Successful Recovery Performance Guarantees of Noisy SOMP

Wei Zhang, Member, IEEE and Taejoon Kim, Senior Member, IEEE

Abstract

The simultaneous orthogonal matching pursuit (SOMP) is a popular, greedy approach for common support recovery of a row-sparse matrix. The support recovery guarantee of SOMP has been extensively studied under the noiseless scenario. Compared to the noiseless scenario, the performance analysis of noisy SOMP is still nascent, in which only the restricted isometry property (RIP)-based analysis has been studied. In this paper, we present the mutual incoherence property (MIP)-based study for performance analysis of noisy SOMP. Specifically, when noise is bounded, we provide the condition on which the exact support recovery is guaranteed in terms of the MIP. When noise is unbounded, we instead derive a bound on the successful recovery probability (SRP) that depends on the specific distribution of noise. Then we focus on the common case when noise is random Gaussian and show that the lower bound of SRP follows Tracy-Widom law distribution. The analysis reveals the number of measurements, noise level, the number of sparse vectors, and the value of MIP constant that are required to guarantee a predefined recovery performance. Theoretically, we show that the MIP constant of the measurement matrix must increase proportional to the noise standard deviation, and the number of sparse vectors needs to grow proportional to the noise variance. Finally, we extensively validate the derived analysis through numerical simulations.

Index Terms

Compressed sensing, simultaneous orthogonal matching pursuit (SOMP), successful recovery probability.

I. INTRODUCTION

The problem of sparse signal recovery appears in various applications of wireless communications and image processing [1]–[9], in which a common linear observation model is assumed

\[ Y = \Phi C + N, \]

W. Zhang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: weizhang@ntu.edu.sg).

T. Kim is with the Department of Electrical Engineering and Computer Science, The University of Kansas, KS 66045, USA (e-mail: taejoonkim@ku.edu).
where $Y \in \mathbb{R}^{M \times d}$ is the observation, $\Phi \in \mathbb{R}^{M \times N}$ with $M \ll N$ is the measurement matrix, $C \in \mathbb{R}^{N \times d}$ is the row-sparse matrix with only $L \ll M$ rows being non-zero, and $N \in \mathbb{R}^{M \times d}$ is the noise matrix. Without loss of generality, we assume that each column of $\Phi$ has unit $\ell_2$-norm. Unlike the single measurement vector (SMV) scenario, in which $d = 1$ in (1), the case with $d > 1$ is commonly referred to as multiple measurement vectors (MMV) model [2], [4], where the columns of $C$ share the same support.

Given the model in (1), the goal is to recover the support set of $C$ from the observations $Y$ and the known measurement matrix $\Phi$. Achieving a nearly optimal recovery performance is possible by employing low-complex, greedy algorithms [4], [8]–[11], such as the orthogonal matching pursuit (OMP) for SMV models ($d = 1$) [3], [8], [10], and simultaneous OMP (SOMP) for MMV models ($d > 1$) [4], [9], [11].

Both OMP and SOMP are iterative algorithms, in which one atom (one column of $\Phi$) is recovered per iteration and added to the recovered support set. To validate the efficacy of OMP and SOMP, the support recovery performances of OMP and SOMP have been analyzed for both noiseless and noisy scenarios [2], [7], [8], [12]–[20]. In the noiseless scenario, OMP can successfully recovery the support set in $L$ iterations when the order-$(L + 1)$ restricted isometry property (RIP) constant $\delta_{L+1}$ and the mutual incoherence property (MIP) constant $\mu$ satisfy, respectively, $\delta_{L+1} < 1/\sqrt{L + 1}$ [14] and $\mu < 1/(2L - 1)$ [16]. About the recovery performance of noisy OMP, sufficient conditions for exact support recovery have been studied in terms of MIP [8], [17] and RIP [15], [18]–[20] constants. For the noiseless SOMP, the exact support recovery is guaranteed when $\mu < 1/(2L - 1)$ [2], which is consistent with the case of noiseless OMP. With regard to noisy SOMP, the successful recovery error probability was analyzed under additive Gaussian noise based on the RIP analysis [7], which reveals that achieving near-zero error is possible when the signal power and the number of sparse vectors $d$ are sufficiently large.

In this paper, we focus on the performance guarantee of noisy SOMP based on the MIP analysis. Compared to the RIP, an advantage of characterizing the MIP constant is in its accountability in terms of being able to capture the property of maximal correlation between different columns of measurement matrix and amenability in terms of being able to evaluate effectively for fixed measurement matrix. However, the support recovery guarantee of noisy SOMP based on the MIP has not been thoroughly studied in the literature. For noisy OMP, MIP-based performance analysis has been studied previously [8], [17]. However, this concept cannot be directly extended to the SOMP due to the new dimension, the number of sparse vectors $d$, added to the problem. By analyzing and evaluating the performance of SOMP based on the MIP, in this paper, we will answer several key research questions: (i) what is the successful recovery probability (SRP) of SOMP under an arbitrary distribution of noise? (ii) what is the
desired condition of the row-sparse matrix $C$ to ensure predefined SRP performance? and (iii) what is the number of sparse vectors (i.e., $d$) to guarantee the exact recovery of the support set?

The established contributions of this paper are summarized as follows:

- We present two SOMP algorithms on the basis of its stopping rule. The first scheme is referred to as SOMP-sparsity (SOMPS), which stops the iteration when the required number of atoms is obtained. The second scheme terminates its iteration based on thresholding residual signal power per each iteration, which we refer to as SOMP-thresholding (SOMPT). When noise power is bounded, i.e., $\|N\|_2 \leq \epsilon$, we identify the conditions on which the exact support recovery is guaranteed for both SOMPS and SOMPT. Specifically, given the MIP constant $\mu < 1/(2L - 1)$, the support recovery of SOMPS is guaranteed when $\ell_2$-norm of the non-zero rows of $C$ is lower bounded by $2\epsilon/(1 - (2L - 1)\mu)$. Similarly, by setting an appropriate threshold value as a stopping criterion of SOMPT, the exact support recovery is also guaranteed on the same condition as SOMPS.

- When noise is unbounded, we derive lower bounds of SRP for both SOMPS and SOMPT algorithms, which depend on a specific distribution of noise. Then we focus on the practical scenario when noise is random Gaussian, and show that the lower bound of SRP follows the Tracy-Widom law distribution [21], [22]. From the identified SRP bound, we establish the required noise level, the number of sparse vectors $d$, the number of measurements, and the value of MIP constant. We reveal that the number of sparse vectors $d$ needs be proportional to the noise variance, and the difference between the MIP constants with and without noise is proportional to the noise standard deviation.

- In the simulation, we corroborate the theoretical results for both SOMPS and SOMPT by benchmarking them with numerical simulation results. The simulation results validate our analysis. We also illustrate the effect of other factors, such as sparsity $L$, the number of measurements $M$, and the number of sparse vectors $d$, on the recovery performance of SOMPS and SOMPT.

A. Paper Organization and Notations

The paper is organized as follows. In Section II, we introduce the support recovery problem, and present the SOMPS and SOMPT algorithms. In Section III, the performances of SOMPS and SOMPT are analyzed for the bounded noise. Then, in Section IV, we present the performance guarantee of SOMPS and SOMPT when the noise is unbounded, and the case with Guassian noise is discussed in detail. The simulation results and conclusions are presented in Section V and Section VI, respectively.

**Notation:** A bold lower case letter $a$ is a vector and a bold capital letter $A$ is a matrix. $A^T$, $A^{-1}$, $\|A\|_F$, $\|A\|_2$ and $\|A\|_\infty$ are, respectively, the transpose, inverse, Frobenius norm, $\ell_2$-norm and $\ell_\infty$-norm.
Algorithm 1 SOMPS

1: Input: The observations $Y$, the measurement matrix $\Phi$, sparsity level $L$.
2: Initialization: Support set $\hat{\Omega}^{(0)} = \emptyset$, residual matrix $R^{(0)} = Y$.
3: for $l = 1$ to $L$ do
4: Select the largest index $\eta = \arg\max_{i=1,\ldots,N} \|\Phi_{:,i}^T R^{(l-1)}\|_2$.
5: Update the support set: $\hat{\Omega}^{(l)} = \hat{\Omega}^{(l-1)} \cup \eta$.
6: Update the residual matrix: $R^{(l)} = Y - [\Phi_{:,\hat{\Omega}^{(l)}}]\hat{\Phi}_{:,\hat{\Omega}^{(l)}} Y$.
7: end for
8: Output: $\hat{\Omega}^{(L)}$.

