Matching Pursuit Covariance Learning

Esa Ollila
Department of Information and Communications Engineering
Aalto University
Espoo, Finland

***Abstract*—In the Multiple Measurements Vector (MMV) model, measurement vectors are connected to unknown, jointly sparse signal vectors through a linear regression model employing a single known measurement matrix (or dictionary). Typically, the number of atoms (columns of the dictionary) is greater than the number measurements and the sparse signal recovery problem is generally ill-posed. In this paper, we treat the signals and measurement noise as independent Gaussian random vectors with unknown signal covariance matrix and noise variance, respectively. Based on the negative log-likelihood function and maximum likelihood principle, we then introduce a matching pursuit covariance learning (CL) algorithm, analogous to popular orthogonal matching pursuit (OMP). Our numerical examples demonstrate effectiveness of the proposed CL strategy in sparse signal recovery where it performs favourably compared to the state-of-the-art algorithms under a broad variety of settings.**

***Index Terms*—compressed sensing, orthogonal matching pursuit, sparse Bayesian learning, sparse signal reconstruction**

I. Introduction

In the *multiple measurements vector* (MMV) [1] model, the measurement vectors $\mathbf{y}_i \in \mathbb{C}^N$ follow the generative model

$$\mathbf{y}_l = \mathbf{A}\mathbf{x}_l + \mathbf{e}_l, \quad l = 1, \ldots, L, \tag{1}$$

where $\mathbf{A} = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_M) \in \mathbb{C}^{N\times M}$ is a fixed (known) overcomplete matrix ($M > N$), called the *dictionary* or *measurement matrix*, $\mathbf{x}_l \in \mathbb{C}^M$ are unobserved source signal vectors, and $\mathbf{e}_l \in \mathbb{C}^N$ are unobserved zero mean white noise vectors, i.e., $\mathrm{cov}(\mathbf{e}_l) = \sigma^2\mathbf{I}$. The column vectors $\mathbf{a}_i$ of $\mathbf{A}$ are referred to as *atoms*. Letting $\mathbf{Y} = (\mathbf{y}_1 \cdots \mathbf{y}_L) \in \mathbb{C}^{N\times L}$ to denote the matrix of measurement vectors, we can write the model (1) in matrix form as

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E}, \tag{2}$$

where matrices $\mathbf{X} = (x_{ml}) \in \mathbb{C}^{M\times L}$ and $\mathbf{E} = (e_{nl}) \in \mathbb{C}^{N\times L}$ contain the signal and error vectors as columns, respectively. The signal matrix $\mathbf{X}$ is assumed to be K-rowsparse, *i.e.,* at most K rows of $\mathbf{X}$ contain non-zero entries. The rowsupport of $\mathbf{X} \in \mathbb{C}^{M\times L}$ is the index set of rows containing non-zero elements:

$$\mathcal{M} = \mathsf{supp}(\mathbf{X}) = \{i \in [\![M]\!] : x_{ij} \neq 0 \ \text{ for some } j \in [\![L]\!]\}$$

where $[\![M]\!] = \{1, \ldots, M\}$. In sparse signal recovery (SSR) problems, the object of interest is identifying the support $\mathcal{M}$, given only the data $\mathbf{Y}$, the dictionary $\mathbf{A}$, and the sparsity level K. See [1]–[6].

The source signal vectors $\mathbf{x}_l$ are modelled as i.i.d. circular Gaussian random vectors with independent elements and zero mean. Additionally, $\mathbf{x}_l$-s are assumed statistically independent of noise vectors $\mathbf{e}_l$, $l = 1, \ldots, L$, which are also assumed to have circular Gaussian distribution. Under the above conditions, $\mathbf{y}_l \sim \mathcal{CN}_N(\mathbf{0}, \boldsymbol{\Sigma})$, where the positive definite Hermitian (PDH) $N \times N$ covariance matrix $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{y}_l)$ is of the form

$$\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I} = \sum_{i=1}^{M} \gamma_i \mathbf{a}_i\mathbf{a}_i^{\mathsf{H}} + \sigma^2\mathbf{I}. \tag{3}$$

where the signal covariance matrix $\boldsymbol{\Gamma} = \mathrm{cov}(\mathbf{x}_l) = \mathrm{diag}(\boldsymbol{\gamma})$ is a diagonal matrix and $\boldsymbol{\gamma} \in \mathbb{R}^M_{\geq 0}$ is a vector of signal powers with only K non-zero elements. Hence, $\mathcal{M} = \mathsf{supp}(\mathbf{X}) = \mathsf{supp}(\boldsymbol{\gamma})$, and *covariance learning* (CL-)based support recovery algorithms can be constructed by minimizing the negative log-likelihood function (LLF) of the data $\mathbf{Y}$, which is given by (after scaling by $1/L$ and ignoring additive constants)

