

A RIP-Based Performance Guarantee of Covariance-Assisted Matching Pursuit

Jiayang Wang, Gen Li, Lucas Rencker, Wenwu Wang, and Yuantao Gu

Abstract—An OMP-like Covariance-Assisted Matching Pursuit (CAMP) method has recently been proposed. Given a prior-knowledge of the covariance and mean of the sparse coefficients, CAMP balances the least squares estimator and the prior-knowledge by leveraging the Gauss-Markov theorem. In this letter, we study the performance of CAMP in the framework of restricted isometry property (RIP). It is shown that under some conditions on RIP and the minimum magnitude of the nonzero elements of the sparse signal, CAMP with sparse level K can recover the exact support of the sparse signal from noisy measurements. l_2 bounded noise and Gaussian noise are considered in our analysis. We also discuss the extreme conditions of noise (e.g. the noise power is infinite) to simply show the stability of CAMP.

Index Terms—Sparse recovery, Covariance-Assisted Matching Pursuit (CAMP), restricted isometry property (RIP), compressed sensing.

I. INTRODUCTION

Sparse recovery refers to the problem of reconstructing a sparse vector from a very limited number of noisy linear measurements [1] [2]. We consider the following model:

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e} \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $M < N$ is the sensing matrix, $\mathbf{x}^* \in \mathbb{R}^N$ is the sparse vector with at most K ($K \ll N$) non-zero elements, and \mathbf{e} is an additive noise term. The aim is to find \mathbf{x}^* given \mathbf{y} and \mathbf{A} .

In some situations the above problem may be well solved by using greedy algorithms, including the well-known orthogonal matching pursuit (OMP) algorithm [3]. Several works have attempted to improve the performance of OMP by leveraging some prior knowledge about the unknown vector to be recovered [4] [5]. The work of [6] employed the prior knowledge of the covariance and mean of the sparse representations

Manuscript received December 19, 2017; revised January 28, 2018; accepted February 24, 2018. Date of publication; date of current version. J. Wang, G. Li, and Y. Gu were funded by the National Natural Science Foundation of China (NSFC 61531166005, 61571263), the National Key Research and Development Program of China (Project No. 2016YFE0201900, 2017YFC0403600), and Tsinghua University Initiative Scientific Research Program (Grant 2014Z01005). L. Rencker and W. Wang were funded by the European Unions H2020 Framework Program (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet. (Corresponding author: Yuantao Gu.)

J. Wang, G. Li, and Y. Gu are with Beijing National Research Center for Information Science and Technology (BNRist) and Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (jy-wang14@mails.tsinghua.edu.cn, g-li16@mails.tsinghua.edu.cn, gyt@tsinghua.edu.cn).

L. Rencker and W. Wang are with Department of Electrical and Electronic Engineering, University of Surrey, Guildford, GU2 7XH, United Kingdom. (lucas.rencker@surrey.ac.uk, w.wang@surrey.ac.uk)

by leveraging Gauss-Markov theorem to obtain performance gains. In contrast to other matching pursuit methods using prior knowledge, the Covariance-Assisted Matching Pursuit (CAMP) doesn't require the explicit prior probabilistic modeling of the sparse coefficients. This makes it better suited for applications such as image restoration [7] and audio processing [8].

For a vector \mathbf{x}^* , we define its support as $\text{supp}(\mathbf{x}^*) = \{i | \mathbf{x}^*(i) \neq 0\}$. In this letter, we consider the exact support recovery of sparse signals with sparsity level K , i.e., $|\text{supp}(\mathbf{x}^*)| = K$. In the noiseless case, CAMP works as OMP does. Thus, we shall consider the noisy case in our letter.

In this letter, we aim to derive an RIP-based condition ensuring the exact support recovery of K -sparse signals for the CAMP algorithm in the noisy case.