II. SOMP Algorithm

Given the MMV observation model in (1), we recall that only $L$ rows of $C$ are non-zero. Here, we denote the row support set of $C$ as $\Omega \subset \{1, 2, \ldots, N\}$ with $|\Omega| = L$. To estimate the $\Omega$ from the observations $Y = \Phi C + N$ in (1), in this work, we mainly focus on two specific SOMP algorithms [2], [4], [23], [24], where Algorithm 1 corresponds to SOMPS and Algorithm 2 corresponds to SOMPT. When the row sparsity of $C$ (i.e., $L$), is known as a priori, the iteration of SOMPS can be terminated when the required number of atoms are selected [2], [4] as described in Algorithm 1. On the other hand, when the row sparsity of $C$ is unavailable as a priori, a threshold $\tau$ can be introduced to evaluate against the amount of power in the residue matrix $R^{(l)}$ at each iteration [2], [4]. Specifically, when $\|R^{(l)}\|_2$ is less than the threshold $\tau$ in Step 3 of Algorithm 2, the iteration of SOMPT is terminated.

As for each iteration of SOMPS and SOMPT, the active index determined in Step 4, i.e.,

$$\eta = \arg\max_{i=1,\ldots,N} \|\Phi_{:,i}^T R^{(l-1)}\|_2,$$

is added to the previously detected support set $\hat{\Omega}^{(l-1)}$ to form $\hat{\Omega}^{(l)}$ in Step 5 in both Algorithm 1 and Algorithm 2. It is crucial to recognize that the updated residue $R^{(l)}$ in Step 6 in Algorithm 1 and Algorithm 2 is orthogonal to the columns of $[\Phi]_{:,\hat{\Omega}^{(l)}}$, which are the already selected $l$ atoms. Precisely, we denote

of matrix $A$. $\|a\|_1$ and $\|a\|_2$ are $\ell_1$-norm and $\ell_2$-norm of vector $a$. $A^\dagger = (A^T A)^{-1} A^T$ denotes the pseudo inverse of a tall matrix $A$. $[A]_{:,i}$, $[A]_{i,:}$, $[A]_{i,j}$, and $[a]_i$ are, respectively, the $i$th column, $i$th row, $i$th row and $j$th column entry of $A$, and $i$th entry of vector $a$. $[A]_{:,S}$ denotes a sub-matrix of $A$ with the columns indexed by set $S$. We use $a \in A$ to denotes a column vector a chosen from the columns of $A$. $|S|$ denotes the cardinality of the set $S$. $\lambda_{\text{min}}(A)$ returns the minimal eigenvalue of a matrix $A$. 
\( \mathbf{P}^{(l)} = [\Phi_\Omega^{(l)} \Phi_\Omega^{(l)\dagger}] \in \mathbb{R}^{M \times M} \) and \( \mathbf{P} = \mathbf{I} - \mathbf{P}^{(l)} \), then the residual of \( l \)th iteration is expressed as
\[
\mathbf{R}^{(l)} = \mathbf{P}^{(l)} \mathbf{Y} = \mathbf{P}^{(l)} (\Phi \mathbf{C} + \mathbf{N}),
\]
where the columns of residue \( \mathbf{R}^{(l)} \) belong to the column subspace of matrix \( \mathbf{P}^{(l)}_\perp \). This is the reason why orthogonal matching pursuit is named.

It is worth noting that Algorithm 1 and Algorithm 2 successfully recover the support if and only if each active index determined in Step 4 is correct, i.e., \( \eta \in \Omega \). In particular, suppose the first \( l \) iterations of Algorithm 1 and Algorithm 2 select \( l \) correct atoms, the following remark gives the condition which guarantees the \((l + 1)\)th iteration selects the correct atom.

**Remark 1:** Assume the first \( l \) iterations of SOMPS in Algorithm 1 and SOMPT in Algorithm 2 select \( l \) correct atoms, i.e., \( \hat{\Omega}^{(l)} \subset \Omega \), and the termination condition is not satisfied at \((l + 1)\)th iteration\(^1\). Then, the \((l + 1)\)th iteration will select the correct atom when the following holds
\[
\max_{d \in [\Phi]_{\Omega^c, \Omega}} \left\| d^T \mathbf{R}^{(l)} \right\|_2 > \max_{d \in [\Phi]_{\Omega^c, \Omega^c}} \left\| d^T \mathbf{R}^{(l)} \right\|_2,
\]
where \( \Omega^c \subset \{1, \ldots, N\} \) with \( |\Omega^c| = N - L \) denotes the compliment of the set \( \Omega \).

Therefore, according to Remark 1, the condition for exact support recovery of SOMP is in fact the one that makes the inequality in (3) hold. This is the essential idea for the analysis for performance guarantee of SOMP in this work.

A. Motivations

The MIP constant has been utilized to measure the maximal coherence of different columns of the measurement matrix [25], [26]. The MIP constant of a matrix \( \Phi \) is defined as
\[
\mu = \max_{i \neq j} \left| \langle [\Phi]_{i,:}, [\Phi]_{j,:} \rangle \right|.
\]
In particular, quantifying the MIP constant of a matrix plays an important role in analyzing and solving various signal processing problems including the Grassmannian line/manifold packing [27], [28], support detection [3], [8], [10], and evaluating the focusing capabilities of imaging systems [29]. As for the support recovery, the MIP is also crucial to guarantee the successful sparse recovery [8], [16], [17], [30]. For example, it has been shown that OMP can successfully recovery the support set in the noiseless

\(^1\)The fact that termination condition is not satisfied at \((l + 1)\)th iteration means that \( l + 1 \leq L \) for Algorithm 1, and \( \left\| \mathbf{R}^{(l)} \right\|_2 \geq \tau \) for Algorithm 2.
scenario when $\mu < 1/(2L-1)$ [16], [30]. Alternatively, the RIP of measurement matrix is also an important characteristic for support recovery. When the RIP satisfies $\delta_{L+1} < 1/\sqrt{L+1}$ [14], the exact support recovery is guaranteed for OMP in noiseless case. However, it is challenging to calculate the RIP constant of a given measurement matrix. Moreover, the MIP constant value of measurement matrix can imply the RIP constant [31], i.e, $\delta_L \leq (L-1)\mu$. In addition, compared to RIP constant, the MIP constant has a lower bound [26], [32] expressed by the dimension of measurement matrix,

$$\mu \geq \sqrt{\frac{N-M}{M(N-1)}}, \tag{5}$$

where the minimal MIP constant can be achieved through the designing techniques for measurement matrix [27], [33], [34]. Though the MIP has the advantages as we mentioned above, as far as we know, the support recovery guarantee of noisy SOMP based on the MIP has not been thoroughly studied in the literature. In this work, our target is to provide the guarantee condition for successful support recovery of Algorithm 1 and Algorithm 2 in terms of MIP.

III. GUARANTEE OF RECOVERY UNDER BOUNDED NOISE

In this section, we discuss the recovery guarantee of SOMP when the noise is upper bounded, i.e., $\|N\|_2 \leq \epsilon$. First of all, we provide several lemmas as preliminaries.

**Lemma 1:** Denote $A \in \mathbb{R}^{m \times n}$, and $X \in \mathbb{R}^{n \times p}$, and let the general norm $\|\cdot\|_w$ defined as follows

$$\|X\|_w = \|X^T\|_{2,\infty} = \max_i \|\{X\}_{i,:}\|_2. \tag{6}$$

then, the following holds

$$\max_X \frac{\|AX\|_w}{\|X\|_w} = \|A\|_{\infty} = \max_i \sum_j |(A)_{i,j}|. \tag{7}$$

**Proof:** See Appendix A.

**Lemma 2:** For the model in (1), we define the constant

$$G = \max_{a \in \Phi} \left\|((\Phi)_{\cdot,\Omega} (\Phi)_{\cdot,\Omega})^{-1} (\Phi)_{\cdot,\Omega} a\right\|_1. \tag{8}$$

Then, the value of $G$ is upper bounded, i.e., $G \leq \frac{L\mu}{1-(L-1)\mu}$, where the constant $\mu$ is the MIP constant of matrix $\Phi$.