$$\begin{aligned}\ell(\boldsymbol{\gamma}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}) = \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\hat{\boldsymbol{\Sigma}}) \\ + \log|\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}|\end{aligned} \tag{4}$$

where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix (SCM),

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}_l\mathbf{y}_l^{\mathsf{H}} = L^{-1}\mathbf{Y}\mathbf{Y}^{\mathsf{H}},$$

and $\mathrm{tr}(\cdot)$ and $|\cdot|$ denote the matrix trace and determinant, respectively.

The objective is to determine the minimizer of (4) along with the identification of the support of non-zero source powers. A popular Bayesian approach, sparse Bayesian learning (SBL) [7], regards signal powers (γ_i) as random hyperparameters governed by a hierarchical prior distribution. Popular M-SBL algorithm [8], [9] employs an empirical Bayes method to construct an EM algorithm for solving (4). M-SBL method is computationally demanding, and is known to have slow convergence, which often prohibits its uses even when the number of atoms M is only moderately large. Although some faster fixed point SBL update rules have been proposed (*e.g.*, [8, eq. (19)]) these often provide worse performance than the original slower rule. The greedy CL algorithm proposed in this paper avoid estimation of $\mathbf{X}$, thus distinguishing itself from methods such as M-SBL, the simultaneous orthogonal matching pursuit (SOMP) [2] or the simultaneous normalized iterative hard thresholding (SNIHT) [6] algorithms, all of which require iterative updates of the unknown source signal matrix. Another disadvantage of M-SBL is the assumption that σ^2 is known. Although the M-step of M-SBL can be modified

for joint estimation of $\boldsymbol{\gamma}$ and σ^2, as mentioned in [7], [10], they tend to provide poor estimation results, which is also attested in our numerical studies. Additionally, the estimate of $\boldsymbol{\gamma}$ is not necessarily very sparse, but is post-processed by pruning (setting to zero all but K-largest elements of final iterate $\hat{\boldsymbol{\gamma}}$). In this paper we design a greedy pursuit CL algorithm, called CL-OMP (*cf.* Section II) that is analogous to SOMP in the classic SSR setting with deterministic source waveforms. Simulation studies (*cf.* Section III) then validate the effectiveness of the proposed estimation scheme.

Notations: $\mathbf{a}_{\mathcal{M}}$ denotes the components of $\mathbf{a}$ corresponding to support set $\mathcal{M}$ with $|\mathcal{M}| = K$. For any $n \times m$ matrix $\mathbf{A}$ we denote by $\mathbf{A}_{\mathcal{M}}$ the $n \times K$ submatrix of $\mathbf{A}$ restricted to the columns of $\mathbf{A}$ indexed by set $\mathcal{M}$. We use $\mathbf{A}^+ = (\mathbf{A}^{\mathsf{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{H}}$ to denote the pseudo-inverse of matrix $\mathbf{A}$ with full column rank. We use $\mathrm{diag}(\cdot)$ in two distinct contexts: $\mathrm{diag}(\mathbf{A})$ denotes an N-vector comprising the diagonal elements of an $N \times N$ matrix $\mathbf{A}$, whereas $\mathrm{diag}(\mathbf{a})$ signifies an $N \times N$ diagonal matrix with its diagonal elements from the N-vector $\mathbf{a}$.

II. Greedy pursuit covariance learning

Define

$$\boldsymbol{\Sigma}_{\setminus i} = \sum_{j \neq i} \gamma_j \mathbf{a}_j \mathbf{a}_j^{\mathsf{H}} + \sigma^2 \mathbf{I} = \boldsymbol{\Sigma} - \gamma_i \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}} \tag{5}$$

as the covariance matrix of $\mathbf{y}_l$-s when the contribution from the i^{th} source signal is removed. Before presenting the CL-OMP algorithm, we recall the following results which will be used in the sequel.

