Majorization-Minimization Covariance Learning Algorithm for Sparse Signal Recovery

Majdoddin Esfandiari*, Esa Ollila*, and Daniel P. Palomar†
*Department of Information and Communications Engineering, Aalto University, Espoo, Finland
†Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong
Email: majdoddin.esfandiari@aalto.fi, esa.ollila@aalto.fi, palomar@ust.hk

***Abstract*—In the multiple measurement vector (MMV) model, sparse signal recovery (SSR) is formulated as identifying the common support set of sparse signal vectors using a collection of measurement vectors generated through a common known overcomplete dictionary. The sparse signal vectors or their powers may be estimated along with the support set. In this paper, we adopt a covariance learning (CL) perspective and focus on estimating the sparse signal powers and their support from the second-order statistics of the measurements. Building on the majorization–minimization (MM) framework, we introduce a CL-based algorithm, named CLMM, which forms a surrogate for the negative log-likelihood function of the measurements. An iterative scheme is then derived to estimate the sparse signal powers. Numerical simulations demonstrate the effectiveness of the proposed CLMM method in SSR, where it consistently outperforms state-of-the-art algorithms, particularly in the face of large sparsity ratios.**

***Index Terms*—Covariance learning, compressed sensing, sparse Bayesian learning, majorization-minimization, sparse signal recovery**

I. Introduction

With significant applications ranging from wireless communications [1]–[3] and biomedical imaging [4] to source localization [5], [6], sparse signal recovery (SSR) has emerged as a fundamental problem in signal processing. In the multiple measurements vector (MMV) case [7], the SSR problem is formulated as estimating $\mathbf{x}_l$-s in:

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

where $\mathbf{y}_l \in \mathbb{C}^M$ is the l-th measurement vector, $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N] \in \mathbb{C}^{M\times N}$ is a known overcomplete matrix (dictionary) ($N > M$) whose columns are referred to as atoms, and $\mathbf{e}_l \in \mathbb{C}^M$ is a zero mean white noise random vector with $\mathrm{cov}(\mathbf{e}_l) = \sigma^2\mathbf{I}$.[1] Additionally, $\mathbf{x}_l \in \mathbb{C}^N$ represents the unobserved sparse random signal vector with K non-zero elements. Letting $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{C}^{M\times L}$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_L] \in \mathbb{C}^{N\times L}$, and $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_L] \in \mathbb{C}^{M\times L}$, the measurement model in (1) can be written in matrix form as

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

The signal vectors $\mathbf{x}_l$-s are assumed statistically independent of the noise vectors $\mathbf{e}_l$-s. They are modeled as zero-mean, parametrized circular Gaussian random vectors with covariance matrix $\mathbf{\Gamma} = \mathrm{cov}(\mathbf{x}_l) = \mathrm{diag}(\boldsymbol{\gamma})$, i.e., $\mathbf{x}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Gamma})$. Here, $\mathrm{diag}(\cdot)$ constructs a diagonal matrix with its diagonal elements from the bracketed vector, and $\boldsymbol{\gamma} \in \mathbb{R}^N_{\geq 0}$ is the vector of the signal powers with K non-zero entries. Consequently, the index set of non-zero rows of $\mathbf{X}$ coincides with the index set of non-zero entries of $\boldsymbol{\gamma}$. This index set, called signal support, is defined as:

$$\begin{aligned}\mathcal{S} &= \mathsf{supp}(\mathbf{X}) = \{i \in [\![N]\!] : x_{ij} \neq 0 \ \text{ for some } j \in [\![L]\!]\} \\ &= \mathsf{supp}(\boldsymbol{\gamma}) = \{i \in [\![N]\!] : \gamma_i \neq 0\}\end{aligned}$$

where $[\![N]\!] = \{1, \dots, N\}$. Under these assumptions, $\mathbf{y}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$, where