**Proof:** See Appendix B.
Algorithm 2 SOMPT

1: Input: The observations \(Y\), the measurement matrix \(\Phi\), the threshold \(\tau\).
2: Initialization: Support set \(\hat{\Omega}(0) = \emptyset\), residual matrix \(R^{(0)} = Y\), iteration number \(l = 1\).
3: \textbf{while} \(\|R^{(l-1)}\|_2 \geq \tau\) \textbf{do}
4: \quad Select the largest index \(\eta = \arg\max_{i=1,\ldots,N} \|\Phi^T_i R^{(l-1)}\|_2\).
5: \quad Update the support set: \(\hat{\Omega}(l) = \hat{\Omega}(l-1) \cup \eta\).
6: \quad Update the residual matrix: \(R^{(l)} = Y - [\Phi;\hat{\Omega}(l)][\Phi]^\dagger \hat{\Omega}(l) Y\), and \(l \leftarrow l + 1\).
7: \textbf{end while}
8: Output: \(\hat{\Omega}^{(l-1)}\).

Though the condition in (3) is necessary and sufficient for the correct selection of the \((l+1)\)th atom, it is not applicable to check whether the inequality in (3) holds, because it depends on priori information of support \(\Omega\). Therefore, to simplify the condition in (3), we plug the expression of \(R^{(l)}\) in (2) into (3), and define the following

\[
Q^{(l,1)} = \max_{d \in [\Phi], \Omega} \|d^T P^{(l)} \Phi C\|_2, \quad (9)
\]

\[
Q^{(l,2)} = \max_{d \in [\Phi], \Omega} \|d^T P^{(l)} \Phi C\|_2, \quad (10)
\]

\[
Z^{(l)} = \max_{d \in \Phi} \|d^T P^{(l)} N\|_2. \quad (11)
\]

Then, we have the following proposition for the relationship between \(Q^{(l,1)}\) in (9) and \(Q^{(l,2)}\) in (10).

**Proposition 1:** Let \(Q^{(l,2)}\) and \(Q^{(l,1)}\) be defined in (9) and (10), respectively, then following inequality holds,

\[
Q^{(l,2)} \leq G Q^{(l,1)}, \quad (12)
\]

where \(G\) is defined in (8).

**Proof:** See Appendix C

Under the conclusion in Proposition 1, we can transform the condition in (3) to the one with respect to \(Q^{(l,1)}\) and \(Z^{(l)}\).

**Lemma 3:** Under the definitions in (9), (10), and (11), if the first \(l\)th iterations select the correct atoms, then the sufficient condition for selecting the correct \((l+1)\)th atom is

\[
Q^{(l,1)} > 2\frac{1 - (L-1)\mu}{1 - (2L-1)\mu} Z^{(l)}. \quad (13)
\]
Thus, when the condition above holds, the right hand side of (3) is bounded as

\[ \max_{d \in [\Phi], \Omega} \| d^T R^{(l)} \|_2 > \max_{d \in [\Phi], \Omega^c} \| d^T R^{(l)} \|_2. \]

By using the triangle inequality and the expression of \( R^{(l)} \) defined in (2), the left hand side of (3) is

\[
\begin{align*}
&\max_{d \in [\Phi], \Omega} \| d^T R^{(l)} \|_2 \\
\geq &\max_{d \in [\Phi], \Omega} \left( \| d^T P^{(l)} \Phi C \|_2 - \| d^T P^{(l)} N \|_2 \right) \\
\geq &\max_{d \in [\Phi], \Omega} \left( \| d^T P^{(l)} \Phi C \|_2 - \max_{d \in [\Phi], \Omega^c} \| d^T P^{(l)} N \|_2 \right) \\
\geq &\max_{d \in [\Phi], \Omega} \left( \| d^T P^{(l)} \Phi C \|_2 - \max_{d \in \Phi} \| d^T P^{(l)} N \|_2 \right) \\
= &Q^{(l,1)} - Z^{(l)}.
\end{align*}
\]

(14)

The right hand side of (3) is bounded as

\[
\begin{align*}
&\max_{d \in [\Phi], \Omega^c} \| d^T R^{(l)} \|_2 \\
\leq &\max_{d \in [\Phi], \Omega^c} \left( \| d^T P^{(l)} \Phi C \|_2 + \| d^T P^{(l)} N \|_2 \right) \\
\leq &\max_{d \in [\Phi], \Omega^c} \left( \| d^T P^{(l)} \Phi C \|_2 + \max_{d \in [\Phi], \Omega} \| d^T P^{(l)} N \|_2 \right) \\
\leq &\max_{d \in [\Phi], \Omega^c} \left( \| d^T P^{(l)} \Phi C \|_2 + \max_{d \in \Phi} \| d^T P^{(l)} N \|_2 \right) \\
= &Q^{(l,2)} + Z^{(l)}.
\end{align*}
\]

(15)

Based on (14) and (15), in order to make the condition in (3) hold, it is sufficient that \(Q^{(l,1)} - Q^{(l,2)} > 2Z^{(l)}\) is true. Combining (12), i.e., \(Q^{(l,2)} \leq GQ^{(l,1)}\), this sufficient condition can be simplified as

\[ Q^{(l,1)} - Q^{(l,2)} \geq (1 - G)Q^{(l,1)} > 2Z^{(l)}. \]

(16)

If \(\mu < 1/(2L - 1)\) is true\(^2\), we plug the bound of \(G\) in Lemma 2 into (16), and obtain the following,

\[ Q^{(l,1)} > 2 \frac{1 - (L - 1)\mu}{1 - (2L - 1)\mu} Z^{(l)}. \]

Thus, when the condition above holds, the \((l + 1)\)th iteration selects the correct atom. This concludes the proof.

\[\]

Though the condition in (13) is about \(Q^{(l,1)}\) and \(Z^{(l)}\), the calculation of \(Q^{(l,1)}\) still requires to know \(\Omega\)

\(^2\)The condition is required to guarantee \(1 - G > 0\).
in advance. To further simplify the condition in (13), the following lemma provides the bound for $Q^{(l,1)}$.

**Lemma 4**: The $Q^{(l,1)}$ defined in (9) is lower bounded as

$$Q^{(l,1)} \geq (L - l)^{-1/2}(1 - (L - 1)\mu)\left\| \left[ \mathbf{C} \right]_{\hat{\Omega}_{c}^{(l),:}} \right\|_F,$$

where $\hat{\Omega}_{c}^{(l)}$ denotes the complement of set $\hat{\Omega}^{(l)}$ over the whole set $\Omega$ with $|\hat{\Omega}_{c}^{(l)}| = L - l$.

**Proof**: See Appendix D.

Based on the results of Lemma 3 and Lemma 4, we can establish the condition which guarantees the successful recovery of support set by using SOMPS in Algorithm 1.

**Theorem 1**: Given signal model in (1), suppose constant $\mu$ is the MIP constant of the measurement matrix $\Phi$ with $\mu < 1/(2L - 1)$. If

$$C_{\min} > \frac{2\|\mathbf{N}\|_2}{1 - (2L - 1)\mu},$$

where $C_{\min} = \min_{i \in \Omega} \| [\mathbf{C}]_{i,:} \|_2$, then SOMPS in Algorithm 1 can successfully recover the support set $\Omega$. In particular, when the noise matrix satisfies $\|\mathbf{N}\|_2 \leq \epsilon$, the condition in (18) can be expressed as

$$C_{\min} > \frac{2\epsilon}{1 - (2L - 1)\mu}.$$

**Proof**: According to Lemma 3 and Lemma 4, in order to guarantee the $(l + 1)$th iteration select the correct atom, the sufficient condition in (13) can be written as,

$$\frac{1 - (L - 1)\mu}{(L - l)^{1/2}}\left\| \left[ \mathbf{C} \right]_{\hat{\Omega}_{c}^{(l),:}} \right\|_F > \frac{2(1 - (L - 1)\mu)}{1 - (2L - 1)\mu} Z^{(l)},$$

which is simplified to

$$\left\| \left[ \mathbf{C} \right]_{\hat{\Omega}_{c}^{(l),:}} \right\|_F > \frac{2\sqrt{L - l}}{1 - (2L - 1)\mu} Z^{(l)}.$$

After standard manipulations, the sufficient condition which guarantees the inequality in (20) hold is

$$\left\| [\mathbf{C}]_{i,:} \right\|_2 > \frac{2Z^{(l)}}{1 - (2L - 1)\mu}, \forall i \in \Omega.$$
Moreover, it is noted from (11) that

\[
Z^{(l)} = \max_{d \in \Phi} \|d^T(I - P^{(l)})N\|_2 \\
\leq \max_d \|d^TN\|_2 \\
= \|N\|_2.
\]  

(22)

Combining (21) and (22), we can find that if

\[
\|\hat{C}_{i,:}\|_2 > \frac{2\|N\|_2}{1 - (2L - 1)\mu}, \forall i \in \Omega,
\]

and the Algorithm 1 terminates for \( l = L \), SOMPS successfully recovers the support \( \Omega \). This concludes the proof.