Lemma 1. *[11] Assume $M = K < N$. Then the parameters $\sigma^2 > 0$ and $\boldsymbol{\gamma} \in \mathbb{R}^K_{\geq 0}$ that minimize $\ell(\boldsymbol{\gamma}, \sigma^2 | \mathbf{Y}, \mathbf{A})$ in (4) are*

$$\hat{\sigma}^2 = \frac{1}{N-K} \mathrm{tr}\big((\mathbf{I} - \mathbf{A}\mathbf{A}^+)\hat{\boldsymbol{\Sigma}}\big) \tag{6}$$

and

$$\hat{\boldsymbol{\gamma}} = \mathrm{diag}(\mathbf{A}^+(\hat{\boldsymbol{\Sigma}} - \hat{\sigma}^2\mathbf{I})\mathbf{A}^{+\mathsf{H}}) \tag{7}$$

where the latter represents an unconstrained MLE, meaning it aligns with MLE when the non-negativity constraint is met, i.e., $\hat{\boldsymbol{\gamma}} \in \mathbb{R}^K_{\geq 0}$.

In scenarios where $\hat{\boldsymbol{\gamma}}$ in (7) contains negative elements, one can calculate the constrained solution using [12, Algorithm I]. A more straightforward method involves setting the negative elements of $\hat{\boldsymbol{\gamma}}$ to zero. While this approach does not precisely yield the MLE, it remains consistent in large samples.

Lemma 2. *[13, Appendix B] Consider the conditional likelihood of (4) where source powers γ_j for $j \neq i$ and the noise variance σ^2 are known. Then the conditional negative log-likelihood function for the unknown i^{th} source, defined as*

$$\begin{aligned}\ell_i(\gamma \mid \mathbf{Y}, \mathbf{A}, \{\gamma_j\}_{j\neq i}, \sigma^2) \\ = \mathrm{tr}((\boldsymbol{\Sigma}_{\setminus i} + \gamma \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}})^{-1}\hat{\boldsymbol{\Sigma}}) + \log|\boldsymbol{\Sigma}_{\setminus i} + \gamma \mathbf{a}_i \mathbf{a}_i^{\mathsf{H}}|,\end{aligned} \tag{8}$$

has a unique optimal value

$$\gamma_i = \max\left(\frac{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}_{\setminus i}^{-1}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_{\setminus i})\boldsymbol{\Sigma}_{\setminus i}^{-1}\mathbf{a}_i}{(\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}_{\setminus i}^{-1}\mathbf{a}_i)^2}, 0\right), \tag{9}$$

where $\boldsymbol{\Sigma}_{\setminus i}$ is defined in (5).

Lemma 2 will serve as the foundation for developing the CL-based Orthogonal Matching Pursuit (CL-OMP) algorithm, presented in algorithm 1. This algorithm follows the generic matching pursuit strategy similar to the one outlined in [14, Table 3.1].

Initialization phase. Set $k = 0$, and $\boldsymbol{\gamma}^{(0)} = \mathbf{0}_{M\times 1}$, $\sigma^{2(0)} = [\mathrm{tr}(\hat{\boldsymbol{\Sigma}})/N]\mathbf{I}$, and $\mathcal{M}^{(0)} = \mathsf{supp}(\boldsymbol{\gamma}^{(0)}) = \emptyset$ as initial solutions of signal and noise powers and the signal support, respectively. Then $\boldsymbol{\Sigma}^{(0)} = \mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma}^{(0)})\mathbf{A}^{\mathsf{H}} + \sigma^{2(0)}\mathbf{I} = \sigma^{2(0)}\mathbf{I}$ is the initial covariance matrix at the start of iterations.

Main Iteration phase consists of the following steps:

1) Sweep: Compute the errors

$$\epsilon_i = \min_{\gamma \geq 0} \ell_i(\gamma \mid \mathbf{Y}, \mathbf{A}, \{\gamma_j^{(k)}\}_{j\neq i}, \sigma^{2(k)}) \tag{10}$$

for each $i \in [\![M]\!] \setminus \mathcal{M}^{(k)}$ using its unique optimal value as given by Lemma 2:

$$\gamma_i = \max\left(\frac{\mathbf{a}_i^{\mathsf{H}}(\boldsymbol{\Sigma}^{(k)})^{-1}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^{(k)})(\boldsymbol{\Sigma}^{(k)})^{-1}\mathbf{a}_i}{(\mathbf{a}_i^{\mathsf{H}}(\boldsymbol{\Sigma}^{(k)})^{-1}\mathbf{a}_i)^2}, 0\right). \tag{11}$$

The sweep stage is comprised of lines 2 and 3 in algorithm 1.