II. PRELIMINARY

A. Notations

We denote the i th column of \mathbf{A} by \mathbf{a}_i , i.e., $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$. Here, we assume that the columns of \mathbf{A} are normalized, i.e., $\|\mathbf{a}_i\|_2 = 1$ for $i = 1, \dots, N$. Let Ω be the support set of a sparse signal \mathbf{x} , i.e., $\Omega = \text{supp}(\mathbf{x})$. Given an index set Γ , we denote by \mathbf{A}_Γ a sub-matrix of \mathbf{A} , which is composed of the columns indexed by Γ . The temporary solution in the t -th matching pursuit stage is denoted by $\mathbf{x}^t \in \mathbb{R}^t$, and the corresponding estimation of \mathbf{x}^* is denoted by $\hat{\mathbf{x}}^t \in \mathbb{R}^N$. $\|\cdot\|$ denotes the ℓ_2 norm.

B. OMP and RIP-based Analysis

The OMP selects the most correlated atom \mathbf{a}_{i_t} with the smallest residual vector at step t and updates the index set $\Omega^t = \Omega^{t-1} \cup i_t$. It then calculates the estimated solution \mathbf{x}^t by projecting \mathbf{y} onto the subspace spanned by the sub-dictionary \mathbf{A}_{Ω^t} :

$$\mathbf{x}^t = (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t})^{-1} \mathbf{A}_{\Omega^t}^T \mathbf{y}, \quad (2)$$

and updates the residual $\mathbf{r}^t = \mathbf{y} - \mathbf{A}_{\Omega^t} \mathbf{x}^t$.

The Restricted Isometry Property (RIP) of a sensing matrix \mathbf{A} is often used to analyze the recovery performance of OMP. A matrix \mathbf{A} is said to satisfy the RIP of order K if there exists a smallest positive constant δ_K such that

$$(1 - \delta_K) \|\mathbf{x}\|^2 \leq \|\mathbf{A}\mathbf{x}\|^2 \leq (1 + \delta_K) \|\mathbf{x}\|^2 \quad (3)$$

for all K -sparse vectors \mathbf{x} , where δ_K is called the Restricted Isometry Constant (RIC).

There are lots of results on analyzing OMP in the noiseless case by using the RIP. Davenport and Wakin proved that

Algorithm 1 Orthogonal Matching Pursuit

1: Input: \mathbf{y} , $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] \in \mathbb{R}^{M \times N}$, K_{\max} , ε .
2: Initialize: $t = 0$, support set $\Omega^0 = \emptyset$, residual $\mathbf{r}^0 = \mathbf{y}$.
3: **while** $t \leq K_{\max}$ or $\|\mathbf{r}^t\| \geq \varepsilon$ **do**
4: Select atom: $j = \arg \max_j |\langle \mathbf{r}^{t-1}, \mathbf{a}_j \rangle|$;
5: Update support set: $\Omega^t = \Omega^{t-1} \cup j$;
6: Update estimation: $\mathbf{x}^t = (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t})^{-1} \mathbf{A}_{\Omega^t}^T \mathbf{y}$;
7: Update residual: $\mathbf{r}^t = \mathbf{y} - \mathbf{A}_{\Omega^t} \mathbf{x}^t$;
8: $t = t + 1$;
9: **end while**
10: Output: $\Omega^{K_{\max}}$, $\mathbf{x}^{K_{\max}}$.

[9] under the condition $\delta_{K+1} < \frac{1}{3\sqrt{K}}$, OMP will accurately recover the sparse signal without noise. To our knowledge, [10] has proven that under the condition $\delta_{K+1} < \frac{1}{\sqrt{K+1}}$ OMP promises to succeed in the noiseless case.

In the noisy case, it has been proved in [11] that under the RIP-based condition $\delta_{K+1} < \frac{1}{\sqrt{K+3}}$ and some assumptions on the minimum magnitude of the nonzero elements of the input signal, OMP will accurately recover the support of the sparse signal with noisy measurements. Further work can be found in [12].

Next we recall some useful properties of RIP to make preparations for further analysis.

Lemma 1. Suppose that matrix \mathbf{A} satisfies RIP of both order K_1 and K_2 , where $K_1 \leq K_2$, then $\delta_{K_1} \leq \delta_{K_2}$.

