ROBUST RECOVERY OF JOINT SPARSE SIGNALS VIA SIMULTANEOUS ORTHOGONAL MATCHING PURSUIT

*Yuxuan Zhang and Jian Wang**

School of Data Science, Fudan University, China

ABSTRACT

Simultaneous orthogonal matching pursuit (SOMP) is a classical algorithm for solving multiple measurement vectors (MMV) problems. In this paper, we analyze the theoretical performance of the SOMP algorithm using the restricted isometry property (RIP). In particular, we show that SOMP can robustly recover any joint K-sparse signal from its noisy measurements if the sensing matrix satisfies the RIP with isometry constant upper bounded by an absolute constant. Our result significantly improves upon some exiting results that require the isometry constant to be at least inversely proportional to $\sqrt{K}$.

Index Terms— Restricted isometry property (RIP), multiple measurement vectors (MMV), joint sparsity.

1. INTRODUCTION

Recently, recovery of joint sparse signals from multiple measurement vectors (MMV) has received much attention in array signal processing [1]. The main goal is to recover a collection of high-dimensional joint K-sparse signals $\mathbf{X} \in \mathbb{R}^{n\times r}$ from the compressed noisy measurements

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{W}, \quad (1)$$

where $\mathbf{\Phi} \in \mathbb{R}^{m\times n}$ is the measurement matrix and $\mathbf{W} \in \mathbb{R}^{m\times r}$ is noise. In general, a collection of high-dimensional signals $\mathbf{X} \in \mathbb{R}^{n\times r}$ is called joint K-sparse if $\mathbf{X}$ has at most K non-zero rows. To recover the signal $\mathbf{X}$, many methods have been put forward by exploiting the joint sparsity [2–4]. For example, simultaneous orthogonal matching pursuit (SOMP) [5] sequentially selects reliable row-support indices according to the correlation between the measurement matrix and residual matrix. Compressive multiple signal classification (CS-MUSIC) [6] combines the greedy search principle with the classical multiple signal classification (MUSIC) method. Signal space matching pursuit (SSMP) [7] iteratively updates the index set by maximizing the power reduction of residual matrix. Among those methods, SOMP has attracted great interest for its simplicity and low computational cost.

E-mail: yuxuanzhang22@m.fudan.edu.cn; jian_wang@fudan.edu.cn

Algorithm 1: The SOMP Algorithm [5]

Input : $\mathbf{\Phi} \in \mathrm{R}^{m\times n}$, $\mathbf{Y} \in \mathrm{R}^{m\times r}$, row-sparsity K, residual tolerance ϵ and maximum iteration number $k_{\max}$.

Initialize: iteration counter $k = 0$, original support $S^0 = \emptyset$, and residual matrix $\mathbf{R}^0 = \mathbf{Y}$.

while $k < k_{\max}$ *and* $\|\mathbf{R}^k\|_F < \epsilon$, **do**

  $k = k + 1$;

  Identify an index $s^k = \underset{i \in \Omega \setminus S^{k-1}}{\arg\max} \|\mathbf{R}^{k-1}\phi_i\|_1$ and merge with previously estimated support: $S^k = S^{k-1} \cup \{s^k\}$;

  Estimate the signal $\hat{\mathbf{X}}^k = \underset{\mathbf{U}:\mathrm{rsupp}(\mathbf{U})=S^k}{\arg\min} \|\mathbf{Y} - \mathbf{\Phi}\mathbf{U}\|_F$;

  Update the residual matrix: $\mathbf{R}^k = \mathbf{Y} - \mathbf{\Phi}\hat{\mathbf{X}}^k$;

end

Output : $\hat{S} = \underset{|T|=K, T\subseteq S^k}{\arg\max} \|(\hat{\mathbf{X}}^k)^T\|_F$ and $\hat{\mathbf{X}} = \underset{\mathbf{U}:\mathrm{rsupp}(\mathbf{U})=\hat{S}}{\arg\min} \|\mathbf{Y} - \mathbf{\Phi}\mathbf{U}\|_F$

The SOMP algorithm is specified in Algorithm 1. Initialized with an empty support set and a residual matrix $\mathbf{R}^0 = \mathbf{Y}$, SOMP iteratively constructs a serial of estimates S^k's for $\mathrm{rsupp}(\mathbf{X})$ via three operations: i) support identification, ii) signal estimation and iii) residual updating operations. Finally, it terminates when the residual matrix satisfies some tolerance and outputs an estimated joint K-sparse signal. This algorithm can be viewed as an extension of the orthogonal matching pursuit (OMP) algorithm [8] from the single measurement vector (SMV) model to MMV.

