \frac{1}{N} \left( \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2 + \| \Phi_{W^{k+1} \setminus T^k} \Phi_{T^k} \|_2 \right)^2 \tag{41}
\]

On the other hand,
\[
\begin{align*}
\beta_{max}^k &\geq \frac{1}{K - c_k} \| \Phi^T_{T \cup T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2^2 \\
&\geq \frac{1}{K - c_k} \| \Phi^T_{T \cup T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2^2 - \| \Phi^T_{T \cup T^k} \Phi^T_{T^k} P_{T^k}^\perp e \|_2^2 \\
&\geq \frac{1}{K - c_k} \left( \| \Phi^T_{T \cup T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2 - \| \Phi^T_{T \cup T^k} \Phi^T_{T^k} P_{T^k}^\perp e \|_2 \right)^2 \\
\end{align*}
\]

Note that
\[
\begin{align*}
\| \Phi^T_{T \cup T^k} \Phi_{T \cup T^k} x_{T \cup T^k} \|_2 &\geq (1 - \delta_{L^k + K - c_k}) \| x_{T \cup T^k} \|_2 \\
&> (1 - \delta_{L^k + K}) \| x_{T \cup T^k} \|_2
\end{align*}
\]

The last step follows because $L + K - c_k \leq (L - 1)(K - 1) + K < LK + N$
Thus,
\[ u_1^k \geq \frac{1}{\sqrt{K - c_k}} \left[ (1 - \delta_{LK+N}) \| x_{T^k,T^{k+1}}^* \|_2 - \sqrt{1 + \delta_{Lk+N}} \| e \|_2 \right] \]
(44)

On the other hand, from Eq. (27)
\[ v_1^k \leq \frac{1}{\sqrt{L}} \| \Phi_{S^{k+1}\setminus T} T^k \|_2 \]

Now, from Eq. (28), \( \forall i \in S^{k+1} \setminus T \)
\[ \| P_{T} \phi_i \|_2^2 \geq \left( 1 - \frac{\delta_{Lk+1}}{1 - \delta_{Lk+1}} \right)^2 \]

and,
\[ \| \Phi_{S^{k+1}\setminus T} T^k \|_2 \]
\[ = \left\| \Phi_{S^{k+1}\setminus T} T^k \right\|_2 \]
\[ = \left\| \Phi_{S^{k+1}\setminus T} T^k \right\|_2 \]
\[ \geq \frac{\delta_{Lk+N} \| x_{T^k,T^{k+1}}^* \|_2 + \sqrt{1 + \delta_{N+Lk}} \| e \|_2} \]

Thus,
\[ v_1^k \leq \frac{\delta_{N+Lk} \| x_{T^k,T^{k+1}}^* \|_2 + \sqrt{1 + \delta_{N+Lk}} \| e \|_2} \]
(45)

From Eq. (44), and Eq. (45), the condition \( h^{k+1} \cap T = \emptyset \) is ensured if
\[ \frac{1}{\sqrt{K - c_k}} \left[ (1 - \delta_{Lk+N}) \| x_{T^k,T^{k+1}}^* \|_2 - \sqrt{1 + \delta_{Lk+N}} \| e \|_2 \right] \geq \frac{\delta_{N+Lk} \| x_{T^k,T^{k+1}}^* \|_2 + \sqrt{1 + \delta_{N+Lk}} \| e \|_2} \]
\[ \sqrt{L \left( 1 - \left( \frac{\delta_{Lk+N}}{1 - \delta_{Lk+N}} \right)^2 \right)} \]

i.e., if,
\[ \frac{\| x_{T^k,T^{k+1}}^* \|_2}{\| e \|_2} \geq \frac{\sqrt{(1 + \gamma)(1 + 2\gamma)}}{\sqrt{L(1 - \gamma^2)} - \gamma \sqrt{K - c_k}} \]
(46)

where \( \gamma := \frac{\delta_{Lk+N}}{1 - \delta_{Lk+N}} \). So, from Eq. (44), and Eq. (46), one can write a sufficient condition for success at the \((k+1)^{th}\) iteration of gOMamOLS, given that iterations 1 through \( k \) are successful, as,
\[ \| x_{T^k,T^{k+1}}^* \|_2 \geq \min \left\{ \sqrt{\frac{N + \sqrt{K - c_k}}{\sqrt{K - c_k} + \sqrt{L(1 - \gamma^2)}}}, \frac{\sqrt{K - c_k} + \sqrt{L(1 - \gamma^2)}}{\sqrt{N - \gamma \sqrt{K - c_k}}} \right\} \]
Furthermore, \[ \|x_{T∪T^k}\|_2 ≥ \|x_T\|_2 \geq |T \setminus T^k| \min_{j \in T} |x_j| \geq \|x\|_2 / \kappa \cdot \sqrt{\frac{K-c_k}{K}} \]