Lemma 2. [13] If matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ satisfies RIP of order K and $\delta_K \leq 1$, then for any vector $\mathbf{u} \in \mathbb{R}^N$ with $|\text{supp}(\mathbf{u})| \leq K$,

$$(1 - \delta_K) \|\mathbf{u}\| \leq \|\mathbf{A}^T \mathbf{A} \mathbf{u}\| \leq (1 + \delta_K) \|\mathbf{u}\|, \quad (4)$$

$$\frac{1}{1 + \delta_K} \|\mathbf{u}\| \leq \|(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{u}\| \leq \frac{1}{1 - \delta_K} \|\mathbf{u}\|. \quad (5)$$

Lemma 3. [14] Suppose that matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ satisfies RIP of order K . Let Γ and Θ be two disjoint index sets, i.e., $\Gamma \cap \Theta = \emptyset$, and $|\text{supp}(\Gamma \cup \Theta)| \leq K$. Then for any vector $\mathbf{u} \in \mathbb{R}^N$ with $\text{supp}(\mathbf{u}) \in \Gamma$,

$$\|\mathbf{A}_{\Theta}^T \mathbf{A} \mathbf{u}\| = \|\mathbf{A}_{\Theta}^T \mathbf{A}_{\Gamma} \mathbf{u}_{\Gamma}\| \leq \delta_K \|\mathbf{u}\|. \quad (6)$$

C. Covariance-Assisted Matching Pursuit (CAMP)

Additional assumptions are made to help the recovery. We assume that \mathbf{e} is a zero-mean random noise component with variance σ^2 , where σ^2 can be estimated from the training data. It is further assumed that the non-zero entries of \mathbf{x} , denoted by \mathbf{x}_{Ω} , are the elements of a vector with mean $\boldsymbol{\mu}_{\Omega} \in \mathbb{R}^{|\Omega|}$ and covariance $\boldsymbol{\Lambda}_{\Omega} \in \mathbb{R}^{|\Omega| \times |\Omega|}$, which are extracted from a single mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times N}$ during the matching pursuit stages.

Based on the above assumption, the temporary solution of CAMP for the t -th matching pursuit stage is solved by [6]

$$\mathbf{x}^t = (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t} + \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1})^{-1} (\mathbf{A}_{\Omega^t}^T \mathbf{y} + \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1} \boldsymbol{\mu}_{\Omega^t}), \quad (7)$$

as described in Algorithm 2. The atom selection and residual update steps are performed in a similar way as OMP. In

Algorithm 2 Covariance-Assisted Matching Pursuit

1: Input: \mathbf{y} , $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] \in \mathbb{R}^{M \times N}$, K_{\max} , ε .
2: Initialize: $t = 0$, support set $\Omega^0 = \emptyset$, residual $\mathbf{r}^0 = \mathbf{y}$.
3: **while** $t \leq K_{\max}$ or $\|\mathbf{r}^t\| \geq \varepsilon$ **do**
4: Select atom: $j = \arg \max_j |\langle \mathbf{r}^{t-1}, \mathbf{a}_j \rangle|$, $\hat{j} \notin \Omega^{t-1}$;
5: Update support set: $\Omega^t = \Omega^{t-1} \cup j$;
6: Update estimation:
 $\mathbf{x}^t = (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t} + \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1})^{-1} (\mathbf{A}_{\Omega^t}^T \mathbf{y} + \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1} \boldsymbol{\mu}_{\Omega^t})$;
7: Update residual: $\mathbf{r}^t = \mathbf{y} - \mathbf{A}_{\Omega^t} \mathbf{x}^t$;
8: $t = t + 1$;
9: **end while**
10: Output: $\Omega^{K_{\max}}$, $\mathbf{x}^{K_{\max}}$.

noiseless case, i.e., $\sigma^2 = 0$, equation (7) will reduce to (2) and CAMP degenerates to OMP. Therefore, in the following analysis, we assume that σ^2 is strictly greater than zero.