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

A class of algorithms known as covariance learning (CL) methods aims to identify the signal support $\mathcal{S}$ and estimate the sparse signal powers $\boldsymbol{\gamma}$ by minimizing the negative log-likelihood function (LLF) of the data $\mathbf{Y}$, that is,

$$\begin{aligned}\underset{\boldsymbol{\gamma},\sigma^2}{\text{minimize}} \quad & \ell(\boldsymbol{\gamma}, \sigma^2) \\ \text{subject to} \quad & \boldsymbol{\gamma} \geq \mathbf{0}, \quad \sigma^2 > 0,\end{aligned} \tag{4}$$

where

$$\begin{aligned}\ell(\boldsymbol{\gamma}, \sigma^2) \equiv \ell(\boldsymbol{\gamma}, \sigma^2 \mid \mathbf{Y}, \mathbf{A}) &= \mathrm{tr}((\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})^{-1}\mathbf{S}) \\ &+ \mathrm{logdet}(\mathbf{A}\,\mathrm{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I})\end{aligned} \tag{5}$$

with $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ respectively denoting the matrix trace and determinant, and $\mathbf{S}$ being the sample covariance matrix (SCM) defined by

$$\mathbf{S} = \frac{1}{L}\sum_{l=1}^{L} \mathbf{y}_l\mathbf{y}_l^{\mathsf{H}}\ .$$

Note that the first term in (5) is convex but the second term is concave, making the problem nonconvex.

Sparse Bayesian learning (SBL) [8] is a widely used framework for SSR. The well-known M-SBL algorithm [9]

This work was supported in part by the Research Council of Finland under grant 368630, and in part by the Hong Kong GRF 16206123 research grant.

[1]Depending on the specific application, the structure of $\mathbf{A}$ varies. For example, in the direction-of-arrival estimation problem, the columns of $\mathbf{A}$ can be the array steering vectors corresponding to N uniformly spaced angular grid points. We do not impose any particular structural assumptions on $\mathbf{A}$ in this paper, which makes the proposed method applicable to various signal processing problems.

employs an empirical Bayesian strategy and derives an expected maximization (EM) procedure to solve (5). However, M-SBL is computationally intensive and typically suffers from slow convergence, making it impractical even for moderately large numbers of atoms N. Although faster fixed-point update rules for SBL have been proposed (e.g., [9, eq. (19)]), they often come with a performance degradation compared to the original, slower method. Other approaches for solving (5) include EM-based schemes [10] and methods based on generalized approximate message passing (GAMP) [11]. In [12], the majorization-minimization (MM) framework [13] has been used to jointly estimate source variances and full-structure noise covariances, called full-structure noise (FUN) learning, particularly in applications like brain source imaging that majorizes the log-determinant term via Taylor expansion while decomposing the trace term using posterior expectations in an EM-like manner. Greedy pursuit methods are another class of algorithms for solving SSR. Popular methods of this class include simultaneous normalized iterative hard thresholding (SNIHT) [14], simultaneous orthogonal matching pursuit (SOMP) [15], and least-squares orthogonal matching pursuit (LSOMP) [16]. In addition, in [17] and [18], greedy CL-based methods respectively referred to as CL-OMP and light-weight sequential SBL (LWS-SBL) have been developed.

In this paper, we exploit MM framework to develop a CL-based method for SSR that estimates $\boldsymbol{\gamma}$ assuming σ^2 is known. The proposed method, named CLMM, first constructs a surrogate optimization problem by forming appropriate tangent upper bounds (majorizers) for both the convex (trace term) and concave (log-determinant term) components of the objective function in (5).[2] It then uses an iterative scheme to estimate the sparse signal powers $\boldsymbol{\gamma}$ and the corresponding support set $\mathcal{S}$ by solving this surrogate optimization problem. Simulation results demonstrate that CLMM outperforms existing methods, particularly when the sparsity ratio K/M approaches 1.

II. The proposed method

To apply the MM framework [13], we first need to derive a suitable majorizer (tangent upper bound) for the objective function in (5). Although the first term of the objective is already convex, it is still useful to majorize it in order to obtain a closed-form update rule at each iteration. The second term is concave and can be conveniently majorized via a simple linearization.