2) Update support: Find a minimizer, i_k of ϵ_i: $\forall i \notin \mathcal{M}^{(k)}$, $\epsilon_{i_k} \leq \epsilon_i$, and update the support $\mathcal{M}^{(k+1)} = \mathcal{M}^{(k)} \cup \{i_k\}$. This step corresponds to line 4 in algorithm 1.

3) Update provisional solution: compute

$$(\hat{\mathbf{g}}, \hat{\sigma}^2) = \arg\min_{\mathbf{g}, \sigma^2} \ell(\mathbf{g}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}_{\mathcal{M}^{(k+1)}}),$$

where ℓ is the LLF defined in (4). These solutions are obtained from Lemma 1 and calculated in lines $5-7$ in algorithm 1, respectively. The obtained signal power is constrained to be non-negative, which is not exactly the MLE. Alternative option is to compute the true MLE (the constrained solution) using [12, Algorithm I].

4) Update the covariance matrix: Compute $\boldsymbol{\Sigma}^{(k+1)} = \mathbf{A}_{\mathcal{M}}\,\mathrm{diag}(\hat{\mathbf{g}})\mathbf{A}_{\mathcal{M}}^{\mathsf{H}} + \hat{\sigma}^2\mathbf{I}$. This stage is implemented by line 8 in algorithm 1.

5) Stopping rule: stop after K iterations, and otherwise increment k by 1 and repeat steps 1)-4).

Note that the sweep stage gives (after some simple algebra) the following value for the error ϵ_i in (10):

$$\epsilon_i = c + \log(1 + \gamma_i \mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i) - \gamma_i \frac{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i}{1 + \gamma_i \mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i} \tag{12}$$

where we have for simplicity of notation written $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{(k)}$ and where c denotes an irrelevant constant that is not dependent on γ_i. Thus without loss of generality, we set $c = 0$. Then using that γ_i is given by (11) we can write

$$\begin{aligned}1 + \gamma_i \mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i &= 1 + \frac{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{a}_i}{(\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i)^2}\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i \\ &= 1 + \frac{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i - \mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i}{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i} \\ &= \frac{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i}{\mathbf{a}_i^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\mathbf{a}_i}\end{aligned}$$

in the case that $\gamma_i > 0$. Substituting this into the denominator of the last term in (12), we obtain

$$\epsilon_i = \log(1 + \gamma_i \mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{a}_i) - \gamma_i \mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{a}_i. \tag{13}$$

If $\gamma_i = 0$, then $\epsilon_i = 0$. This explains line 3 in algorithm 1.

In the provisional solution update, one finds the minimizer of $\ell(\boldsymbol{\gamma}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}_{\mathcal{M}})$ where for notational simplicity we have written $\mathcal{M} = \mathcal{M}^{(k+1)}$ for the current support with $|\mathcal{M}| = k$. Thus the problem to be solved is

$$\underset{\boldsymbol{\gamma} \in \mathbb{R}^k_{\geq 0}, \sigma^2 > 0}{\text{minimize}} \; \operatorname{tr}((\mathbf{A}_{\mathcal{M}} \operatorname{diag}(\boldsymbol{\gamma}) \mathbf{A}_{\mathcal{M}}^{\mathsf{H}} + \sigma^2 \mathbf{I})^{-1} \hat{\boldsymbol{\Sigma}}) + \log |\mathbf{A}_{\mathcal{M}} \operatorname{diag}(\boldsymbol{\gamma}) \mathbf{A}_{\mathcal{M}}^{\mathsf{H}} + \sigma^2 \mathbf{I}|.$$

Taking the derivative of this equation w.r.t. γ_i and setting it to zero, gives the likelihood equation:

$$0 = \mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \mathbf{a}_i \quad \forall i \in \mathcal{M}.$$

This in turn implies that in the next iteration γ_i in (11) will take value $\gamma_i \approx 0$ for $i \in \mathcal{M}$. Thus it is unlikely that the atoms that have been already selected will be chosen again in the next iterations. This explains the name of this greedy pursuit algorithm: the method is similar to conventional OMP [15] or SOMP [2] where no atom is ever chosen twice in the sweep stage.