III. ANALYSIS OF CAMP

To show how CAMP works, we reshape its update equation. By splitting (7) into two items, we have

$$\mathbf{x}^t = (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t} + \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1})^{-1} \mathbf{A}_{\Omega^t}^T \mathbf{y} + (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t} + \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1})^{-1} \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1} \boldsymbol{\mu}_{\Omega^t}. \quad (8)$$

Applying the equalities of

$$(\mathbf{A} + \mathbf{B})^{-1} = (\mathbf{I} + \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{A}^{-1} \quad (9)$$

$$= (\mathbf{B}^{-1} \mathbf{A} + \mathbf{I})^{-1} \mathbf{B}^{-1}, \quad (10)$$

where \mathbf{A} and \mathbf{B} denote any invertible matrices, to the two items in the RHS of (8), respectively, we have

$$\mathbf{x}^t = (\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t})^{-1} \mathbf{A}_{\Omega^t}^T \mathbf{y} + (\mathbf{H}_{\Omega^t} + \mathbf{I})^{-1} (\sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1})^{-1} \sigma^2 \boldsymbol{\Lambda}_{\Omega^t}^{-1} \boldsymbol{\mu}_{\Omega^t}, \quad (11)$$

where $\mathbf{H}_{\Omega^t} = \sigma^{-2} \boldsymbol{\Lambda}_{\Omega^t} \mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t}$. Then (11) is further shaped to

$$\mathbf{x}^t = (\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} \mathbf{x}_{\text{LS}}^t + (\mathbf{I} + \mathbf{H}_{\Omega^t})^{-1} \boldsymbol{\mu}_{\Omega^t}, \quad (12)$$

by denoting $\mathbf{x}_{\text{LS}}^t = (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t})^{-1} \mathbf{A}_{\Omega^t}^T \mathbf{y}$ as the least-squares solution.

According to the equality of

$$(\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} + (\mathbf{I} + \mathbf{H}_{\Omega^t})^{-1} = \mathbf{I}, \quad (13)$$

\mathbf{x}^t in the LHS of (12) can be understood as a weighted average between the least-squares solution \mathbf{x}_{LS}^t and the prior mean $\boldsymbol{\mu}_{\Omega^t}$.

It is clear that $\sigma^{-2} \boldsymbol{\Lambda}_{\Omega^t}$ and $\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t}$ are both positive semi-definite matrices, and $\mathbf{H}_{\Omega^t} = \sigma^{-2} \boldsymbol{\Lambda}_{\Omega^t} \mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t}$ has the same eigenvalues as $\sigma^{-2} \mathbf{A}_{\Omega^t}^T \boldsymbol{\Lambda}_{\Omega^t} \mathbf{A}_{\Omega^t}$, which can be derived from the identity $\det(t\mathbf{I} - \mathbf{A}\mathbf{B}) = \det(t\mathbf{I} - \mathbf{B}\mathbf{A})$. Thus, all the eigenvalues of the matrix \mathbf{H}_{Ω^t} , which are denoted by λ_i , satisfy $\lambda_i \geq 0$. It is clear that the matrices \mathbf{H} , $(\mathbf{I} + \mathbf{H})^{-1}$ and $(\mathbf{I} + \mathbf{H}^{-1})^{-1}$ have the corporate eigenvectors, and the eigenvalues of $(\mathbf{I} + \mathbf{H})^{-1}$ and $(\mathbf{I} + \mathbf{H}^{-1})^{-1}$, which are denoted by $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$, can be expressed as

$$\lambda_i^{(1)} = \lambda_i ((\mathbf{I} + \mathbf{H})^{-1}) = \frac{1}{1 + \lambda_i},$$

$$\lambda_i^{(2)} = \lambda_i((\mathbf{I} + \mathbf{H}^{-1})^{-1}) = \frac{\lambda_i}{1 + \lambda_i}.$$

Since $\lambda_i \geq 0$, we can easily conclude that

$$0 \leq \lambda_i^{(1)} \leq 1, \quad 0 \leq \lambda_i^{(2)} \leq 1. \quad (14)$$

As noise power σ^2 increases, the weight of the prior mean $\boldsymbol{\mu}_{\Omega^t}$ increases while the weight of least-squares estimation \mathbf{x}_{LS}^t decreases. Under the condition that the noise power is infinite, the estimation \mathbf{x}^t would converge to the prior mean $\boldsymbol{\mu}_{\Omega^t}$, i.e., $\mathbf{x}_{\sigma^2 \rightarrow \infty}^t \rightarrow \boldsymbol{\mu}_{\Omega^t}$. In this way, CAMP will be able to stay stable under noisy conditions by employing the prior knowledge.