In analyzing the theoretical performance of SOMP, much effort has recently been made. In [9], Determe *et al.* allocated different weights to the measurement vectors when identifying the support of $\mathbf{X}$ and proposed a lower bound on the probability of full support recovery via SOMP. By assuming the noise to be additive Gaussian, they further established an upper bound for the probability that SOMP fails to achieve full support recovery within K iterations [10]. This bound indicates robustness of SOMP when the signal-to-noise ratio (SNR) is sufficiently large.

While the above results were obtained from the probabilistic perspective, there have also been works characterizing the theoretical performance of SOMP in the restricted isometry property (RIP) framework [11, 12]. A measurement matrix $\mathbf{\Phi}$ satisfies the RIP of order K, if there exists a constant $\delta \in [0,1)$ such that

$$(1-\delta)\|\mathbf{x}\|_2^2 \le \|\mathbf{\Phi}\mathbf{x}\|_2^2 \le (1+\delta)\|\mathbf{x}\|_2^2. \quad (2)$$

holds for any K-sparse vector $\mathbf{x}$. The minimum of all δ's satisfying (2), denoted as δ_K, is the restricted isometry constant (RIC). An advantage of the RIP analysis (over the probabilistic arguments) is that it offers a uniform recovery guarantee for all K-sparse signals, rather than for a fixed signal.

In [13], it is shown that SOMP accurately recovers the support of $\mathbf{X}$ in K iterations if the measurement matrix obeys

$$\delta_{K+1} < \frac{1}{\sqrt{K+1}} \quad \text{or} \quad \delta_K < \frac{\sqrt{K-1}}{\sqrt{K-1}+K}. \quad (3)$$

The condition cannot be fundamentally improved since there exists counterexamples that OMP (i.e., the SOMP algorithm with $r=1$) fails to recover some K-sparse signals in K iterations when $\delta_{K+1} = \frac{1}{\sqrt{K+1}}$ [14]. For Gaussian random measurements where each entry of $\mathbf{\Phi}$ is drawn *i.i.d.* from Gaussian distribution $\mathcal{N}(0, \frac{1}{m})$, $\delta_K < \varepsilon$ can be satisfied when

$$m \ge \mathcal{O}\left(\frac{K \log \frac{n}{K}}{\varepsilon^2}\right). \quad (4)$$

Therefore, the required number of measurements for satisfying (3) is $m \ge \mathcal{O}(K \log^2 \frac{n}{K})$. This, however, is significantly worse than the information theoretic bound $m \ge \mathcal{O}(K \log \frac{n}{K})$ of sampling complexity for the SMV model [15]. Note that the information theoretic bound is attained by many SMV-based methods (e.g., ℓ_1-minimization [16]).

In this paper, we aim to improve the exact recovery condition of SOMP in (3), and thus reduce the required sampling complexity. The key idea is to allow the SOMP algorithm to perform more than K iterations. Our result shows that if the measurement matrix satisfies the RIP with RIC

$$\delta_{\lceil 3.8K \rceil} < 2 \cdot 10^{-5}, \quad (5)$$

SOMP robustly recovers any joint K-sparse signal within $\lceil 2.8K \rceil$ iterations. The significance of our result lies not only in that it elucidates the robustness of SOMP, but also in that condition (5) can be satisfied when there are $m \ge \mathcal{O}(K \log \frac{n}{K})$ Gaussian random measurements available. This improves the result in (3) by a factor of K, and achieves the fundamental lower bound of sampling complexity in SMV.