A. Majorizer of the convex term of (5)

We begin by presenting two useful observations in Lemma 1 and Corollary 1.

Lemma 1 (Matrix upper bound for the PD case)**.** *The following matrix inequality holds for $\boldsymbol{D} \succ \mathbf{0}$ and $\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^{\mathsf{H}} \succ \mathbf{0}$,*

$$\left(\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^{\mathsf{H}}\right)^{-1} \preceq \left(\mathbf{F}^{(k)}\right)^{-1} \boldsymbol{A}\boldsymbol{D}^{(k)}\boldsymbol{D}^{-1}\boldsymbol{D}^{(k)}\boldsymbol{A}^{\mathsf{H}} \left(\mathbf{F}^{(k)}\right)^{-1} \quad (6)$$

with equality satisfied at $\boldsymbol{D} = \boldsymbol{D}^{(k)}$, where $\mathbf{F}^{(k)} = \boldsymbol{A}\boldsymbol{D}^{(k)}\boldsymbol{A}^{\mathsf{H}}$.

[2]Unlike FUN learning that only majorizes the concave term.

Proof. See [19, Proposition 4] and [13, Example 19]. □

Corollary 1 (Matrix upper bound for positive powers)**.** *Suppose $\boldsymbol{\Sigma} = \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I} \succ \mathbf{0}$, with $\boldsymbol{\gamma} > \mathbf{0}$ and $\sigma^2 > 0$. Then,*

$$\left(\boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}\right)^{-1} \preceq \tilde{\mathbf{G}}^{(k)}\tilde{\boldsymbol{D}}^{-1}\left(\tilde{\mathbf{G}}^{(k)}\right)^{\mathsf{H}}, \quad (7)$$

where

$$\tilde{\mathbf{G}}^{(k)} = \left(\boldsymbol{\Sigma}^{(k)}\right)^{-1}\left[\boldsymbol{A}\,\mathrm{diag}\left(\boldsymbol{\gamma}^{(k)}\right) \quad \sigma^{2(k)}\boldsymbol{I}\right]$$
$$\boldsymbol{\Sigma}^{(k)} = \boldsymbol{A}\,\mathrm{diag}\left(\boldsymbol{\gamma}^{(k)}\right)\boldsymbol{A}^{\mathsf{H}} + \sigma^{2(k)}\boldsymbol{I}$$
$$\tilde{\boldsymbol{D}} = \begin{bmatrix}\mathrm{diag}(\boldsymbol{\gamma}) & \mathbf{0}\\ \mathbf{0} & \sigma^2\boldsymbol{I}\end{bmatrix}.$$

Equality holds at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(k)}$ and $\sigma^2 = \sigma^{2(k)}$.

Proof. Lemma 1 can be applied to the matrix $\left(\boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}\right)^{-1}$ by defining $\tilde{\boldsymbol{A}} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}\end{bmatrix}$ since $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}\tilde{\boldsymbol{A}}^{\mathsf{H}} = \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}$. □

The assumptions made in Corollary 1 do not completely match the constraints in (4) as we cannot assume $\boldsymbol{\gamma} > \mathbf{0}$ (only $\sigma^2 > 0$ can be assumed). Note that by a limiting argument, the inequality in Lemma 1 still holds for $\boldsymbol{D} \succeq \mathbf{0}$ (and Corollary 1 for $\boldsymbol{\gamma} \geq \mathbf{0}$), but the upper bound becomes loose and not useful in practice. Fortunately, this bound can be improved because $\sigma^2 > 0$ as follows.