IV. EXACT SUPPORT RECOVERY VIA CAMP

In this section, we analyze the conditions on RIP and sparse signals for the exact support recovery via CAMP. Both the l_2 bounded noise and the Gaussian noise are considered.

A. l_2 Bounded Noise

Theorem 1. Suppose that $\|\mathbf{e}\| \leq \epsilon_1$ and \mathbf{A} satisfies the RIP condition that

$$\delta_K < \frac{1}{2 + \sqrt{K-1}} \quad (15)$$

in model (1). Define the differential vector $\mathbf{x}_e = \mathbf{x}^* - \boldsymbol{\mu}_{\Omega}$ and suppose that $\|\mathbf{x}_e\| \leq E$. Then CAMP with stopping rule $t = K$ will exactly recover the support Ω of the sparse signal \mathbf{x}^* with sparsity level K , if the minimum magnitude of nonzero elements of \mathbf{x}^* satisfies

$$\min_{i \in \Omega} |\mathbf{x}^*(i)| \geq \frac{2\delta_{K+1}(1 - \delta_{K+1})E + \left(1 - \delta_{K+1} + (1 + \delta_{K+1})^{\frac{3}{2}}\right) \epsilon_1}{1 - (2 + \sqrt{K})\delta_{K+1}} \quad (16)$$

Proof: As above, we have already known the sparsity level K , i.e., $|\Omega| = K$. Since CAMP employs prior knowledge during each iteration, the differential vector \mathbf{x}_e deserves some attention. According to its definition, \mathbf{x}_e has the same support as \mathbf{x}^* , i.e., $\text{supp}(\mathbf{x}_e) = \Omega$. \mathbf{x}_e with high energy means that the prior knowledge reduces the algorithm's accuracy.

Under the assumption that the sparsity level K is known, it remains to be shown that CAMP will select the correct atom at each stage. Our proof applies the mathematical induction method [15]. We first suppose that CAMP selects correct indices at the first t iterations, i.e., $\Omega_t \subset \Omega$. It naturally holds true when $t = 1$. Then, the condition

$$\|\mathbf{A}_{\Omega^c}^T \mathbf{r}^t\|_{\infty} < \|\mathbf{A}_{\Omega \setminus \Omega_t}^T \mathbf{r}^t\|_{\infty} \quad (17)$$

and

$$\|\mathbf{A}_{\Omega_t}^T \mathbf{r}^t\|_{\infty} < \|\mathbf{A}_{\Omega \setminus \Omega_t}^T \mathbf{r}^t\|_{\infty} \quad (18)$$

can guarantee that CAMP selects a correct index, i.e., $i_t \in \Omega$ and $i_t \notin \Omega_t$, at the t -th iteration for $t = 0, 1, \dots, K-1$. The inequality (17) guarantees that CAMP won't select a wrong atom and (18) promises to avoid repetition of the same atom.

We first consider condition (18). LHS of (18) can be directly derived from the update solution (8):

$$\mathbf{A}_{\Omega_t}^T \mathbf{r}^t = \sigma^2 \boldsymbol{\Lambda}_{\Omega_t}^{-1} (\mathbf{x}^t - \boldsymbol{\mu}_{\Omega^t}). \quad (19)$$

Unlike OMP, CAMP loses some energy at each iteration and will not guarantee that the same atom won't be selected twice. However, we have little knowledge about the estimated covariance \mathbf{A} , which means that we cannot make more assumptions on the matrix $\boldsymbol{\Lambda}_{\Omega_t}^{-1}$. Thus, the upper bound of $\|\mathbf{A}_{\Omega_t}^T \mathbf{r}^t\|_{\infty}$ and the solution to (18) won't be that meaningful. Here, we follow [6] to avoid multiple selection of the same atom by selecting atoms from $\Omega \setminus \Omega_t$ at the $(t+1)$ -th stage and we appreciate the work of [16] that solves this problem by adding a multivariate Gaussian prior on the sparse signal \mathbf{x}^* .