Algorithm 1: CL-OMP algorithm

Input : $\mathbf{Y}$, $\mathbf{A}$, K

Initialize: $\hat{\boldsymbol{\Sigma}} = L^{-1}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, $\boldsymbol{\Sigma} = [\operatorname{tr}(\hat{\boldsymbol{\Sigma}})/p]\mathbf{I}$, $\mathcal{M} = \emptyset$

1 **for** $k = 1, \ldots, K$ **do**

2 $\quad \boldsymbol{\gamma} = (\gamma_i)_{M\times 1}$, $\gamma_i \leftarrow \max\left(\frac{\mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \mathbf{a}_i}{(\mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{a}_i)^2}, 0\right)$

3 $\quad \boldsymbol{\epsilon} = (\epsilon_i) \leftarrow \left(\log(1 + \gamma_i \mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{a}_i) - \gamma_i \mathbf{a}_i^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} \mathbf{a}_i\right)_{M\times 1}$

4 $\quad \mathcal{M} \leftarrow \mathcal{M} \cup \{i_k\}$ with $i_k \leftarrow \arg\min_{i \notin \mathcal{M}} \epsilon_i$

5 $\quad \sigma^2 \leftarrow \frac{1}{N-k} \operatorname{tr}\left((\mathbf{I} - \mathbf{A}_{\mathcal{M}} \mathbf{A}_{\mathcal{M}}^{+}) \hat{\boldsymbol{\Sigma}}\right)$

6 $\quad \boldsymbol{\gamma}_{\mathcal{M}} \leftarrow \max\left(\operatorname{diag}\left(\mathbf{A}_{\mathcal{M}}^{+} (\hat{\boldsymbol{\Sigma}} - \sigma^2 \mathbf{I}) \mathbf{A}_{\mathcal{M}}^{+\mathsf{H}}\right), 0\right)$

7 $\quad \boldsymbol{\gamma}_{\mathcal{M}^{\complement}} \leftarrow \mathbf{0}$

8 $\quad \boldsymbol{\Sigma} \leftarrow \mathbf{A} \operatorname{diag}(\boldsymbol{\gamma}) \mathbf{A}^{\mathsf{H}} + \sigma^2 \mathbf{I}$

Output : $\mathcal{M}$, $\boldsymbol{\gamma}$, σ^2

III. Simulation studies

A. Simulation setting

We compare the performance of CL-OMP algorithm against the traditional greedy SSR algorithms, the **SOMP** [2, Algorithm 3.1] and the **SNIHT** [6, Algorithm 1]. Both SOMP and SNIHT algorithms are designed for MMV model, and return an estimated support $\mathcal{M}$ of K-rowsparse signal matrix given the measurement matrix $\mathbf{Y}$, the dictionary $\mathbf{A}$, and the desired rowsparsity level K.

We also compare with other CL algorithms designed for solving (4). We note that comparison to M-SBL is challenging as it assumes that σ^2 is known. However, update rules for σ^2 can be integrated into the EM algorithm by replacing the M-step with a joint maximization over σ^2 and $\boldsymbol{\gamma}$, resulting in the addition of the σ^2 update [8, Eq. (21)] to the M-step. In [8], it was noted that joint estimation can result in severely biased estimates of σ^2, thereby affecting the accuracy of $\boldsymbol{\gamma}$ estimate. Consequently, the authors in [8] recommend using alternatives such as grid search or other heuristics to estimate σ^2. For fair comparison, we report values of M-SBL that utilize the true σ^2, denoted as **M-SBL-σ^2**, while the version that jointly estimates both σ^2 and $\boldsymbol{\gamma}$ is termed **M-SBL**. We also compare with coordinatewise optimization (CWO) method of [16, Algorithm 1]. This algorithm also assumes that the noise power σ^2 in the LLF (4) is known. Hence we denote the method as **CWO-σ^2**, where again the suffix σ^2 is used to indicate that the method has oracle knowledge of σ^2. Naturally, M-SBL-σ^2 and CWO-σ^2 are not realisable in practise in our scenario as σ^2 is unknown and their performance is only shown as benchmark for best possible performance.

In our simulations, the matrix $\mathbf{A}$ is a Gaussian random measurement matrix, i.e., the elements of $\mathbf{A}$ are drawn from $\mathcal{CN}(0, 1)$ distribution and the columns are unit-norm normalised as is common in compressed sensing. To form the K-rowsparse source signal matrix $\mathbf{X} \in \mathbb{C}^{M\times L}$, support $\mathcal{M} = \operatorname{supp}(\mathbf{X})$ is randomly chosen from $\{1, \ldots, M\}$ without replacement for each Monte-Carlo (MC-)trial. The noise $\mathbf{E} \in \mathbb{C}^{N\times L}$ have elements that are i.i.d. circular complex Gaussian with unit variance (i.e., $\sigma^2 = 1$).