**Remark 2:** According to Theorem 1, the exact recovery of the support set is guaranteed for Algorithm 1 if \( C_{\text{min}} \) is lower bounded by the value given in (18). The lower bound is dependent on the noise level \( \|N\|_2 \), sparsity level \( L \), and the MIP constant \( \mu \). When the value of \( L \) and \( \mu \) are fixed, to guarantee the successful recovery of the support set, the value of \( C_{\text{min}} \) should be proportional to the value of \( \|N\|_2 \). In particular, when the number of sparse vector \( d = 1 \), the derived result in (18) is consistent with the case of OMP in [8], which validates that our derived result is more general.

**Remark 3:** Based on the results of Theorem 1, we can obtain the condition of \( \mu \) to guarantee the successful recovery of SOMPS as follows

\[
\mu < \frac{1 - 2\|N\|_2/C_{\text{min}}}{2L - 1}.
\]  

(23)

It is noted that when the noise exists, the value of MIP constant \( \mu \) should be reduced compared to the noiseless case such as \( \mu < 1/(2L - 1) \) [16], [30]. In other words, when the \( N = 0 \), the condition in (23) will be equivalent to the noiseless case. When the noise is bounded, i.e., \( \|N\|_2 \leq \epsilon \), the condition in (23) will become \( \mu < \frac{1 - 2\epsilon/C_{\text{min}}}{2L - 1} \).

In the following, we will discuss the conclusions of Theorem 1 under one type of special measurement matrix.

**Remark 4:** Since the value of \( \mu \) depends on the measurement matrix \( \Phi \), and recall that it has a lower bound in (5) expressed by its dimension,

\[
\mu \geq \sqrt{\frac{N - M}{M(N - 1)}}.
\]  

(24)
Therefore, we can combine (23) and (24), and obtain the condition for the performance guarantee of SOMPS when measurement matrix $\Phi$ achieves the minimal MIP constant. Specifically, when the following holds

$$\sqrt{\frac{N - M}{M(N - 1)}} < \frac{1 - 2\|N\|_2/C_{\min}}{2L - 1},$$

the successful recovery of SOMPS is guaranteed. We can simplify the formula above with respect to $M$ as follows

$$M > \frac{N}{\left(\frac{1 - 2\|N\|_2/C_{\min}}{2L - 1}\right)^2(N - 1) + 1}.$$  \hspace{1cm} (25)

It means that when the number of measurements $M$ is larger than the value provided in (25), the successful recovery of support can be guaranteed. Interestingly, it can be noted that when the noise is zero, i.e., $N = 0$, and $N \gg L$, the condition in (25) can be approximated as $M > (2L - 1)^2$. Compared to the well-known bound of number of measurements $M = O(L\log(N))$ [1], [10], the derived bound of number of measurements is independent of $N$ and more strict especially when the sparsity level $L$ is small.

As for SOMPT in Algorithm 2, it is different from SOMPS that has fixed iterations, instead, the stopping criterion of SOMPT is determined by the thresholding parameter $\tau$. To guarantee the successful recovery of support by using SOMPT, the following theorem provides the sufficient condition.

**Theorem 2:** Given the signal model in (1), we suppose constant $\mu$ is MIP constant of the measurement matrix $\Phi$ with $\mu < 1/(2L - 1)$, the noise is bounded with $\|N\|_2 \leq \epsilon$, and the thresholding parameter is $\tau = \epsilon$ for SOMPT in Algorithm 2. Then if the following condition holds,

$$C_{\min} > \frac{2\epsilon}{1 - (2L - 1)\mu},$$

Algorithm 2 can successfully recover the support set $\Omega$.

**Proof:** Given the results of Theorem 1, we only need to prove the Algorithm 2 will terminate correctly, i.e., $l = L$. In other words, it is sufficient to show that when $l < L$, we will have $\|R^{(l)}\|_2 > \epsilon$, and if
when $l = L$, we will have $\|R^{(l)}\|_2 \leq \epsilon$. First of all, when $l = L$, we have

$$\begin{align*}
\|R^{(L)}\|_2 &= \|P^{(L)}_\perp \Phi C + P^{(L)}_\perp N\|_2 \\
&= \|P^{(L)}_\perp N\|_2 \\
&= \|N\|_2 \\
&\leq \epsilon.
\end{align*}$$

(27)

For the case of $l < L$, we have

$$\begin{align*}
\|R^{(l)}\|_2 &= \|P^{(l)}_\perp \Phi C + P^{(l)}_\perp N\|_2 \\
&\geq \|P^{(l)}_\perp \Phi C\|_2 - \|P^{(l)}_\perp N\|_2 \\
&\geq \|P^{(l)}_\perp \Phi C\|_2 - \epsilon \\
&= \|P^{(l)}_\perp \Phi C\|_2 - \epsilon \\
&\geq \sqrt{\lambda_{\min} \left( [\Phi^T]^L_{\hat{\Omega}^{(l)}_1} P^{(l)}_\perp [\Phi^L_{\hat{\Omega}^{(l)}_1} \hat{\Omega}^{(l)}_2] \right) \|C\|_{\hat{\Omega}^{(l)}_1} \|_2 - \epsilon} \\
&\geq \sqrt{\lambda_{\min} \left( [\Phi^T]^L_{\hat{\Omega}^{(l)}_1} \hat{\Omega}^{(l)}_2 \right) \|C\|_{\hat{\Omega}^{(l)}_1} \|_2 - \epsilon} \\
&\geq \sqrt{1 - \frac{(L-1)\mu}{\|C\|_{\hat{\Omega}^{(l)}_1} \|_2 - \epsilon}}.
\end{align*}$$

(28)

where $(a)$ is from Lemma 5 in [8], and $(b)$ holds from the definition of $\mu$ in (4). According to the assumption in (26), the first part in (28) has the bound given by

$$\begin{align*}
\sqrt{1 - \frac{(L-1)\mu}{\|C\|_{\hat{\Omega}^{(l)}_1} \|_2 - \epsilon}} &\geq \sqrt{1 - \frac{(L-1)\mu}{\|C\|_{\hat{\Omega}^{(l)}_1} \|_2 - \epsilon}} \\
&> \frac{2\sqrt{1 - (L-1)\mu\epsilon}}{1 - (2L-1)\mu} \\
&> 2\epsilon.
\end{align*}$$

(29)

Therefore, after combining (28) and (29), we have $\|R^{(l)}\|_2 > \epsilon, \forall l < L$. This concludes the proof.

**Remark 5:** As we can see from Theorem 2, when the noise is bounded with $\|N\|_2 \leq \epsilon$ and the thresholding parameter in Algorithm 2 is $\tau = \epsilon$, then the SOMPT in Algorithm 2 can achieve the exact recovery of support set on the same condition as SOMPS in Algorithm 1, i.e., $C_{\min} > \frac{2\epsilon}{1 - (2L-1)\mu}$.

**Remark 6:** Similarly, we suppose the measurement matrix $\Phi$ achieves the lower bound of MIP constant in (24). According to the conclusions of Theorem 2 and Remark 4, if the number of measurements satisfies
the following,

\[ M > \frac{N}{\left(\frac{1-2\epsilon/C_{\text{min}}}{2L-1}\right)^2(N-1)+1}, \]

and the thresholding parameter is \( \tau = \epsilon \) in Algorithm 2, then SOMPT can successfully recover the support set \( \Omega \).

IV. GUARANTEE OF RECOVERY UNDER GENERAL UNBOUNDED NOISE

In Section III, we assume the noise is bounded with \( \|N\|_2 \leq \epsilon \). In this section, we will discuss the recovery performance of SOMPS and SOMPT in the case when the noise \( N \) in (1) is generally unbounded.

A. Analysis of SOMPS with unbounded noise

The following theorem shows the SRP of SOMPS in Algorithm 1 with general random noise.

Theorem 3: For the model in (1), we suppose \( \mu \) is MIP constant of the measurement matrix \( \Phi \) with \( \mu < 1/(2L-1) \). Define \( f_N(\cdot) \) as the CDF of \( \|N\|_2 \) given by

\[
\Pr(\|N\|_2 \leq x) = f_N(x), \forall x > 0, \quad (30)
\]

then the SRP of SOMPS in Algorithm 1 satisfies

\[
P_s \geq f_N \left( \frac{C_{\text{min}}(1-(2L-1)\mu)}{2} \right), \quad (31)
\]

where \( C_{\text{min}} = \min_{i \in \Omega} \|[C]_{i,:}\|_2 \).