We explain some notations used in this paper. Let $\Omega := \{1, \cdots, n\}$ and $S := \text{rsupp}(\mathbf{X})$ denote the row support of $\mathbf{X}$. Define $\mathbf{X}^S$ (or $\mathbf{X}_S$) as a sub-matrix of $\mathbf{X}$ with rows (or columns) indexed by S. For $T \subset \Omega$, $S \backslash T$ denotes the set that contains elements in S but not in T. $|S|$ is the cardinality of S. If $\mathbf{\Phi}_T$ has full column rank, $\mathbf{\Phi}_T^\dagger = (\mathbf{\Phi}'_T \mathbf{\Phi}_T)^{-1} \mathbf{\Phi}_T$ is the Moore-Penrose pseudo-inverse of $\mathbf{\Phi}_T$. $\mathcal{R}(\mathbf{\Phi})$ is the vector space spanned by the columns of $\mathbf{\Phi}$ and $\mathbf{P}_T = \mathbf{\Phi}_T \mathbf{\Phi}_T^\dagger$ represents the orthogonal projection onto $\mathcal{R}(\mathbf{\Phi}_T)$.

2. RESIDUAL ANALYSIS OF SOMP

In this section, we discuss the residual reduction behavior of SOMP. Before proceeding, we first introduce a specific division for the remaining support set $\Lambda^k := S \backslash S^k$ that haven't been selected during k iterations of SOMP. Without loss of generality, assume that $\Lambda^k = \{1, 2, \cdots, \lambda^k\}$ where $\lambda^k := |\Lambda^k|$ and that the rows in X^{Λ^k} are sorted in descending order of their ℓ_2-norms (i.e., $\|\mathbf{X}^1\|_2 \ge \|\mathbf{X}^2\|_2 \ge \cdots \ge \|\mathbf{X}^{\lambda^k}\|_2$). Then, define the subset Λ_τ^k of Λ^k as

$$\Lambda_\tau^k = \begin{cases} \emptyset, & \tau = 0, \\ \{1, 2, \cdots, 2^\tau - 1\}, & \tau = 1, 2, \cdots, \lfloor \log_2 \lambda^k \rfloor, \\ \Lambda^k, & \tau = \lfloor \log_2 \lambda^k \rfloor + 1. \end{cases}$$

Since Λ^k may have less than $2^{\lfloor \log_2 \lambda^k \rfloor + 1}$ elements, we set Λ_τ^k as the whole set when $\tau = \lfloor \log_2 \lambda^k \rfloor + 1$.

Next, we quantitively analyze the residual reduction in one iteration of SOMP.

Proposition 1. *Suppose that there are $\lambda^k = |\Lambda^k|$ remaining support indices after k iterations of SOMP. Let $j \ge k$ be an arbitrary integer. Then, for any $1 \le \tau \le \lfloor \log_2 \lambda^k \rfloor + 1$, the following inequality holds.*

$$\begin{aligned} \|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2 &\ge \frac{\left(1 - \delta_{|\Lambda_\tau^k \cup S^j|}\right)^2}{\max\{1, |\Lambda_\tau^k \backslash S^j|\}} \\ &\times \left(\|\mathbf{R}^j\|_F^2 - \left\| \mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k} \mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W} \right\|_F^2 \right). \end{aligned} \quad (6)$$

Proposition 1 offers a lower bound on the energy reduction, indicating that SOMP makes a non-trivial progress in each iteration. The proof is left to Supplementary [17].