Now we can apply previous work [17] on OMP to obtain an upper bound for the LHS of (17)

$$\|\mathbf{A}_{\Omega^c}^T \mathbf{r}^t\|_{\infty} \leq \delta_{K+1} \|\mathbf{x}^* - \hat{\mathbf{x}}^t\| + \epsilon_1. \quad (20)$$

Using the triangle inequality, we can relax the upper bound as

$$\|\mathbf{A}_{\Omega^c}^T \mathbf{r}^t\|_{\infty} \leq \delta_{K+1} \|\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t\| + \delta_{K+1} \|\mathbf{x}_{\Omega \setminus \Omega^t}^*\| + \epsilon_1. \quad (21)$$

Next, we consider the RHS of (17). Since $|\text{supp}(\mathbf{A}_{\Omega \setminus \Omega_t}^T \mathbf{r}^t)| = K - t$, we have

$$\|\mathbf{A}_{\Omega \setminus \Omega_t}^T \mathbf{r}^t\|_{\infty} \geq \frac{\|\mathbf{A}_{\Omega \setminus \Omega_t}^T \mathbf{r}^t\|}{\sqrt{K-t}}. \quad (22)$$

Notice that

$$\begin{aligned} \|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{r}^t\| &= \|\mathbf{A}_{\Omega \setminus \Omega^t}^T (\mathbf{A}(\mathbf{x}^* - \hat{\mathbf{x}}^t) + \mathbf{e})\| \\ &= \|\mathbf{A}_{\Omega \setminus \Omega^t}^T (\mathbf{A}_{\Omega \setminus \Omega^t} \mathbf{x}_{\Omega \setminus \Omega^t}^* + \mathbf{A}_{\Omega^t} (\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t) + \mathbf{e})\| \\ &\geq \|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{A}_{\Omega \setminus \Omega^t} \mathbf{x}_{\Omega \setminus \Omega^t}^*\| \\ &\quad - \|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{A}_{\Omega^t} (\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t)\| - \|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{e}\|. \end{aligned} \quad (23)$$

By Lemma 3, it holds that

$$\|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{A}_{\Omega \setminus \Omega^t} \mathbf{x}_{\Omega \setminus \Omega^t}^*\| \geq (1 - \delta_K) \|\mathbf{x}_{\Omega \setminus \Omega^t}^*\|, \quad (24)$$

and by Lemma 4, we have

$$\|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{A}_{\Omega^t} (\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t)\| \leq \delta_K \|\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t\|. \quad (25)$$

It has been proven in [17] that

$$\|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{e}\| \leq \|\mathbf{A}_{\Omega \setminus \Omega^t}^T \|\mathbf{e}\| \leq \sqrt{1 + \delta_K} \epsilon_1. \quad (26)$$

Based on (24), (25), (26), and (23), we can easily verify that

$$\begin{aligned} \|\mathbf{A}_{\Omega \setminus \Omega^t}^T \mathbf{r}^t\|_{\infty} &\geq \frac{1 - \delta_K}{\sqrt{K-t}} \|\mathbf{x}_{\Omega \setminus \Omega^t}^*\| \\ &\quad - \frac{\delta_K}{\sqrt{K-t}} \|\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t\| - \frac{\sqrt{1 + \delta_K}}{\sqrt{K-t}} \epsilon_1. \end{aligned} \quad (27)$$

It follows from (12) and (14) that

$$\begin{aligned} &\|\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t\| \\ &= \|(\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} (\mathbf{x}_{\Omega^t}^* - \mathbf{x}_{\text{LS}}^t) + (\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} (\mathbf{x}_{\Omega^t}^* - \boldsymbol{\mu}_{\Omega^t})\| \\ &\leq \|(\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} (\mathbf{x}_{\Omega^t}^* - \mathbf{x}_{\text{LS}}^t)\| + \|(\mathbf{I} + \mathbf{H}_{\Omega^t}^{-1})^{-1} (\mathbf{x}_{\Omega^t}^* - \boldsymbol{\mu}_{\Omega^t})\| \\ &\leq \|\mathbf{x}_{\Omega^t}^* - \mathbf{x}_{\text{LS}}^t\| + \|\mathbf{x}_{\Omega^t}^* - \boldsymbol{\mu}_{\Omega^t}\|. \end{aligned} \quad (28)$$