Corollary 2 (Matrix upper bound for nonnegative powers)**.** *Suppose $\boldsymbol{\Sigma} = \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I} \succ \mathbf{0}$, with $\boldsymbol{\gamma} \geq \mathbf{0}$ and $\sigma^2 > 0$. Define $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ as the Cholesky decomposition of $\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}} = \eta\boldsymbol{I} - \boldsymbol{A}\boldsymbol{A}^{\mathsf{H}} \succeq \mathbf{0}$ for some $\eta > \lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}})$, where $\lambda_{\max}(\cdot)$ computes the maximum eigenvalue of the bracketed matrix. Then,*

$$\left(\boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}\right)^{-1} \preceq \bar{\mathbf{G}}^{(k)}\bar{\boldsymbol{D}}^{-1}\left(\bar{\mathbf{G}}^{(k)}\right)^{\mathsf{H}}, \quad (8)$$

where

$$\bar{\mathbf{G}}^{(k)} = \left(\boldsymbol{\Sigma}^{(k)}\right)^{-1}\left[\boldsymbol{A}\,\mathrm{diag}\left(\bar{\boldsymbol{\gamma}}^{(k)}\right) \quad \bar{\sigma}^{2(k)}\boldsymbol{B}\right]$$
$$\bar{\boldsymbol{D}} = \begin{bmatrix}\mathrm{diag}(\bar{\boldsymbol{\gamma}}) & \mathbf{0}\\ \mathbf{0} & \bar{\sigma}^2\boldsymbol{I}\end{bmatrix}, \quad \bar{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \sigma^2/\eta,$$
$$\bar{\boldsymbol{\gamma}}^{(k)} = \boldsymbol{\gamma}^{(k)} + \sigma^{2(k)}/\eta, \quad \bar{\sigma}^2 = \sigma^2/\eta.$$

Observe that since $\sigma^2 > 0$, then $\bar{\sigma}^2 > 0$ and $\bar{\boldsymbol{\gamma}} > \mathbf{0}$. Equality holds at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{(k)}$ and $\sigma^2 = \sigma^{2(k)}$.

Proof. We want to have an expression similar to $\boldsymbol{\Sigma} = \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}$ but where we only need to assume $\boldsymbol{\gamma} \geq \mathbf{0}$. We can achieve that by absorbing part of σ^2 into $\boldsymbol{\gamma}$ as follows. From the relation of η and $\boldsymbol{B}$, we have $\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}} + \boldsymbol{B}\boldsymbol{B}^{\mathsf{H}} = \eta\boldsymbol{I}$. We can then write[3]

$$\begin{aligned}\boldsymbol{\Sigma} &= \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}\\ &= \boldsymbol{A}\,\mathrm{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \frac{\sigma^2}{\eta}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}} + \boldsymbol{B}\boldsymbol{B}^{\mathsf{H}})\\ &= \boldsymbol{A}\left(\mathrm{diag}(\boldsymbol{\gamma}) + \frac{\sigma^2}{\eta}\boldsymbol{I}\right)\boldsymbol{A}^{\mathsf{H}} + \frac{\sigma^2}{\eta}\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}},\end{aligned}$$

[3]The authors would like to thank Amirhossein Javaheri for a discussion on this idea.

which can then be used in Lemma 1 by defining:

$$\bar{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} \text{ and } \bar{\boldsymbol{D}} = \begin{bmatrix} \operatorname{diag}(\boldsymbol{\gamma}) + \frac{\sigma^2}{\eta}\boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2}{\eta}\boldsymbol{I} \end{bmatrix}.$$

□

Thus, our final matrix upper bound reads

$$\left(\boldsymbol{A}\operatorname{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}} + \sigma^2\,\boldsymbol{I}\right)^{-1} \preceq \bar{\mathbf{G}}^{(k)}\bar{\boldsymbol{D}}^{-1}\left(\bar{\mathbf{G}}^{(k)}\right)^{\mathsf{H}}, \quad (9)$$

where

$$\bar{\mathbf{G}}^{(k)} = \left(\boldsymbol{\Sigma}^{(k)}\right)^{-1}\left[\boldsymbol{A}\operatorname{diag}\left(\boldsymbol{\gamma}^{(k)} + \tfrac{\sigma^{2(k)}}{\eta}\right) \quad \tfrac{\sigma^{2(k)}}{\eta}\boldsymbol{B}\right] \quad (10)$$

$$\bar{\boldsymbol{D}} = \begin{bmatrix} \operatorname{diag}\left(\boldsymbol{\gamma} + \frac{\sigma^2}{\eta}\right) & \mathbf{0} \\ \mathbf{0} & \frac{\sigma^2}{\eta}\boldsymbol{I} \end{bmatrix}. \quad (11)$$