Proof: When \( N \) is a random noise matrix, \( \|N\|_2 \) in (18) will also be a random variable. Thus, by using the results of Theorem 1, when the event in (18) happens, SOMPS in Algorithm 1 can guarantee the successful recovery of support. Precisely, suppose \( x > 0 \) is arbitrary, and we define the events \( \mathcal{X}_1 \) and \( \mathcal{X}_2 \) as follows

\[
\mathcal{X}_1 = \left\{ \|[C]_{i,:}\|_2 > \frac{2x}{1-(2L-1)\mu}, \forall i \in \Omega \right\}
\]
\[
\mathcal{X}_2 = \{ \|N\|_2 \leq x \}.
\]

Then, according to Theorem 1, the SRP of SOMPS is lower bounded by \( \Pr(\mathcal{X}_1 \cap \mathcal{X}_2) \). Moreover, since \( x \)
is arbitrary, we let \( x = C_{\min}(1 -(2L -1)\mu)/2 \), then
\[
\Pr(\mathcal{X}_1 \cap \mathcal{X}_2) = \Pr(\mathcal{X}_2) = f_{N}\left(\frac{C_{\min}(1 -(2L -1)\mu)}{2}\right).
\]
Thus, the SRP of Algorithm 1 satisfies
\[
P_s \geq f_{N}\left(\frac{C_{\min}(1 -(2L -1)\mu)}{2}\right),
\]
where \( C_{\min} = \min_{i \in \Omega} \|C_i\|_2 \). This concludes the proof.

Compared to the case of bounded noise in Theorem 1, the successful support recovery with unbounded noise will only be guaranteed from the probabilistic perspective. According to Theorem 3, the lower bound of the SRP is determined by the distribution of \( \|N\|_2 \), i.e., \( f_{N}(\cdot) \). In particular, by using the theory of random matrix, we can obtain the expression of \( f_{N}(\cdot) \) for a given distribution of N. As an example, the following proposition gives the expression of \( f_{N}(\cdot) \) when the entries in N are i.i.d. Gaussian \( \mathcal{N}(0, \sigma^2) \).

**Proposition 2:** Suppose random matrix \( N \in \mathbb{R}^{M \times d} \) has entries being i.i.d. according to \( \mathcal{N}(0, \sigma^2) \). The CDF of the largest singular value of N can be approximated as
\[
\Pr(\|N\|_2 \leq x) \approx TW_1\left(\frac{x^2 - \mu_{M,d}}{\sigma_{M,d}}\right),
\]
where the function \( TW_1(\cdot) \) is the CDF of Tracy-Widom law [21], [22], \( \mu_{M,d} = (M^{1/2} + d^{1/2})^2 \), and \( \sigma_{M,d} = (M^{1/2} + d^{1/2})(M^{-1/2} + d^{-1/2})^{1/3} \).

**Proof:** The detailed proof can be found in [21], [22], thus we omit it here.

Under the conclusions of Proposition 2 and Theorem 3, we can obtain the lower bound of SRP of SOMPS in Algorithm 1 when the entries in N are i.i.d. with \( \mathcal{N}(0, \sigma^2) \).

**Corollary 1:** Given signal model in (1), we suppose constant \( \mu \) is MIP constant of the measurement matrix \( \Phi \) with \( \mu < 1/(2L -1) \). When the entries in N are i.i.d. with \( \mathcal{N}(0, \sigma^2) \), then the SRP of Algorithm

\[TW_1(s) = \exp\left(-\frac{1}{2} \int_s^\infty q(x) + (x - s)q(x)dx\right),\]

where \( q(x) \) is the solution of Painlevé equation of type II:
\[q''(x) = xq(x) + 2q(x)^3, \quad q(x) \sim Ai(x), \quad x \to \infty,\]
where \( Ai(x) \) is the Airy function [21], [22]. To save the computational complexity, we admit the table lookup method [35] to obtain the value of \( TW_1(\cdot) \).
1 satisfies

\[ P_s \geq TW_1 \left( \frac{(1 - (2L - 1)\mu)^2 C_{\min}^2 - 4\sigma^2 \mu M_d}{4\sigma^2 \sigma_{M,d}} \right), \]  

(33)

where \( C_{\min} = \min_{i \in \Omega} \| [C]_{i,:} \|_2 \).

**Proof:** We plug the expression of \( f_N(\cdot) \) of (32) into Theorem 3, which obtains the results in (33). □

In some applications, we require the SRP of support set should be close to one. Then, which condition should we impose on the model in (1) to achieve this requirement? To address this question, the following theorem is an extension of Theorem 3, which gives the condition for \( C_{\min} \) to guarantee the required SRP of SOMPS in Algorithm 1.

**Theorem 4:** Given signal model in (1), we assume \( \mu \) is MIP constant of the measurement matrix \( \Phi \) with \( \mu < 1/(2L - 1) \). Let \( x_\delta \) and \( \delta \) be defined as

\[ \Pr(\|N\|_2 \leq x_\delta) \geq 1 - \delta, \]  

(34)

where \( \delta > 0 \) is a small number. Then, if

\[ C_{\min} > \frac{2x_\delta}{1 - (2L - 1)\mu}, \]  

(35)

then SOMPS in Algorithm 1 can successfully recover the support set \( \Omega \) with probability higher than \( 1 - \delta \).

**Proof:** The proof is similar as the procedures in Theorem 3, so we omit it here. □

According to the conclusion of Theorem 4, the row-sparse matrix \( C \) in (1) with \( C_{\min} \) satisfying the condition in (35) can guarantee the required SRP, i.e., \( 1 - \delta \). In particular, when the entries in \( N \) are i.i.d with distribution of \( \mathcal{N}(0, \sigma^2) \), we can obtain the performance guarantee of SOMPS in Algorithm 1 in the following corollary.

**Corollary 2:** Given signal model in (1), it is assumed that \( \mu < 1/(2L - 1) \) with \( \mu \) being MIP constant of \( \Phi \), and the noise \( N \) has i.i.d. entries according to \( \mathcal{N}(0, \sigma^2) \). We have that the noise part \( N \) satisfies the following

\[ \Pr \left( \|N\|_2 \leq \sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu M_d)\sigma^2} \right) \geq 1 - \delta, \]  

(36)

where \( \delta > 0 \) is a small number. Then, if

\[ C_{\min} > \frac{2\sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu M_d)\sigma^2}}{1 - (2L - 1)\mu}, \]  

(37)
or similarly
\[
\sigma < \frac{C_{\min}(1 - (2L - 1)\mu)}{2\sqrt{(TW^{-1}_1(1 - \delta)\sigma_{M,d} + \mu_{M,d})}},
\]
then Algorithm 1 can successfully recover the support set \( \Omega \) with probability higher than \( 1 - \delta \).

**Proof:** According to the expression of \( f_N(\cdot) \) in Proposition 2, we can check that (36) holds with 
\[
x_\delta = \sqrt{(TW^{-1}_1(1 - \delta)\sigma_{M,d} + \mu_{M,d})}\sigma^2 \text{ in (34).}
\]
Then, the condition in (37) holds from (35). This concludes the proof.

In the following, we discuss the performance guarantee of SOMPS when we let \( \delta \) close to zero.

**Remark 7:** From [36], when the entries in \( N \) are i.i.d with distribution of \( \mathcal{N}(0, \sigma^2) \) and 
\[
x_\delta \geq (\sqrt{M} + \sqrt{d})\sigma,
\]
we have that \( \Pr(\|N\|_2 \leq x_\delta) \approx 1 \) in (32). Thus, according to (35), the SRP is close to one when the following holds
\[
C_{\min} > \frac{2(\sqrt{M} + \sqrt{d})\sigma}{1 - (2L - 1)\mu},
\]
or equivalently
\[
\mu < \frac{1 - 2(\sqrt{M} + \sqrt{d})\sigma/C_{\min}}{2L - 1}.
\]

Compared to the sufficient condition \( \mu < 1/(2L - 1) \) for support recovery in noiseless case [16], the equivalent condition for \( \mu \) in (39) is more strict due to the noisy observations. Moreover, we calculate the difference between them,
\[
\frac{1}{2L - 1} - \frac{2(\sqrt{M} + \sqrt{d})\sigma/C_{\min}}{2L - 1} = \frac{2(\sqrt{M} + \sqrt{d})\sigma/C_{\min}}{2L - 1}.
\]

It is interesting to find that the difference is proportional to the noise standard derivation \( \sigma \).