Recall Lemma 3 and Lemma 4, it holds that

$$\begin{aligned} \|\mathbf{x}_{\Omega^t}^* - \mathbf{x}_{\text{LS}}^t\| &= \|\mathbf{x}_{\Omega^t}^* - (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t})^{-1} \mathbf{A}_{\Omega^t} (\mathbf{A} \mathbf{x}^* + \mathbf{e})\| \\ &= \|\mathbf{x}_{\Omega^t}^* - (\mathbf{A}_{\Omega^t}^T \mathbf{A}_{\Omega^t})^{-1} \mathbf{A}_{\Omega^t} (\mathbf{A}_{\Omega \setminus \Omega^t} \mathbf{x}_{\Omega \setminus \Omega^t}^* + \mathbf{e})\| \\ &\leq \frac{\delta_K}{1 - \delta_K} \|\mathbf{x}_{\Omega \setminus \Omega^t}^*\| + \frac{\sqrt{1 + \delta_K}}{1 - \delta_K} \epsilon_1. \end{aligned} \quad (29)$$

Then we have

$$\|\mathbf{x}_{\Omega^t}^* - \mathbf{x}^t\| \leq \frac{\delta_K}{1 - \delta_K} \|\mathbf{x}_{\Omega \setminus \Omega^t}^*\| + \frac{\sqrt{1 + \delta_K}}{1 - \delta_K} \epsilon_1 + E. \quad (30)$$

To satisfy the condition (17), it's sufficient to prove

$$\begin{aligned} & \left(\frac{1 - \delta_K}{\sqrt{K - t}} - \delta_{K+1} - \frac{\delta_{K+1}^2(1 + \sqrt{K - t})}{(1 - \delta_{K+1})\sqrt{K - t}} \right) \|\mathbf{x}_{\Omega \setminus \Omega^t}^*\| \\ & \geq \left(\frac{\delta_K}{\sqrt{K - t}} + \delta_{K+1} \right) E \\ & + \left(1 + \frac{\sqrt{1 + \delta_K}}{\sqrt{K - t}} + \left(\frac{\delta_K}{\sqrt{K - t}} + \delta_{K+1} \right) \frac{\sqrt{1 + \delta_K}}{1 - \delta_K} \right) \epsilon_1, \end{aligned} \quad (31)$$

where the inequality follows from (17), (21), (27), (30) and a derivative fact of Lemma 1 that $\delta_K \leq \delta_{K+1}$. Notice that $\sqrt{K - t} \leq \sqrt{K}$, $\frac{1}{\sqrt{K - t}} \leq 1$ and

$$\|\mathbf{x}_{\Omega \setminus \Omega^t}^*\| \geq \sqrt{K - t} \min_{i \in \Omega} |\mathbf{x}^*(i)|, \quad (32)$$

we can give another sufficient condition:

$$\begin{aligned} \frac{1 - (2 + \sqrt{K})\delta_{K+1}}{1 - \delta_{K+1}} \min_{i \in \Omega} |\mathbf{x}^*(i)| & \geq \frac{2\delta_{K+1}(1 - \delta_{K+1})}{1 - \delta_{K+1}} E \\ & + \frac{1 - \delta_{K+1} + \sqrt{1 + \delta_{K+1}}(1 + \delta_{K+1})}{1 - \delta_{K+1}} \epsilon_1. \end{aligned} \quad (33)$$

With the assumption of (15) that $\delta_K < \frac{1}{2 + \sqrt{K - 1}}$, LHS of (33) is guaranteed to be positive. Therefore, we finally obtain a sufficient condition for (17):

$$\begin{aligned} \min_{i \in \Omega} |\mathbf{x}^*(i)| & \geq \\ & \frac{2\delta_{K+1}(1 - \delta_{K+1})E + \left(1 - \delta_{K+1} + (1 + \delta_{K+1})^{\frac{3}{2}}\right) \epsilon_1}{1 - (2 + \sqrt{K})\delta_{K+1}}. \end{aligned} \quad (34)$$

Thus, we complete the proof. \blacksquare

B. Gaussian Noise

Assume that the *i.i.d* noise in model (1) obeys that $e_i \sim N(0, \sigma^2)$. It has been shown in [18] that

$$P\left(\|\mathbf{e}\| \leq \sigma\sqrt{m + 2\sqrt{m \log m}}\right) \geq 1 - \frac{1}{m}. \quad (35)$$

By using the above conclusion together with Theorem 1, we obtain the following result.