B. Majorizer of the concave term of (5)

Following [13, Example 2], the logdet term in (5) can be upper-bounded using its Taylor expansion as

$$\operatorname{logdet}(\boldsymbol{\Sigma}) \le \operatorname{tr}\left(\left(\boldsymbol{\Sigma}^{(k)}\right)^{-1}\boldsymbol{\Sigma}\right) + \text{const}, \quad (12)$$

where $\boldsymbol{\Sigma} = \mathbf{A}\operatorname{diag}(\boldsymbol{\gamma})\mathbf{A}^{\mathsf{H}} + \sigma^2\mathbf{I}$.

C. MM-based algorithm

Algorithm 1 CLMM algorithm

Input : $\mathbf{S}$, $\mathbf{A}$, σ^2, and $I_{\max}$.
Initialize : $\boldsymbol{\gamma}_{\text{init}} = \mathbf{0}$.

1 $\boldsymbol{\gamma}^{(0)} = \boldsymbol{\gamma}_{\text{init}}$, $\delta = 10^{-3}$.
2 $\eta = \operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)$.
for $k = 0, \ldots, I_{\max}$ **do**
3 $\quad \bar{\boldsymbol{\Sigma}}^{(k)} = \boldsymbol{A}\operatorname{diag}\left(\boldsymbol{\gamma}^{(k)}\right)\boldsymbol{A}^{\mathsf{H}} + \sigma^2\,\boldsymbol{I}$
4 $\quad \boldsymbol{M}^{(k)} = \left(\bar{\boldsymbol{\Sigma}}^{(k)}\right)^{-1}\boldsymbol{A}\operatorname{diag}\left(\boldsymbol{\gamma}^{(k)} + \frac{\sigma^2}{\eta}\right)$
5 $\quad$ For all $i \in \{1, \ldots, N\}$, compute:

$$T_i^{(k)} = \left(\boldsymbol{m}_i^{(k)}\right)^{\mathsf{H}}\boldsymbol{S}\boldsymbol{m}_i^{(k)} \quad \text{and} \quad U_i^{(k)} = \mathbf{a}_i^{\mathsf{H}}\left(\bar{\boldsymbol{\Sigma}}^{(k)}\right)^{-1}\mathbf{a}_i$$

6 $\quad$ For all $i \in \{1, \ldots, N\}$, compute:

$$\gamma_i^{(k+1)} = \left[\sqrt{\frac{T_i^{(k)}}{U_i^{(k)}}} - \frac{\sigma^2}{\eta}\right]_+$$

7 $\quad$ **if** $\frac{\left\|\boldsymbol{\gamma}^{(k+1)} - \boldsymbol{\gamma}^{(k)}\right\|_2}{\left\|\boldsymbol{\gamma}^{(k)}\right\|_2} < \delta$ **then**
$\quad\quad$ break
Output : $\boldsymbol{\gamma}$

The focus is to estimate $\boldsymbol{\gamma}$ for a known σ^2 in this paper, leaving joint estimation of $(\boldsymbol{\gamma}, \sigma^2)$ to future works. Towards this end, under the MM framework [13], one has to sequentially solve a surrogate problem obtaining the sequence of solutions $\boldsymbol{\gamma}^{(0)}, \boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, \ldots, \boldsymbol{\gamma}^{(I_{\max})}$ where $I_{\max}$ is the maximum number of iterations. At iteration $k+1$, exploiting (9)-(12), the surrogate problem of the original problem (4) with respect to (w.r.t) $\boldsymbol{\gamma}$ (constant terms are removed) is