As performance measure we use the empirical *probability of exact recovery*, PER $= \frac{1}{T}\sum_{t=1}^{T} \mathrm{I}\big(\hat{\mathcal{M}}^{(t)} = \mathcal{M}^{(t)}\big)$, where $\mathrm{I}(\cdot)$ denotes the indicator function, and $\hat{\mathcal{M}}^{(t)}$ denotes the estimate of the true signal support $\mathcal{M}^{(t)}$ for t^{th} MC trial. The number of MC trials is $T = 2000$, the number of atoms is $M = 256$, and the sparsity level is $K = 4$. The dimensionality of the measurements N or the number of measurement vectors L may vary. Let $\mathcal{M} = \{i_1, \ldots, i_K\}$ be the true support set where $K = 4$, and let $\sigma_1^2 = \gamma_{i_1}$ denote the power of the 1^{st} Gaussian source signal. Define the SNR of the first non-zero source sequence as $10 \log_{10} \sigma_1^2/\sigma^2$. For the second, third, and fourth Gaussian sources, we set their SNR levels to be 1 dB, 2 dB, and 4 dB lower than that of the first source, respectively.

B. Results

The upper panel of Figure 1 displays the PER rates when the number of measurement vectors is fixed ($L = 16$) while the dimensionality of measurements N grows ($N = 8, 16, 32$ to $N = 64$). As can be noted, CL-OMP has the best performance for all N and SNR levels. For example, at SNR $=$ 5dB and $N = 32$, the PER rates are 60.1%, 47.2%, 54.8%, 32.8%, 57.45% and 56.55% for the CL-OMP, SOMP, SNIHT, M-SBL, M-SBL-σ^2 and CWO-σ^2, respectively. Note that M-SBL (which estimates σ^2 in M-step) has 25% drop in accuracy compared with M-SBL-σ^2 (that uses true σ^2) illustrating the instability of estimating $\boldsymbol{\gamma}$ and σ^2 jointly in the EM algorithm. It can also be noted that the performance of M-SBL deteriorates as N increases. SNIHT has better performance compared to SOMP, yet SOMP gradually improves, approaching the