Sketch of Proof: Our proof consists of two steps. First, we show that the energy reduction can be expressed as the projection of energy onto a subspace, which has a lower bound:

$$\begin{aligned} \|\mathbf{R}^j\|_F^2 - \|\mathbf{R}^{j+1}\|_F^2 &= \|\mathbf{P}_{\mathcal{R}(\mathbf{P}_{S^j}^\perp \phi_{s^{j+1}})} \mathbf{R}^j\|_F^2 \\ &\ge \max_{i \in \Omega \backslash S^j} \|\phi_i^T \mathbf{R}^j\|_2^2, \end{aligned} \quad (7)$$

We then estimate this bound by relating it to the inner product $\langle \mathbf{\Phi}^T \mathbf{R}^j, \mathbf{H} \rangle_F$, where $\mathbf{H} \in \mathbb{R}^{n \times r}$ obeys $\mathbf{H}^{S \cap S^k \cup \Lambda_\tau^k} = \mathbf{X}^{S \cap S^k \cup \Lambda_\tau^k}$ and $\mathbf{H}^{(S \cap S^k \cup \Lambda_\tau^k)^c} = \mathbf{0}$. To be specific,

$$\begin{aligned} \langle \mathbf{\Phi}^T \mathbf{R}^j, \mathbf{H} \rangle_F &\le \max\{1, |\Lambda_\tau^k \backslash S^j|^{1/2}\} \\ &\times \max_{i \in \Omega \backslash S^j} \left\|\phi_i^T \mathbf{R}^j\right\|_2 \|\mathbf{H}^{\Omega \backslash S^j}\|_F. \end{aligned} \quad (8)$$

On the other hand, this inner product also has a lower bound:

$$\begin{aligned} \langle \mathbf{\Phi}^T \mathbf{R}^j, \mathbf{H} \rangle_F &\ge \left(1 - \delta_{|\Lambda_\tau^k \cup S^j|}\right) \|\mathbf{H}^{\Omega \backslash S^j}\|_F \\ &\times \left(\|\mathbf{R}^j\|_F^2 - \|\mathbf{\Phi}_{\Lambda^k \backslash \Lambda_\tau^k} \mathbf{X}^{\Lambda^k \backslash \Lambda_\tau^k} + \mathbf{W}\|_F^2\right)^{1/2}. \end{aligned} \quad (9)$$

Combining (8) and (9) yields:

$$\max_{i\in\Omega\setminus S^j}\|\phi_i^T\mathbf{R}^j\|_2 \geq \frac{1-\delta_{|\Lambda_\tau^k\cup S^j|}}{\max\{1,|\Lambda_\tau^k\setminus S^j|^{1/2}\}} \times\left(\|\mathbf{R}^j\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_\tau^k}+\mathbf{W}\|_F^2\right)^{1/2}, \quad (10)$$

which, together with (7), establishes this proposition. $\square$

Then, we extend Proposition 1 to characterize the residual reduction after multiple iterations of SOMP.

Proposition 2. *Suppose that there are $\lambda^k=|\Lambda^k|$ remaining support indices after k iterations of SOMP. Let $j\geq k$ be an arbitrary integer. Then, for any $1\leq\tau\leq\lfloor\log_2\lambda^k\rfloor+1$ and any integer $\Delta_j>0$, we have*

$$\begin{aligned}&\|\mathbf{R}^{j+\Delta_j}\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_\tau^k}+\mathbf{W}\|_F^2\\ &\leq C_{\tau,j,\Delta_j}\left(\|\mathbf{R}^j\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_\tau^k}+\mathbf{W}\|_F^2\right), \quad (11)\end{aligned}$$

where $C_{\tau,j,\Delta_j}=\exp\left(-\frac{\Delta_j(1-\delta_{|\Lambda_\tau^k\cup S^{j+\Delta_j}|})^2}{\max\{1,|\Lambda_\tau^k\setminus S^j|\}}\right)$ is a constant.

We can rewrite (11) as a lower bound of residual reduction after Δ_j iterations, which is simpler to interpret:

$$\begin{aligned}&\|\mathbf{R}^j\|_F^2-\|\mathbf{R}^{j+\Delta_j}\|_F^2\\ &\geq\left(1-C_{\tau,j,\Delta_j}\right)\left(\|\mathbf{R}^j\|_F^2-\|\mathbf{\Phi}_{\Lambda^k\setminus\Lambda_\tau^k}\mathbf{X}^{\Lambda^k\setminus\Lambda_\tau^k}+\mathbf{W}\|_F^2\right).\end{aligned}$$