$$\begin{aligned} \underset{\boldsymbol{\gamma}}{\text{minimize}} \quad & \operatorname{tr}\left(\boldsymbol{M}^{(k)}\left(\operatorname{diag}\left(\boldsymbol{\gamma} + \tfrac{\sigma^2}{\eta}\right)\right)^{-1}\left(\boldsymbol{M}^{(k)}\right)^{\mathsf{H}}\boldsymbol{S}\right) \\ & + \operatorname{tr}\left(\left(\bar{\boldsymbol{\Sigma}}^{(k)}\right)^{-1}\boldsymbol{A}\operatorname{diag}(\boldsymbol{\gamma})\boldsymbol{A}^{\mathsf{H}}\right) \\ \text{subject to} \quad & \boldsymbol{\gamma} \ge \mathbf{0}, \end{aligned} \quad (13)$$

where

$$\begin{aligned} \bar{\boldsymbol{\Sigma}}^{(k)} &= \boldsymbol{A}\operatorname{diag}\left(\boldsymbol{\gamma}^{(k)}\right)\boldsymbol{A}^{\mathsf{H}} + \sigma^2\,\boldsymbol{I} \\ \boldsymbol{M}^{(k)} &= \left(\bar{\boldsymbol{\Sigma}}^{(k)}\right)^{-1}\boldsymbol{A}\operatorname{diag}\left(\boldsymbol{\gamma}^{(k)} + \frac{\sigma^2}{\eta}\right). \end{aligned} \quad (14)$$

Using the cyclic property of the trace operator ($\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$), the simplified surrogate problem is

$$\begin{aligned} \underset{\boldsymbol{\gamma}}{\text{minimize}} \quad & \bar{\ell}^{(k+1)}\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(k)}, \sigma^2\right) \\ \text{subject to} \quad & \boldsymbol{\gamma} \ge \mathbf{0}, \end{aligned} \quad (15)$$

where

$$\begin{aligned} \bar{\ell}^{(k+1)}\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(k)}, \sigma^2\right) &= \sum_{i=1}^{N}\frac{T_i^{(k)}}{\gamma_i + \sigma^2/\eta} + \sum_{i=1}^{N}\gamma_i\,U_i^{(k)} \\ T_i^{(k)} &= \left[\left(\boldsymbol{M}^{(k)}\right)^{\mathsf{H}}\boldsymbol{S}\boldsymbol{M}^{(k)}\right]_{ii} = \left(\boldsymbol{m}_i^{(k)}\right)^{\mathsf{H}}\boldsymbol{S}\boldsymbol{m}_i^{(k)}, \\ & \quad \text{for} \quad i = 1, \ldots, N \\ U_i^{(k)} &= \left[\boldsymbol{A}^{\mathsf{H}}\left(\bar{\boldsymbol{\Sigma}}^{(k)}\right)^{-1}\boldsymbol{A}\right]_{ii} = \mathbf{a}_i^{\mathsf{H}}\left(\bar{\boldsymbol{\Sigma}}^{(k)}\right)^{-1}\mathbf{a}_i, \\ & \quad \text{for} \quad i = 1, \ldots, N \end{aligned} \quad (16)$$

with $\boldsymbol{m}_i^{(k)}$ denoting the i-th column of $\boldsymbol{M}^{(k)}$. Taking the derivative of $\bar{\ell}^{(k+1)}\left(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}^{(k)}, \sigma^2\right)$ in (15) w.r.t γ_i, we obtain