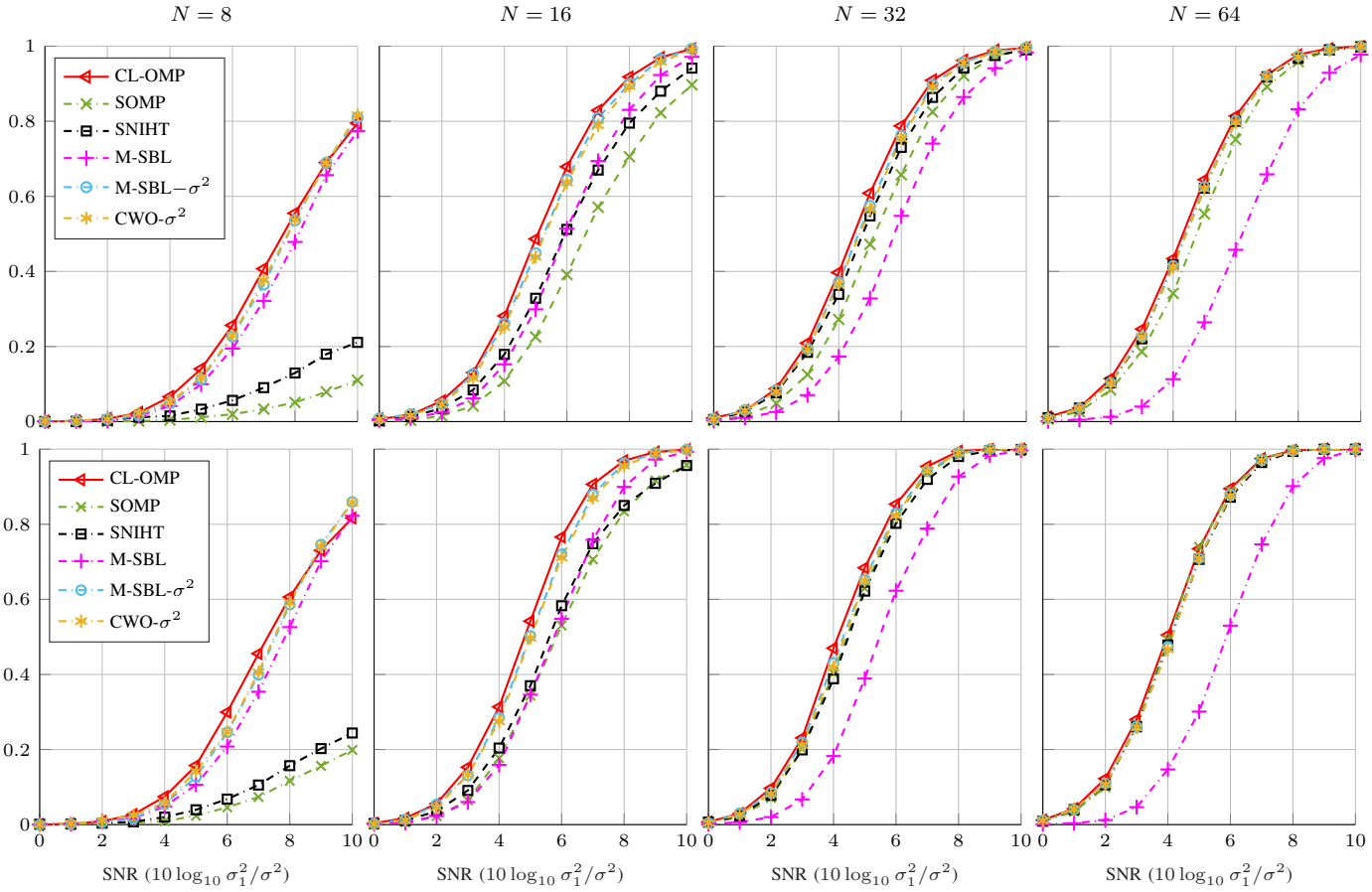

Fig. 1. Comparison of PER rates of different sparse recovery algorithms under a variety of SNR levels when the sources are Gaussian (top panel) and non-Gaussian (bottom panel) and the number of measurements increases from $N = 8$ to $N = 64$; $K = 4$, $M = 256$, $L = 16$ and dictionary is a Gaussian measurement matrix.

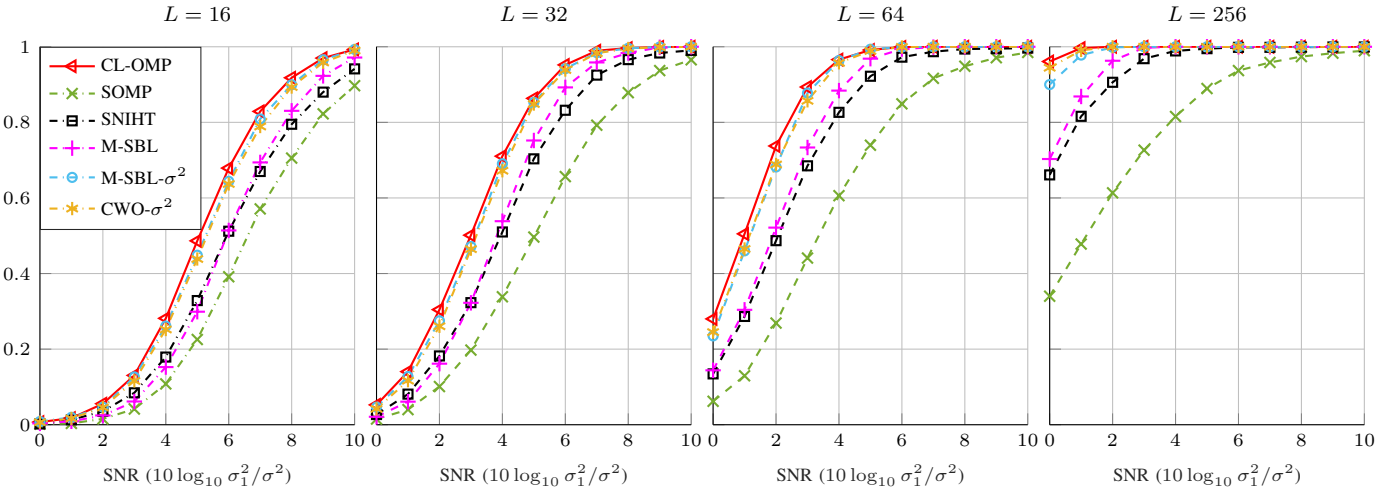

Fig. 2. Comparison of PER rates of different sparse recovery algorithms under a variety of SNR levels when the sources are Gaussian and the number of snapshots increases from $L = 16$ to $L = 256$; $N = 16$, $K = 4$, $M = 256$, and the dictionary is a Gaussian measurement matrix.

performance level of SNIHT as the number of measurements N increases. When comparing M-SBL-σ^2 and CWO-σ^2 we notice that M-SBL-σ^2 often has slight advantage over CWO-σ^2 but this comes with extra computational burden.