3. ROBUSTNESS OF SOMP

3.1. Preliminaries

Recall that the remaining support Λ^k is divided into $\lfloor\log_2\lambda^k\rfloor+1$ subsets Λ_τ^k's, from small to large. To characterize the energy distribution of $\mathbf{X}^{\Lambda^k}$, we sequentially choose larger subsets Λ_τ^k, $\tau=1,\cdots,\lfloor\log_2\lambda^k\rfloor+1$, to embrace the energy in $\mathbf{X}^{\Lambda^k}$ until a sharp decline in the incremental growth. To describe this critical point, we introduce a parameter L.

Definition 1. *Given an integer $k\geq0$ and a constant $\beta\geq1$, $L\in\{1,2,\cdots,\lfloor\log_2\lambda^k\rfloor+1\}$ is the minimal positive integer satisfying the following conditions:*

$$\|X^{\Lambda^k\setminus\Lambda_0^k}\|_F^2<\beta\|X^{\Lambda^k\setminus\Lambda_1^k}\|_F^2, \quad (12a)$$

$$\|X^{\Lambda^k\setminus\Lambda_1^k}\|_F^2<\beta\|X^{\Lambda^k\setminus\Lambda_2^k}\|_F^2, \quad (12b)$$

$$\vdots$$

$$\|X^{\Lambda^k\setminus\Lambda_{L-2}^k}\|_F^2<\beta\|X^{\Lambda^k\setminus\Lambda_{L-1}^k}\|_F^2, \quad (12c)$$

$$\|X^{\Lambda^k\setminus\Lambda_{L-1}^k}\|_F^2\geq\beta\|X^{\Lambda^k\setminus\Lambda_L^k}\|_F^2. \quad (12d)$$

We take $L=1$ and ignore (12a)–(12c) when (12d) always holds true. It is not hard to see that (12d) holds at least for $L=\lfloor\log_2(\lambda^k)\rfloor+1$, since $\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{\lfloor\log_2(\lambda^k)\rfloor+1}}\|=0$. Therefore, L always exists. From (12a)–(12d), we directly have

$$\|X^{\Lambda^k\setminus\Lambda_\tau^k}\|_F^2\leq\beta^{L-1-\tau}\|X^{\Lambda^k\setminus\Lambda_{L-1}^k}\|_F^2,\ \tau=0,\cdots,L. \quad (13)$$

In the subsequent analysis, we fix $\beta=\frac{1}{2}\exp\left(\frac{c}{4}(1-\delta_{p_2})^2\right)$ to determine L, where c and p_2 are defined in Theorem 1.

3.2. Robustness

We consider the typical MMV case where the noise is bounded in Frobenius norm. With the above analysis, we can establish a robustness result for the residual of SOMP.

Theorem 1. *Suppose that there are $\lambda^k=|\Lambda^k|$ remaining support indices after k iterations of SOMP. Also, suppose that the measurement matrix $\mathbf{\Phi}$ has ℓ_2-normalized columns and satisfies the RIP of order $\max\{p_1,p_2\}$, where $p_1=\lambda^k+k+1$ and $p_2=k+\lceil(c+1)\lambda^k\rceil$. Then, the residual of SOMP obeys*

$$\left\|\mathbf{R}^{k+\lceil c\lambda^k\rceil}\right\|_F\leq C_0\left\|\mathbf{W}\right\|_F \quad (14)$$

if $c>\max\{c',c''\}$ for some constant $t>0$, where C_0 is an absolute constant and

$$c'=\frac{-4}{(1-\delta_{p_2})^2}\ln\left(\frac{1}{2}-\frac{1}{2(1-\delta_{p_1})^2}+\frac{1}{2(1-\delta_{p_1})(1+t)(1+\delta_{\lambda^k})}\right),$$

$$c''=\frac{-4}{(1-\delta_{p_2})^2}\ln\left(\frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{1-\delta_{p_2}}{(1+t)(1+\delta_{\lambda^k})}}\right).$$