**Remark 8:** In general, the value of \( C_{\min}^2 \) is related to the number sparse vectors \( d \). For simplicity, here, we assume \( C_{\min}^2 \) is proportional to the number of sparse vectors \( d \), i.e., \( C_{\min}^2 = dc_m \), where \( c_m \) is a constant. Then we manipulate (38) with respect to \( d \) as follows,
\[
d > \frac{M}{\left(\sqrt{c_m\frac{1-(2L-1)\mu}{2\sigma}} - 1\right)^2}.
\]
If the noise variance is small such that \( \sqrt{c_m\frac{1-(2L-1)\mu}{2\sigma}} \gg 1 \), we can write the expression above as
\[
d > \frac{4M\sigma^2}{(1 - (2L - 1)\mu)^2c_m}.
\]
It means that the required number of required sparse vectors is proportional to the noise variance $\sigma^2$.

B. Analysis of SOMPT with unbounded noise

In this subsection, we discuss the performance of SOMPT in Algorithm 2 when $N$ is unbounded noise. First of all, the following theorem shows the lower bound of SRP for SOMPT.

**Theorem 5:** Given the signal model in (1) and $\mu < 1/(2L - 1)$ with $\mu$ being MIP constant of $\Phi$, we suppose $N$ satisfies $\Pr(\|N\|\leq x) = f_N(x)$, and the thresholding parameter is

$$\tau = \frac{C_{\min}(1 - (2L - 1)\mu)}{2},$$

then the SRP of Algorithm 2 is lower bounded,

$$P_s \geq f_N \left(\frac{C_{\min}(1 - (2L - 1)\mu)}{2}\right), \quad (40)$$

where $C_{\min} = \min_{i \in \Omega} \|[C]_{i,:}\|_2$.

**Proof:** According to Theorem 2, SOMPT in Algorithm 2 can guarantee the support recovery if the following holds,

$$\|[C]_{i,:}\|_2 > \frac{2x}{1 - (2L - 1)\mu}, \forall i \in \Omega$$

$$\tau = x, \quad \|N\|\leq x. \quad (41)$$

Using the similar arguments of Theorem 3, we just let $x = C_{\min}(1 - (2L - 1)\mu)/2$, which concludes the proof.

Similarly, according to Theorem 5, for the specific noise, we can substitute the expression of $f_N(\cdot)$ into (40), and then obtain the lower bound of SRP of SOMPT in Algorithm 2. In particular, when the entries in $N$ are i.i.d. Guassian with $\mathcal{N}(0, \sigma^2)$, the SRP of SOMPT is bounded as the following corollary shows.

**Corollary 3:** Given signal model in (1), we assume $\mu$ is MIP constant of the measurement matrix $\Phi$ with $\mu < 1/(2L - 1)$. When the noise part $N$ are i.i.d. Gaussian $\mathcal{N}(0, \sigma^2)$, and the thresholding parameter of SOMPT in Algorithm 2 is

$$\tau = \frac{C_{\min}(1 - (2L - 1)\mu)}{2},$$

then the SRP of SOMPT Algorithm 2 satisfies

$$P_s \geq TW_1 \left(\frac{(1 - (2L - 1)\mu)^2C_{\min}^2 - 4\sigma^2\mu_{M,d}}{4\sigma^2\sigma_{M,d}}\right), \quad (42)$$
where \( C_{\text{min}} = \min_{i \in \Omega} \| [C]_{i,:} \|_2 \).

**Proof:** We plug the distribution function \( f_{\mathbf{N}} (\cdot) \) provided in (32) into (40), and this completes the proof. 

For recovery performance guarantee of SOMPT, we would require that the SRP is close to one. To address this, the following theorem provides the condition of \( C_{\text{min}} \) to guarantee the required SRP of SOMPT in Algorithm 2.

**Theorem 6:** Given signal model in (1) and \( \mu < 1/(2L - 1) \) with \( \mu \) being MIP constant of \( \Phi \), we assume \( x_\delta \) and \( \delta \) satisfy \( \Pr (\|\mathbf{N}\|_2 \leq x_\delta) > 1 - \delta \). If we have

\[
C_{\text{min}} > \frac{2x_\delta}{1 - (2L - 1)\mu}
\]

with \( C_{\text{min}} = \min_{i \in \Omega} \| [C]_{i,:} \|_2 \) and the thresholding parameter is \( \tau = x_\delta \), then SOMPT in Algorithm 2 can successfully recover the support set \( \Omega \) with probability higher than \( 1 - \delta \).

**Proof:** The proof is straightforward by using similar arguments as the proof in Theorem 5, so we omit it here. 

In particular, when the noise matrix \( \mathbf{N} \) has i.i.d. \( \mathcal{N}(0, \sigma^2) \) entries, the following corollary gives the performance guarantee of SOMPT under the Gaussian noise.

**Corollary 4:** Given signal model in (1), we assume constant \( \mu \) is MIP constant of the measurement matrix \( \Phi \) with \( \mu < 1/(2L - 1) \), and the entries in \( \mathbf{N} \) are i.i.d. according to \( \mathcal{N}(0, \sigma^2) \). We let thresholding parameter of SOMPT in Algorithm 2 be

\[
\tau = \sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu_{M,d})\sigma^2}.
\]

Then, if the following holds,

\[
C_{\text{min}} > \frac{2\sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu_{M,d})\sigma^2}}{1 - (2L - 1)\mu},
\]

or similarly

\[
\sigma < \frac{C_{\text{min}}(1 - (2L - 1)\mu)}{2\sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu_{M,d})\sigma^2}},
\]

then SOMPT in Algorithm 2 can successfully recover the support set \( \Omega \) with probability higher than \( 1 - \delta \).

**Proof:** We substitute \( f_{\mathbf{N}} (\cdot) \) of Proposition 2 into Theorem 6 with \( x_\delta = \sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu_{M,d})\sigma^2} \), and obtain the result in (44), which concludes the proof.
Remark 9: According to the results in Corollary 4, if the thresholding parameter of the SOMPT is given in (43), we can obtain the conditions of $C_{\text{min}}$ and $\sigma$ on which SOMPT can guarantee the required SRP of support set. Moreover, the conditions of $\mu$ and $d$ for performance guarantee can also be obtained in the similar way as in Remark 7 and Remark 8. Specifically, if the following

$$\mu < \frac{1 - 2(\sqrt{M} + \sqrt{d})\sigma/C_{\text{min}}}{2L - 1}$$

or

$$d > \frac{4M\sigma^2}{(1 - (2L - 1)\mu)^2c_m}$$

holds, SOMPT can achieve SRP close to one.

V. SIMULATION RESULTS

In this section, we validate the main results of SOMPS and SOMPT in Section III and Section IV.

A. Verify the analyzed SRPs of SOMPS and SOMPT

1) SRP with bounded noise: In Fig. 1, we validate the analysis of SOMPS and SOMPT when the noise is bounded. The simulated parameters are $M = 50, N = 100, L = 4, d = 4, C_{\text{min}} = 2$. Here, we let the noise $\|N\|_2 \leq \epsilon$. The simulation curves are plotted by averaging 1000 trials. According to Theorem 1 and Theorem 2, when the $\epsilon$ satisfies $\epsilon < C_{\text{min}}(1 - (2L - 1)\mu)/2$, SOMPS and SOMPT can successfully recover the support set. Thus, the theory curve is plotted through $\epsilon = C_{\text{min}}(1 - (2L - 1)\mu)/2$. As we can
Fig. 2: SRP of SOMP under Gaussian noise ($M = 50, N = 100, L = 4, d = 4, C_{\text{min}} = 2, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$).

Fig. 3: SRP of SOMPS (Algorithm 1) with different $L$ ($M = 160, N = 200, C_{\text{min}} = 2, d = 4, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$).

see from Fig. 1, when $\epsilon < C_{\text{min}}(1 - (2L - 1)\mu)/2$, SOMPS and SOMPT can both guarantee the successful recovery of the support set. This verifies that simulation results are consistent with our analysis. Moreover, it is interesting to find that SOMPS and SOMPT can achieve the similar SRP. The performance of SOMPS is slightly better than SOMPT because of the priori information of the sparsity level.

2) SRP with Gaussian noise: In Fig. 2, we validate our analysis when the noise is random Gaussian with $[N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$. The simulated parameters are $M = 50, N = 100, L = 4, d = 4, \delta = 0.05, C_{\text{min}} = 2$. In Fig. 2, we plot the simulated SRPs of SOMPS and SOMPT, and compare them with the
Fig. 4: SRP of SOMPT (Algorithm 2) with different \( L \) \((M = 160, N = 200, C_{\text{min}} = 2, d = 4, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2))\).