Next, we explore the impact of the Gaussianity of source signals on method performances. We let the signal sequences

x_{il} possess a fixed amplitude $\gamma_i = |x_{il}|^2 > 0$ for all $l = 1, \ldots, L$, where $\gamma_i > 0$ for $i \in \mathcal{M}$ and $\gamma_i = 0$ otherwise. The phases of source signals are generated as random variables from a uniform distribution on $[0, 2\pi)$, i.e., $\mathrm{Arg}(x_{il}) \sim Unif(0, 2\pi)$. In other words, $\{x_{il}\}_{l=1}^{L}$ for $i \in \mathcal{M}$ is an i.i.d sample from a uniform distribution on the complex circle of fixed radius γ_i. The bottom panel of Figure 1 depicts the performance in this scenario. A comparison between the upper and bottom panel figures reveals that the performance of all CL methods (CL-OMP, M-SBL or CWO) remain consistent even with non-Gaussian sources. This illustrates some robustness of the CL-methods against the violation of the Gaussianity assumption of source signals.

In the next study, dimensionality N is fixed ($N = 16$), while the number of measurement vectors L varies. Otherwise the setting is as earlier. Again we note from Figure 2 that CL-OMP uniformly outperforms the other methods for all SNR levels and sample lengths. Consider, for instance, the scenario with $L = 256$ and a very low SNR of 1 dB. In this case, CL-OMP stands out as the only method achieving a perfect 100% PER rate. Meanwhile, SNIHT and SOMP exhibit PER rates of 81.6% and 47.8%, respectively. The PER rates for M-SBL, M-SBL-σ^2 and CWO-σ^2 are 86.8%, 97.8%, 99.0%, respectively. Overall (for all values of L) CL-OMP is distinctively more robust to low SNR than conventional greedy methods, the SOMP and the SNIHT. CL-OMP is also consistently better than M-SBL-σ^2 and CWO-σ^2 which are the only methods that have oracle knowledge of the noise power σ^2. Another interesting observation is that as L increases the performance gap between the CL methods and the traditional greedy pursuit methods (the SNIHT and SOMP) becomes slightly larger. This can be due to the fact that as L increases the SCM $\hat{\mathbf{\Sigma}}$ better estimates the true covariance matrix $\mathbf{\Sigma}$ giving an additional boost to CL methods. We repeated the simulation for non-Gaussian sources scenario, but the results were similar as for Gaussian sources case, and are omitted.

IV. Conclusions

In this paper, we proposed a matching pursuit CL algorithm. Our simulation studies demonstrated that the proposed CL-OMP method outperforms traditional SSR methods, the SOMP and the SNIHT, often with a large margin. Especially, when N or L are small, the SNIHT and the SOMP performed poorly compared to CL methods. Remarkably, CL-OMP also outperformed the CWO(-σ^2) and M-SBL(-σ^2) algorithms which minimize the LLF in (4) under the assumption that σ^2 is known.

There are many open questions to be addressed in future works. For example, estimation of number of sources K or addressing the non-Gaussian scenarios. A journal version [17] of this paper also considers the source localization problem where the method performs favourably compared to competing methods. There are also several applications where the proposed CL-OMP can be useful. For instance, it can be used as atom selection method in coupled dictionary learning algorithm [18]. These explorations are left as future works.

References

[1] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, 2011.

[2] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, 2006.

[3] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: convex relaxation," *Signal Processing*, vol. 86, pp. 589–602, 2006.

[4] 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.

[5] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519, 2010.

[6] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jin, "Greedy algorithms for joint sparse recovery," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1694 – 1704, 2014.

[7] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[8] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[9] D. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 1023–1030.

[10] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[11] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," *Circuits, Systems and Signal Processing*, vol. 14, no. 5, pp. 669–674, 1995.

[12] Y. Bresler, "Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing," in *Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*. IEEE, 1988, pp. 172–175.

[13] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 425–443, 2010.

[14] M. Elad, *Sparse and redundant representations*. New York: Springer, 2010.

[15] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, pp. 57–98, 1997.

[16] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2925–2951, 2021.

[17] E. Ollila, "Sparse signal recovery and source localization via covariance learning," *arXiv preprint, arXiv:2401.13975 [stat.ME]*, 2024.

[18] F. G. Veshki, N. Ouzir, S. A. Vorobyov, and E. Ollila, "Multimodal image fusion via coupled feature learning," *Signal Processing*, vol. 200, p. 108637, 2022.