Sketch of Proof: The proof is by mathematical induction on the number λ^k of remaining support indices, which is inspired from [18]. When $\lambda^k=0$, all support indices are already selected ($S\subseteq S^k$), and consequently, the energy of the residual matrix satisfies:

$$\begin{aligned}\|\mathbf{R}^k\|_F&=\min_{\mathbf{U}:\text{rsupp}(\mathbf{U})\subseteq S^k}\|\mathbf{Y}-\mathbf{\Phi}\mathbf{U}\|_F\\ &\leq\min_{\mathbf{U}:\text{rsupp}(\mathbf{U})\subseteq S}\|\mathbf{Y}-\mathbf{\Phi}\mathbf{U}\|_F\\ &\leq\|\mathbf{Y}-\mathbf{\Phi}\mathbf{X}\|_F=\|\mathbf{W}\|_F.\end{aligned}$$

Next, assume that the robustness result holds up to an integer $N-1$ (i.e., $\lambda^k\leq N-1$). Then, we consider $\lambda^k=N$ and prove the result in two cases (i.e., $L=1$ and $L\geq2$). Intuitively, $L=1$ refers to that the energy of $\mathbf{X}^{\Lambda^k}$ is well concentrated on the subset Λ_1^k. Whereas, $L\geq2$ indicates that Λ_1^k is not large enough to embrace the major energy of $\mathbf{X}^{\Lambda^k}$ and we need Λ_2^k or even larger subset.

Case 1: $L=1$**.** We prove this case by first showing that

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F<\|\mathbf{X}^{\Lambda^k}\|_F. \quad (15)$$

To this end, we employ Proposition 1 to obtain

$$\|\mathbf{R}^{k+1}\|_F\leq\mu_0\|\mathbf{X}^{\Lambda^k}\|_F+\left(1+\frac{1}{t}\right)^{1/2}\|\mathbf{W}\|_F, \quad (16)$$

where $\mu_0 := (1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}(1-(1-\frac{1}{\beta})(1-\delta_{p_1})^2)^{1/2}$. Thus,

$$\|\mathbf{X}^{\Lambda^{k+1}}\|_F \le \frac{\|\mathbf{R}^{k+1}\|_F + \|\mathbf{W}\|_F}{\sqrt{1-\delta_{p_1}}} \le \mu_1\|\mathbf{X}^{\Lambda^k}\|_F + \mu_2\|\mathbf{W}\|_F,$$

where $\mu_1 := \mu_0(1-\delta_{p_1})^{-1/2}$ and $\mu_2 := (1-\delta_{p_1})^{-1/2}\big(1 + (1+\frac{1}{t})^{1/2}\big)$. Since $\mu_1 < 1$ whenever $c > c'$, it is clear that (15) holds when $\|\mathbf{W}\|_F < \frac{1-\mu_1}{\mu_2}\|\mathbf{X}^{\Lambda^k}\|_F$. The remaining case (i.e., $\|\mathbf{W}\|_F \ge \frac{1-\mu_1}{\mu_2}\|\mathbf{X}^{\Lambda^k}\|_F$) is trivial as it easily leads to the robustness result.

Then, from the definition of Λ^k, we have $\Lambda^{k+1} \subset \Lambda^k$, which together with (15) implies $\lambda^{k+1} < \lambda^k = N$, and hence

$$\begin{aligned} k+1+\lceil c\lambda^{k+1}\rceil &= k + \lceil 1 + c\lambda^{k+1}\rceil \\ &\le k + \lceil c(\lambda^{k+1}+1)\rceil \le k + \lceil cN\rceil. \end{aligned} \tag{17}$$

Since the residual energy is always non-increasing, we have

$$\|\mathbf{R}^{k+\lceil cN\rceil}\|_F \le \|\mathbf{R}^{k+1+\lceil c\lambda^{k+1}\rceil}\|_F \le C_0\|\mathbf{W}\|_F.$$

where the second inequality uses the induction hypothesis.