Since, by assumption, \( c_k ≤ K-1 \), \( \|x_{T∪T^k}\|_2 ≥ \kappa \|x\|_2 / \sqrt{K} \), using the fact that \( c_k ≥ 0 \), allows the following to be another sufficient condition for success at \((k+1)\)th iteration, \[ \|x\|_2 ≤ \sqrt{\frac{K}{K-\kappa \sqrt{K}}} \]

Noting that \( \|\Phi x\|_2 ≤ \sqrt{1+\delta_K} \|x\|_2 < \sqrt{1+\delta_{LK+N}} \|x\|_2 \)

and \( N ≥ L \), and putting the expression of \( \gamma \), we find \[ \sqrt{(1+\gamma)(1+2\gamma)} \sqrt{\frac{N}{N-\gamma \sqrt{K}}} \]

Also note that \[ \sqrt{(1+\gamma)(1+2\gamma)} \sqrt{\frac{K}{K-(\gamma \sqrt{K})}} \]

Now, since \[ \sqrt{N-(\sqrt{N}+\sqrt{K}) \delta_{K+N}} \]

we see that the following is another sufficient condition for success at the \((k+1)\)th iteration, \[ \sqrt{\text{snr}} ≥ \frac{(\sqrt{N}+\sqrt{K})(1+\delta_{LK+N})\sqrt{K}}{\kappa \left( \sqrt{L(1-2\delta_{LK+N})-\delta_{LK+N} \sqrt{K}} \right)} \]

Since this condition does not depend on \( k \), condition in Eq. (50) is sufficient to guarantee success at all iterations \( k ≥ 2 \). Furthermore, from Eq. (49), we see that under the condition in Eq. (50) the first iteration is also guaranteed to succeed.

Thus, from Eq. (40) and Eq. (50), we see that the overall sufficient condition for perfect recovery in \( K \) iterations is given by Eq. (50).
Fig. 2: Recovery probability vs number of measurements

Fig. 3: Recovery probability vs number of measurements
Sparsity $K$

# of iterations for exact recovery

$T = 0$

(a) $T = 0$

Sparsity $K$

Mean runtime (seconds) per iteration

$T = 4$

(b) $T = 4$

Sparsity $K$

Mean runtime (seconds) per iteration

$T = 8$

(c) $T = 8$

Fig. 4: No. of iteration for exact recovery vs sparsity

Fig. 5: Mean runtime per iteration vs sparsity

mOLS ($L = 3$)
gOMP ($N = 3$)
m$^2$OLS ($N = 48, L = 3$)

mOLS ($L = 3$)
gOMP ($N = 3$)
m$^2$OLS ($N = 48, L = 3$)
VII. CONCLUSION AND FUTURE WORK

In this paper we have proposed a greedy algorithm for sparse signal recovery named m²OLS which preselects a few possibly “good” indices using gOMP and then uses an mOLS step to identify indices to be included in the estimated support set. We have carried out a theoretical analysis of the algorithm using RIP and have shown that if the sensing matrix satisfies the RIP condition \( \delta_{LK+N} < \frac{\sqrt{L}}{\sqrt{L}+\sqrt{K}} \), then the m²OLS algorithm is guaranteed to exactly recover a \( K \) sparse unknown vector, satisfying the measurement model, in exactly \( K \) steps. Also, we have extended our analysis to the noisy measurement setup and analytically provided bounds on the measurement SNR, and the unknown signal \( \text{MAR} \), under which recovery of the support of the unknown sparse vector is possible. Through numerical simulations, we have verified that introduction of m²OLS relative to gOMP, especially in correlated dictionaries. A future direction for this work could be to understand why exactly mOLS performs better than gOMP in presence of correlated dictionary and thereafter propose a further refinement of m²OLS based on that exploration.