Theory bound provided in (37) of Corollary 2 and (44) of Corollary 4. Theoretically, when the noise level \( \sigma \) satisfies the condition in (37) and (44) such that

\[
\sigma < \frac{C_{\min}(1 - (2L - 1)\mu)}{2\sqrt{(TW_1^{-1}(1 - \delta)\sigma_{M,d} + \mu_{M,d})}},
\]

the recovery probability is at least \( 1 - \delta \). As can be seen from Fig. 2, when the noise level \( \sigma \) is lower than the theory bound, both SOMPS and SOMPT can achieve SRP of near probability one, which validates our analysis. Similarly, the recovery performance of SOMPS is slightly better than SOMPT from the same reason as Fig. 1.

B. Illustrate the effects of factors on the guarantee for SRPs of SOMPS and SOMPT

1) The sparsity level \( L \): In Fig. 3 and Fig. 4, we illustrate our analysis under different levels of sparsity. The simulated parameters are \( M = 160, N = 200, d = 4, C_{\text{min}} = 2, \) and \( L = 1, 2, \ldots, 10 \). The theory bounds in Fig. 3 and Fig. 4 are plotted by letting \( \delta = 0.05 \) in (37) of Corollary 2 and (44) of Corollary 4. For the simulation results, we use the colormap to define the SRP in the range of \([0, 1]\), where the black color means SRP = 0, and white color means SRP = 1. According to (37) of Corollary 2 and (44) of Corollary 4, the required noise level to guarantee the recovery performance of SOMPS and SOMPT will decrease as sparsity \( L \) increases, which is shown by the red lines of theory bounds in Fig. 3 and Fig. 4. For the simulation results, it can be found from Fig. 3 and Fig. 4 that as the sparsity level increases, to guarantee exact recovery, the required noise level will decrease, which is consistent with our analysis.
2) The number of sparse vectors $d$: In Fig. 5-8, we illustrate our analysis with different values of $d$. For a complete comparison, we show the following two cases:

- For the first case, we assume $C_{\text{min}}^2 = dc_m$.
- For the second case, we have $C_{\text{min}}$ is a constant for different values of $d$.

In Fig. 5 and Fig. 6, we show the first case of $C_{\text{min}}^2 = dc_m$, where the simulation parameters are $M = 100, N = 200, L = 4, c_m = 1, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$. As can be seen in Fig. 5 and Fig. 6, the simulated noise level which guarantees the SRPs of SOMPS and SOMPT becomes higher as $d$ increases. This
Fig. 7: SRP of SOMPS (Algorithm 1) with different $d$ ($M = 100, N = 200, L = 4, C_{\text{min}} = 2, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$).

Fig. 8: SRP of SOMPT (Algorithm 2) with different $d$ ($M = 100, N = 200, L = 4, C_{\text{min}} = 2, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$).

validates the analyzed results in Remark 8 and Remark 9, where the number of vectors should satisfy $d > \frac{4M\sigma^2}{(1-(2L-1)\mu)^2 C_{\text{min}}}$ to guarantee the performance of support recovery.

In Fig. 7 and Fig. 8, we show the SRPs of SOMPS and SOPM in the second case when $C_{\text{min}}$ is a constant for different values of $d$. The simulation parameters are $M = 100, N = 200, L = 4, [N]_{i,j} \sim \mathcal{N}(0, \sigma^2)$. It is verified that the value of $\frac{C_{\text{min}}(1-(2L-1)\mu)}{2\sqrt{(TW_1)^2(1-\eta)\sigma_{T,d}+\mu_{T,d}}}$ in (37) and (44) will decrease as $d$ increases, which is shown in the theory bounds in Fig. 7 and Fig. 8. For the simulation results, in Fig. 7 and Fig. 8, it is found that the simulated noise level which guarantees the SRPs of SOMPS and SOMPT also decreases.
as $d$ increases, which is consistent with the trend of theory bound of our analysis.

3) The number of measurements $M$: In Fig. 9 and Fig. 10, we illustrate our analysis under different number of measurements. The simulated parameters are $L = 4, N = 200, d = 4, C_{\min} = 10$, and $M = 50, 60, \ldots, 160$. The noise is bounded $\|N\|_2 \leq \epsilon$ with $\epsilon = 0.01$. According to Remark 4 and Remark 6, the required number of measurements to guarantee the recovery performance of SOMPS and SOMPT will increase as the sparsity level $L$ increases, which is shown by the red lines of theory bounds. For the simulations in Fig. 9 and Fig. 10, as the sparsity level $L$ increases, the required number of measurements will also increase in the similar trend as the theory bound. Interestingly, we can find in Fig. 9 and Fig.
that the theory bound is tight when the sparsity level is small, which is consistent with our discussion in Remark 4 and Remark 6.

VI. CONCLUSION

In this paper, we have analyzed the performance guarantee of SOMP with respect to the MIP constant of measurement matrix when the observations are corrupted by noise. Specifically, when the noise is bounded, we have showed that if the $\ell_2$-norm of non-zero rows of the row-sparse matrix is lower bounded, the successful recovery of support set is guaranteed. On the other hand, when the noise is unbounded, the closed-form lower bound of the SRP of support set has been derived. Based on the derived lower bound, we have showed the conditions for the number of measurements, noise level, the number of sparse vectors, and MIP constant, on which the required SRP can be achieved. Finally, the simulation results validated our derived analysis.

APPENDIX A

PROOF OF LEMMA 1

Proof: According to the definition in Lemma 1, we have

$$\max_{\mathbf{X}} \frac{\|\mathbf{AX}\|_w}{\|\mathbf{X}\|_w} = \max_{\|\mathbf{X}\|_w=1} \|\mathbf{AX}\|_w$$

$$= \max_{\|\mathbf{X}\|_w=1} \max_i \left( \sum_{j=1}^p \left( \sum_{k=1}^n [\mathbf{A}]_{i,k} [\mathbf{X}]_{k,j} \right)^2 \right)^{1/2}$$

$$= \max_i \max_{\|\mathbf{X}\|_w=1} \left( \sum_{j=1}^p \left( \sum_{k=1}^n [\mathbf{A}]_{i,k} [\mathbf{X}]_{k,j} \right)^2 \right)^{1/2}$$

$$= \max_i \max_{\|\mathbf{X}\|_w=1} \left\| \mathbf{X}^T [\mathbf{A}]^T_{i,:} \right\|_2$$

$$= \max_i \max_{\|\mathbf{X}\|_w=1} \left\| \mathbf{X}^T [\mathbf{A}]^T_{i,:} \right\|_2.$$  (46)

It is worth noting that the following holds

$$\max_{\|\mathbf{X}\|_{k,:} \|_2=1} \left\| \mathbf{X}^T [\mathbf{A}]^T_{i,:} \right\|_2 = \sum_{k=1}^n |[\mathbf{A}]_{i,k}|,$$  (47)

where the maximum is achieved if the condition below is satisfied,

$$[\mathbf{X}]_{k,:} = \frac{[\mathbf{A}]_{i,k}}{|[\mathbf{A}]_{i,k}|} v^T, \forall k,$$
with \( v \in \mathbb{R}^{p \times 1} \) being any \( \ell_2 \)-norm vector, i.e., \( \|v\|_2 = 1 \). Then, plugging (47) into (46) gives

\[
\max_X \frac{\|AX\|_w}{\|X\|_w} = \max_i \sum_{k=1}^n |A_{i,k}|
\]

\[
= \|A\|_{\infty}.
\]

This concludes the proof.

\[\square\]

**APPENDIX B**

**PROOF OF LEMMA 2**

**Proof:** The proof is based on the arguments of Theorem 3.5 of [30], for the compactness of this paper, we provide the details of the proof as follows. First of all, we have

\[
G = \max_{a \in [\Phi],_{\Omega^c}} \left\| \left( [\Phi]_{\Omega}^T [\Phi]_{\Omega} \right)^{-1} [\Phi]_{\Omega}^T a \right\|_1
\]

\[
= \left\| \left( [\Phi]_{\Omega}^T [\Phi]_{\Omega} \right)^{-1} \right\|_\infty \max_{a \in [\Phi],_{\Omega^c}} \left\| [\Phi]_{\Omega}^T a \right\|_1. \tag{48}
\]

According to the definition of \( \mu \) in (4), the second part of (48) satisfies the following,

\[
\max_{a \in [\Phi],_{\Omega^c}} \left\| [\Phi]_{\Omega}^T a \right\|_1 \leq L \mu. \tag{49}
\]