Case 2: $L \ge 2$**.** The proof for this case is similar to that of Case 1. Specifically, we first show that

$$\|\mathbf{X}^{\Lambda^{k_L}}\|_F \le \|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F, \tag{18}$$

where $k_L := k + \sum_{\tau=1}^{L}\lceil\frac{c}{4}|\Lambda^k_\tau|\rceil$. From the definition of k_L, and also noting that $|\Lambda^k_\tau| \le 2^\tau - 1$, we can easily get

$$k_L \le k + \sum_{\tau=1}^{L}\left\lceil\frac{c}{4}(2^\tau-1)\right\rceil \overset{[18,\text{ Appendix B}]}{\le} k + \lceil c2^{L-1}\rceil - 1. \tag{19}$$

To prove (18), we first use Proposition 2 to bound the energy of residual $\mathbf{R}^{k_L}$ as

$$\|\mathbf{R}^{k_L}\|_F \le \tau_0\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F + \left(1+\frac{1}{t}\right)^{1/2}\|\mathbf{W}\|_F,$$

where $\tau_0 := \beta^{-1/2}(1+t)^{1/2}(1+\delta_{\lambda^k})^{1/2}(2-\frac{1}{\beta})^{1/2}$. Then, it follows that

$$\begin{aligned} \|\mathbf{X}^{\Lambda^{k_L}}\|_F &\le \frac{\|\mathbf{R}^{k_L}\|_F + \|\mathbf{W}\|_F}{(1-\delta_{p_2})^{1/2}}, \\ &\le \tau_1\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F + \tau_2\|\mathbf{W}\|_F, \end{aligned} \tag{20}$$

where $\tau_1 := (1-\delta_{p_2})^{-1/2}\tau_0$ and $\tau_2 := (1-\delta_{p_2})^{-1/2}\big(1 + (1+\frac{1}{t})^{1/2}\big)$. Note that $\tau_1 < 1$ when $c > c''$, (18) holds true whenever $\|\mathbf{W}\|_F < \frac{1-\tau_1}{\tau_2}\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F$. For the remaining case where $\|\mathbf{W}\|_F \ge \frac{1-\tau_1}{\tau_2}\|\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}\|_F$, we can directly prove the robustness conclusion in (14).

Then, recall that rows in $\mathbf{X}^{\Lambda^k}$ are sorted in descending order in their ℓ_2-norms. Hence, $\mathbf{X}^{\Lambda^k\setminus\Lambda^k_{L-1}}$ includes the elements with the lowest energy in $\mathbf{X}^{\Lambda^k}$. This, together with (18), indicates

$$|\Lambda^{k_L}| < |\Lambda^k\setminus\Lambda^k_{L-1}| \le N - 2^{L-1} + 1. \tag{21}$$

We thus have

$$\begin{aligned} k_L + \lceil c\lambda^{k_L}\rceil &\overset{(19)}{\le} k + \lceil c2^{L-1}\rceil - 1 + \lceil c(N-2^{L-1})\rceil \\ &\le k + \lceil cN\rceil. \end{aligned} \tag{22}$$

Based on the non-increasing nature of residual energy, we conclude that

$$\|\mathbf{R}^{k+\lceil cN\rceil}\|_F \le \|\mathbf{R}^{k_L+\lceil c\lambda^{k_L}\rceil}\|_F \le C_0\|\mathbf{W}\|_F. \tag{23}$$

Combining the cases of $L = 1$ and $L \ge 2$ completes the proof. □

From Theorem 1, we can further derive the robustness results for the recovered signal by letting $k = 0$ and $\lambda^k = K$.