Theorem 2. Suppose that the *i.i.d* noise in model (1) obeys $e_i \sim N(0, \sigma^2)$ and \mathbf{A} satisfies the RIP condition that

$$\delta_K < \frac{1}{2 + \sqrt{K - 1}} \quad (36)$$

in model (1). Define the differential vector $\mathbf{x}_e = \mathbf{x}^* - \boldsymbol{\mu}_{\Omega}$ and suppose that $\|\mathbf{x}_e\| \leq E$. Then CAMP with stopping rule $t = K$ will exactly recover the support Ω of the sparse signal \mathbf{x}^* with probability at least $1 - \frac{1}{m}$, if the minimum magnitude of nonzero elements of \mathbf{x}^* satisfies

$$\begin{aligned} \min_{i \in \Omega} |\mathbf{x}^*(i)| & \geq \frac{2\delta_{K+1}(1 - \delta_{K+1})}{1 - (2 + \sqrt{K})\delta_{K+1}} E \\ & + \frac{1 - \delta_{K+1} + (1 + \delta_{K+1})^{\frac{3}{2}}}{1 - (2 + \sqrt{K})\delta_{K+1}} \sigma\sqrt{m + 2\sqrt{m \log m}}. \end{aligned} \quad (37)$$

V. CONCLUSION

In this letter, we have studied the performance of Covariance-Assisted Matching Pursuit under the framework of RIP. A sufficient condition on RIP and sparse signals is given under l_2 bounded noise and Gaussian noise respectively to guarantee the exact support recovery of CAMP. As shown, CAMP is able to stay stable under noisy conditions. Though prior knowledge improves the performance of CAMP, the uncertainty caused by extra information increases as well. Thus, a sharp condition is required by CAMP to recover the sparse signal exactly. The accuracy of the estimation of prior knowledge influences the performance of CAMP to a great extent.

REFERENCES

- [1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [4] H. Zhang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Joint-structured-sparsity-based classification for multiple-measurement transient acoustic signals," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 6, pp. 1586–1598, 2012.
- [5] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse bayesian learning for recovery of block-sparse signals," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 360–372, 2014.
- [6] A. Adler, "Covariance-assisted matching pursuit," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 149–153, 2015.
- [7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [8] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, 2011.
- [9] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.
- [10] Q. Mo, "A sharp restricted isometry constant bound of orthogonal matching pursuit," *arXiv:1501.01708 [cs.IT]*, 2015.
- [11] Y. Shen and S. Li, "Sparse signals recovery from noisy measurements by orthogonal matching pursuit," *Inverse Problems and Imaging*, vol. 9, no. 1, pp. 231–238, 2015.
- [12] J. Lin and S. Li, "Nonuniform support recovery from noisy random measurements by orthogonal matching pursuit," *J. Approx. Theory*, vol. 165, pp. 20–40, 2013.
- [13] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [14] S. Foucart, "Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants," *Neamtu M., Schumaker L. (eds) Approximation Theory XIII: San Antonio 2010. Springer Proceedings in Mathematics*, vol. 13, pp. 65–77, 2012.
- [15] G. Boole, I. Grattan-Guinness, and G. Borner, *Elementary Treatise on Logic not Mathematical*, pp. 40–41. Birkhuser Verlag, Berlin, 1849.
- [16] L. Rencker, W. Wang, and M. D. Plumbley, "A greedy algorithm with learned statistics for sparse signal reconstruction," in *ICASSP*, 2017.
- [17] R. Wu, W. Huang, and D. R. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 403–406, 2012.
- [18] T. T. Cai and L. Wang, "Sparse signals recovery from noisy measurements by orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.