For the first part of (48), we can express \([\Phi]_{\Omega}^T [\Phi]_{\Omega} = I + E\) with \( E = [\Phi]_{\Omega}^T [\Phi]_{\Omega} - I \). Then, if \( \|E\|_\infty < 1 \),

\[
\left\| (I + E)^{-1} \right\|_\infty \overset{(a)}{=} \left\| \sum_{k=0}^{\infty} (-E)^k \right\|_\infty
\]

\[
\leq \sum_{k=0}^{\infty} \|E\|^k_\infty
\]

\[
\overset{(b)}{=} \sum_{k=0}^{\infty} \|E\|^k_\infty = \frac{1}{1 - \|E\|_\infty}
\]

\[
\overset{(c)}{=} \frac{1}{1 - (L - 1)\mu}, \tag{50}
\]

where equality \((a)\) holds from the fact that Neumann series \( \sum_{k=0}^{\infty} (-E)^k \) converge to \((I + E)^{-1}\) when \( \|E\|_\infty < 1 \) [37], the inequality \((b)\) holds from the Cauchy-Schwarz inequality, and the inequality \((c)\) is from the definition of \( \mu \) in (4). Then, we can combine (49) and (50), and this concludes the proof.

\[\square\]
APPENDIX C

PROOF OF PROPOSITION 1

Proof: Considering the definitions of $Q^{(l,1)}$ and $Q^{(l,2)}$, we can rewrite them as following

\[ Q^{(l,1)} = \| [\Phi_{l}^{T} \Phi C]_{w} \], \]
\[ Q^{(l,2)} = \| [\Phi_{l}^{T} (I - P^{(l)}) \Phi C]_{w} \].

Therefore, by denoting $S^{(l)} = (I - P^{(l)}) \Phi C$, it has $Q^{(l,2)}_{\sqrt{G}} = \| [\Phi_{l}^{T} \Phi C]_{w} \].$ Moreover, the following is true

\[ S^{(l)} = [\Phi_{l}]_{\omega} (\Phi_{l})^{-1} \| [\Phi_{l}^{T} \Phi C]_{w} \],

which is because the columns of $S^{(l)}$ belong to the subspace spanned by the columns of $[\Phi_{l}]_{\omega}$. Then, we can express the ratio between $Q^{(l,2)}$ and $Q^{(l,1)}$ as

\[ \frac{\| [\Phi_{l}^{T} \Phi C]_{w} \|}{\| [\Phi_{l}^{T} \Phi C]_{w} \|} \leq \| [\Phi_{l}^{T} \Phi C]_{w} \| \leq \| [\Phi_{l}^{T} \Phi C]_{w} \| \]

where the inequality holds from Lemma 1 and $G$ is defined in (8). This concludes the proof.

APPENDIX D

PROOF OF PROPOSITION 4

Proof: Based on the definition of $Q^{(l,1)}$ in (9), we have

\[ Q^{(l,1)} = \| [\Phi_{l}^{T} \Phi C]_{w} \|

where the inequality (a) holds from $P^{(l)} = I - [\Phi_{l}]_{\omega} [\Phi_{l}]_{\omega}^{T}$, the inequality (b) holds from the fact that $\| A \|_{w} \geq \sqrt{1/m} \| A \|_{F}$ for any $A \in \mathbb{R}^{m \times n}$, and the inequality (c) is from the definition of $\mu$ in (4). This
concludes the proof.

REFERENCES

[1] D. L. Donoho, “Compressed sensing.” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[2] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. part i: Greedy pursuit,” Signal Process., vol. 86, no. 3, pp. 572 – 588, 2006.
[3] Q. Duan, T. Kim, L. Dai, and E. Perrins, “Coherence statistics of structured random ensembles and support detection bounds for OMP,” IEEE Signal Process. Lett., vol. 26, no. 11, pp. 1638–1642, 2019.
[4] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” IEEE Trans. Signal Process., vol. 54, no. 12, pp. 4634–4643, Dec 2006.
[5] W. Zhang, T. Kim, D. J. Love, and E. Perrins, “Leveraging the restricted isometry property: Improved low-rank subspace decomposition for hybrid millimeter-wave systems,” IEEE Trans. Commun., vol. 66, no. 11, pp. 5814–5827, Nov 2018.
[6] Y. Chi, Y. M. Lu, and Y. Chen, “Nonconvex optimization meets low-rank matrix factorization: An overview,” IEEE Trans. Signal Process., vol. 67, no. 20, pp. 5239–5269, 2019.
[7] J. Determe, J. Louveaux, L. Jacques, and F. Horlin, “On the noise robustness of simultaneous orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 65, no. 4, pp. 864–875, 2017.
[8] T. T. Cai and L. Wang, “Orthogonal matching pursuit for sparse signal recovery with noise,” IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4680–4688, July 2011.
[9] E. van den Berg and M. P. Friedlander, “Theoretical and empirical results for recovery from multiple measurements,” IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2516–2527, May 2010.
[10] 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, Dec 2007.
[11] K. Lee, Y. Bresler, and M. Junge, “Subspace methods for joint sparse recovery,” IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3613–3641, June 2012.
[12] Q. Mo and Y. Shen, “A remark on the restricted isometry property in orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3654–3656, 2012.
[13] J. Wang and B. Shim, “On the recovery limit of sparse signals using orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 60, no. 9, pp. 4973–4976, 2012.
[14] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, “A sharp condition for exact support recovery with orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 65, no. 6, pp. 1370–1382, 2017.
[15] C. Liu, Y. Fang, and J. Liu, “Some new results about sufficient conditions for exact support recovery of sparse signals via orthogonal matching pursuit,” IEEE Trans. Signal Process., vol. 65, no. 17, pp. 4511–4524, 2017.
[16] T. T. Cai, L. Wang, and G. Xu, “Stable recovery of sparse signals and an oracle inequality,” IEEE Trans. Inf. Theory, vol. 56, no. 7, pp. 3516–3522, 2010.
[17] E. Miandji, M. Emadi, J. Unger, and E. Afshari, “On probability of support recovery for orthogonal matching pursuit using mutual coherence,” IEEE Signal Process. Lett., vol. 24, no. 11, pp. 1646–1650, 2017.
[18] L. Chang and J. Wu, “An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 60, no. 9, pp. 5702–5715, 2014.
[19] H. Ge, L. Wang, J. Wen, and J. Xian, “An RIP condition for exact support recovery with covariance-assisted matching pursuit,” IEEE Signal Process. Lett., vol. 26, no. 3, pp. 520–524, 2019.
[20] X. Cai, Z. Zhou, Y. Yang, and Y. Wang, “Improved sufficient conditions for support recovery of sparse signals via orthogonal matching pursuit,” *IEEE Access*, vol. 6, pp. 30,437–30,443, 2018.

[21] K. Johansson, “Shape fluctuations and random matrices,” *Commun. Math. Phys.*, vol. 209, no. 2, pp. 437–476, Feb 2000.

[22] I. M. Johnstone et al., “On the distribution of the largest eigenvalue in principal components analysis,” *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.

[23] S. F. Cotter, B. D. Rao, Kjersti Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, 2005.

[24] D. Leviatan and V. N. Temlyakov, “Simultaneous approximation by greedy algorithms,” *Adv. Comput. Math.*, vol. 25, no. 1-3, pp. 73–90, 2006.

[25] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization,” *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, 2003.

[26] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[27] T. Strohmer and R. W. Heath Jr, “Grassmannian frames with applications to coding and communication,” *Appl. Comp. Harmonic Anal.*, vol. 14, no. 3, pp. 257–275, 2003.

[28] D. Love, R. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[29] R. Obermeier and J. A. Martinez-Lorenzo, “Sensing matrix design via mutual coherence minimization for electromagnetic compressive imaging applications,” *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 217–229, 2017.

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

[31] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. Springer New York, 2013.

[32] L. Welch, “Lower bounds on the maximum cross correlation of signals (corresp.),” *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, 1974.

[33] M. A. Herman and T. Strohmer, “High-resolution radar via compressed sensing,” *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, 2009.

[34] M. Elad, “Optimized projections for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec 2007.

[35] B. Nadler, “On the distribution of the ratio of the largest eigenvalue to the trace of a wishart matrix,” *J. Multivar. Anal.*, vol. 102, no. 2, pp. 363–371, 2011.

[36] M. Rudelson and R. Vershynin, “Non-asymptotic theory of random matrices: extreme singular values,” in *Proc. Int. Congr. Math.*, 2010.

[37] E. Kreyszig, *Introductory functional analysis with applications*. Wiley New York, 1978, vol. 1.