Theorem 2. *Consider the MMV model in (1). Then, SOMP robustly recover any joint K-sparse signal $\mathbf{X}$ with*

$$\|\mathbf{X} - \mathbf{X}^{\lceil cK\rceil}\|_F \le C_1\|\mathbf{W}\|_F,$$
$$\|\mathbf{X} - \hat{\mathbf{X}}\|_F \le C\,\|\mathbf{W}\|_F,$$

if $c > \max\Big\{-\frac{4}{(1-\delta)^2}\ln\big(\frac{1}{2} - \frac{1}{2(1-\delta)^2} + \frac{1}{2(1-\delta)(1+t)(1+\delta)}\big), -\frac{4}{(1-\delta)^2}\ln\Big(\frac{1}{2} - \frac{1}{2}\sqrt{1-\frac{1-\delta}{(1+t)(1+\delta)}}\Big)\Big\}$ *for some constant $t > 0$, and C_1, C are constants depending on $\delta := \delta_{\lceil(c+1)K\rceil}$.*

Remark 1. *One can interpret from Theorem 2 that when $\delta, t \to 0$, $c > 4\ln 2$ is sufficient to guarantee the robustness of residual. Indeed, when $c = 2.8$ (i.e., when SOMP runs at most $\lceil 2.8K\rceil$ iterations) and $t \to 0$, the robust recovery can be ensured under $\delta_{\lceil 3.8K\rceil} < 2\cdot 10^{-5}$. This matches the existing result of OMP [18], which considered exact support recovery of sparse signals from the SMV model, but with extension to the noisy MMV scenario. To the best of our knowledge, Theorem 2 is the first robustness result for recovering joint sparse signals from the MMV model with the minimum sampling complexity $m \ge \mathcal{O}(K\log\frac{n}{K})$.*

4. CONCLUSION

In this paper, with an aim of improving its sampling complexity for signal recovery from the MMV model, we have investigated the theoretical performance of SOMP. Our result shows that SOMP performs robust recovery of any joint K-sparse n-dimensional signal from its noisy measurements within $\lceil 2.8K\rceil$ iterations under a mild RIP condition $\delta_{\lceil 3.8K\rceil} < 2\cdot 10^{-5}$. In particular, for Gaussian random measurements, the RIP condition can be satisfied when the number of measurements is $\mathcal{O}(K\log\frac{n}{K})$, which attains the information theoretic bound of sampling complexity in the SMV literature. Whether it is possible to further reduce the required iteration number $\lceil 2.8K\rceil$ to, e.g., no more than $2K$ remains an interesting open question.

5. REFERENCES

[1] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[2] J. A. Tropp, "Algorithms for simultaneous sparse approximation. part ii: Convex relaxation," *Signal Processing*, vol. 86, no. 3, pp. 589–602, 2006.

[3] J. Ding, L. Chen, and Y. Gu, "Robustness of orthogonal matching pursuit for multiple measurement vectors in noisy scenario," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3813–3816.

[4] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1135–1146, 2012.

[5] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[6] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive music: Revisiting the link between compressive sensing and array signal processing," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 278–301, 2012.

[7] J. Kim, J. Wang, L. T. Nguyen, and B. Shim, "Joint sparse recovery using signal space matching pursuit," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 5072–5096, 2020.

[8] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1993, vol. 1, pp. 40–44.

[9] J. F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "Simultaneous orthogonal matching pursuit with noise stabilization: Theoretical analysis," 2015.

[10] J. F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "On the noise robustness of simultaneous orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 4, pp. 864–875, 2017.

[11] J. F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "Improving the correlation lower bound for simultaneous orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1642–1646, 2016.

[12] H. Li, Y. Ma, and Y. Fu, "An improved rip-based performance guarantee for sparse signal recovery via simultaneous orthogonal matching pursuit," *Signal Processing*, vol. 144, pp. 29–35, 2018.

[13] J. F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "On the exact recovery condition of simultaneous orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 164–168, 2016.

[14] J. Wen, Z. Zhou, J. Wang, X. Tang, and M. Qun, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1370–1382, 2017.

[15] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[16] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[17] Y. Zhang and J. Wang, "Robust recovery of joint sparse signals via simultaneous orthogonal matching pursuit–supplementary," Online avaliable: `https://github.com/xuanxuan202/Supplementary`, 2023.

[18] J. Wang and B. Shim, "Exact recovery of sparse signals using orthogonal matching pursuit: How many iterations do we need?," *IEEE Transactions on Signal Processing*, vol. 64, no. 16, pp. 4194–4202, 2016.