APPENDIX A

PROOF OF LEMMAS

A. Lemma 1.1

Proof: At the \( k+1 \)th step of m²OLS the identification is performed by solving the following optimization problem

\[
h^{k+1} = \arg \min_{S \subset S^{k+1} : |S| = L} \sum_{i \in S} \| P_{T \cup \{i\}}^1 y \|^2_2
\]

Let us use, for any support \( T \subset \{1, 2, \ldots, n\} \), the following:

\[
W_T = R(\Phi_T)
\]

where \( R(A) \) denotes the range space or column space of matrix \( A \) i.e. the span of the columns of the matrix \( A \). Then, \( P_{T_k} \) is the projection operator on the space \( W_{T_k} \). Then,

\[
W_{T_k \cup \{i\}} = \text{span} (\phi_i, \ldots, \phi_i, \phi_i)
\]

where \( T_k = \{i_1, i_2, \ldots, i_k\} \). Then we can write

\[
W_{T_k \cup \{i\}} = \text{span} (\phi_i, \ldots, \phi_i) \oplus \text{span} (P_{T_k}^1 \phi_i)
\]

\[
=W_{T_k} \oplus \text{span} (P_{T_k}^1 \phi_i)
\]

where \( \oplus \) denotes direct sum. Hence from properties of projection operator it follows

\[
P_{T_k \cup \{i\}} y = P_{T_k} y + \langle P_{T_k}^1 \phi_i, y \rangle P_{T_k}^1 \phi_i
\]

\[
\Rightarrow P_{T_k \cup \{i\}}^1 y = P_{T_k}^1 y - \langle P_{T_k}^1 \phi_i, y \rangle P_{T_k}^1 \phi_i
\]

\[
\Rightarrow \| P_{T_k \cup \{i\}}^1 y \|^2_2 = \langle y, P_{T_k}^1 y - \langle P_{T_k}^1 \phi_i, y \rangle P_{T_k}^1 \phi_i \rangle
\]

\[
= \langle y, P_{T_k}^1 y - \langle P_{T_k}^1 \phi_i, y \rangle P_{T_k}^1 \phi_i \rangle
\]

\[
= \| P_{T_k}^1 y \|^2_2 - \langle \phi_i, r^k \rangle^2
\]

Thus \( \forall S \subset S^{k+1} : |S| = N \),

\[
\sum_{i \in S} \| P_{T_k \cup \{i\}}^1 y \|^2_2 = N \| P_{T_k}^1 y \|^2_2 - \sum_{i \in S} \| \langle \phi_i, r^k \rangle \|^2_2
\]

from which the second part of the claim is established.

To prove the first part, we prove the following claim, from which the result will follow immediately:

**Lemma 1.1.** Let \( \{a_1, \ldots, a_n\} \) be a set of \( n \) positive real numbers, where \( n \geq 1 \). Let \( |n| := \{1, 2, \ldots, n\} \). Let \( 1 \leq K \leq n \) be a positive number. Let

\[
h = \arg \max_{S : |S| = K, S \subset [n]} \sum_{i \in S} a_i
\]

and

\[
H = \arg \max_{S : |S| = K, S \subset [n]} \sum_{i \in S} a_i^2
\]

Then, \( \sum_{i \in h} a_i = \sum_{i \in H} a_i \).

Proof: We first show that \( \min_{i \in H} a_i \geq \max_{i \notin [n] \setminus H} a_i \). Assume that this is not true. Then, \( \exists i \in H, j \in [n] \setminus H \) such that \( a_i < a_j \) and consequently \( a_i^2 < a_j^2 \). Then, we can form a new set \( H' = H \cup \{j\} \setminus \{i\} \), with the sum of \( a_i^2 \)'s on the set as \( \sum_{k \in H'} a_k^2 = \sum_{k \in H} a_k^2 - a_i^2 + a_j^2 > \sum_{k \in H} a_k^2 \), which contradicts the fact that \( \sum_{k \in H} a_k^2 \) is maximal.

Thus, \( \sum_{i \in H} a_i \geq K \min_{i \in H} a_i \geq \sum_{j \in H} a_j \). Also, by assumption, \( \sum_{i \in H} a_i \leq \sum_{j \in H} a_j \). Thus the claim follows.

\[ \square \]
B. Lemma 4.2
Proof:
Let $N \leq K$. Then, according to the definition of $S^1$, for all $\Lambda \subset T$, such that $|\Lambda| = N$,

$$\|\Phi^T_{S^1} y\|_2^2 \geq \|\Phi^T_{\Lambda} y\|_2^2$$

Since there are $\binom{K}{N}$ such subsets of $T$, labelled, $\Lambda_i$, $1 \leq i \leq \binom{K}{N}$, we have

$$\frac{K}{N} \|\Phi^T_{S^1} y\|_2^2 \geq \sum_{i=1}^{\binom{K}{N}} \|\Phi^T_{\Lambda_i} y\|_2^2$$

(51)

Now, take any $j \in T$, and note that it appears in one of the $\Lambda_i$'s in exactly $\binom{K}{N-1}$ different ways. Thus, from the summation in Eq. (51), we find,

$$\frac{K}{N} \|\Phi^T_{S^1} y\|_2^2 \geq \frac{K-1}{N-1} \|\Phi^T_{\Lambda} y\|_2^2$$

$$\Rightarrow \|\Phi^T_{S^1} y\|_2^2 \geq \frac{N}{K} \|\Phi^T_{\Lambda} y\|_2^2$$

From which Eq. (5) follows.

Now, let $N > K$. Then, we can take any subset $\Sigma \subset \{1, 2, \cdots, n\}$, such that $|\Sigma| = N$ and $T \subset \Sigma$. Then, from definition, $\|\Phi^T_{S^1} y\|_2^2 \geq \|\Phi^T_{\Sigma} y\|_2^2 \geq \|\Phi^T_{\Lambda} y\|_2^2$ from which Eq. (4) follows.

To prove Eq. (5), first note that $T^1$ consists of indices corresponding to the largest $L$ absolute values $|\langle \phi_i, y \rangle|^2$, for $i \in S^1$. But since $S^1$ consists of indices corresponding to the $N$ largest absolute values $|\langle \phi_i, y \rangle|^2$ with $i \in \{1, 2, \cdots, n\} =: \mathcal{H}$, and since $N \geq L$, we have, $\min_{i \in \mathcal{H} \setminus T^1} |\langle \phi_i, y \rangle|^2 \geq \max_{i \in \mathcal{H} \setminus T^1} |\langle \phi_i, y \rangle|^2$. Since, $L \leq K$, for each $|\Gamma| = L$, such that $|\Gamma| = L$, we have

$$\|\Phi^T_{\Gamma} y\|_2^2 \geq \|\Phi^T_{\Lambda} y\|_2^2$$

Since there are $\binom{K}{L}$ such subsets, we can write

$$\frac{K}{L} \|\Phi^T_{\Gamma} y\|_2^2 \geq \sum_{|\Gamma| = L} \|\Phi^T_{\Gamma} y\|_2^2$$

Now any index $i \in T$ is contained in exactly $\binom{K-1}{L-1}$ of such $L$ cardinality subsets. Hence

$$\frac{K}{L} \|\Phi^T_{\Lambda} y\|_2^2 \geq \frac{K-1}{L-1} \|\Phi^T_{\Lambda} y\|_2^2$$

from which Eq. (5) follows.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, 2006.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” SIAM J. Sci. Comput., vol. 20, no. 1, pp. 33–61, 1998.

[3] G. Davis, S. Mallat, and M. Avellaneda, “Adaptive greedy approximations,” Constr. Approx., vol. 13, no. 1, pp. 57–98, 1997.

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

[5] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification,” Int. J. Control, vol. 50, no. 5, pp. 1873–1896, 1989.

[6] B. K. Natarajan, “Sparse approximate solutions to linear systems,” SIAM J. Sci. Comput., vol. 24, no. 2, pp. 227–234, 1995.

[7] J. Wang, S. Kwon, and B. Shim, “Generalized orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 60, no. 12, pp. 6202–6216, 2012.

[8] J. Wang and P. Li, “Recovery of Sparse Signals Using Multiple Orthogonal Least Squares,” arXiv preprint arXiv:1410.2505, Oct. 2014.

[9] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1094–1121, 2012.

[10] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” IEEE J. Sel. Topics Signal Process., vol. 4, no. 2, pp. 310–316, 2010.

[11] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 301–321, 2009.

[12] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” IEEE Trans. Inf. Theory, vol. 55, no. 5, pp. 2230–2249, 2009.

[13] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, “Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares,” IEEE Trans. Inf. Theory, vol. 59, no. 5, pp. 3158–3174, 2013.

[14] S. Satpathi, R. L. Das, and M. Chakraborty, “Improving the bound on the rip constant in generalized orthogonal matching pursuit,” IEEE Signal Process. Lett., vol. 20, no. 11, pp. 1074–1077, 2013.

[15] A. K. Fletcher and S. Rangan, “Orthogonal matching pursuit: A brownian motion analysis,” IEEE Trans. Signal Process., vol. 60, no. 3, pp. 1010–1021, 2012.