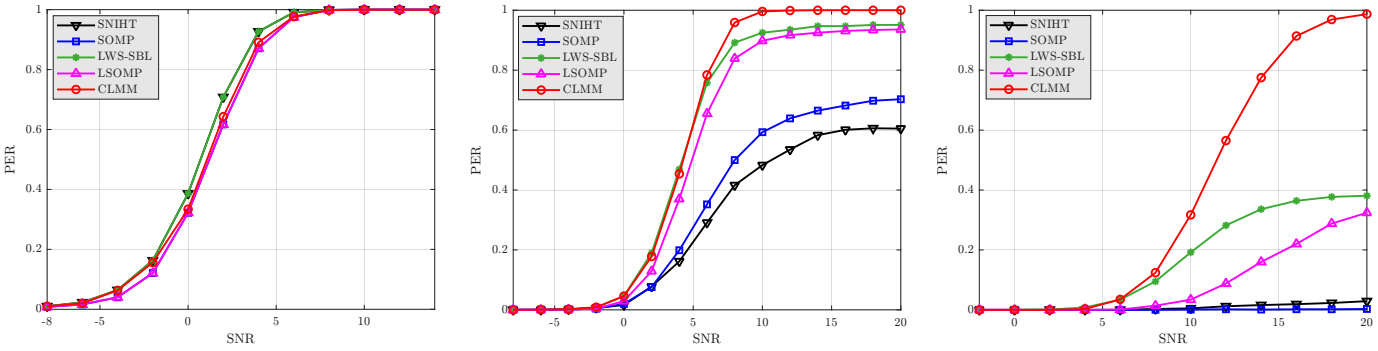
$$\begin{aligned} \frac{\partial\bar{\ell}^{(k+1)}}{\partial\gamma_i} &= -\frac{T_i^{(k)}}{(\gamma_i + \sigma^2/\eta)^2} + U_i^{(k)}, \\ & \quad \text{for} \quad i = 1, \ldots, N. \end{aligned} \quad (17)$$

Equating (17) along with taking into consideration the constraint in (15), we find $\gamma_i^{(k+1)}$ as

$$\gamma_i^{(k+1)} = \left[\sqrt{\frac{T_i^{(k)}}{U_i^{(k)}}} - \frac{\sigma^2}{\eta}\right]_+, \qquad i = 1, \ldots, N, \quad (18)$$

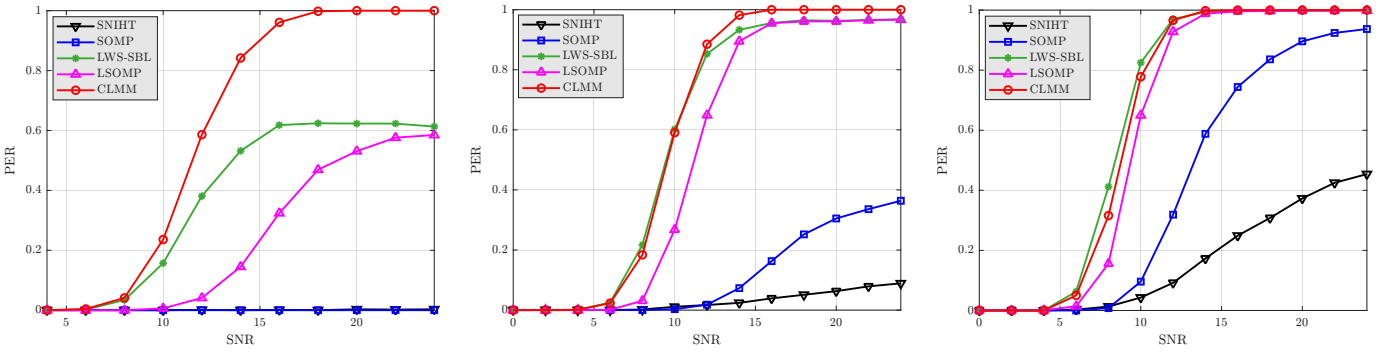
where $[\cdot]_+ = \max(\cdot, 0)$. The steps of the proposed CLMM algorithm are outlined in Algorithm 1. Note that the choice of η must satisfy the inequality $\eta > \lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}})$ based on Corollary 2. The choice of $\eta = \operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right)$ in Step 2 of Algorithm 1 satisfies this inequality as $\operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathsf{H}}\right) > \lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{H}})$.

The computational cost of Step 2 of CLMM (Algorithm 1) is $\mathcal{O}(MN)$. Computational complexity of Steps 3-6 are $\mathcal{O}(M^2N)$, $\mathcal{O}(M^3 + M^2N)$, $\mathcal{O}(M^2N)$, and $\mathcal{O}(N)$. Let I be the number of iterations in CLMM (bounded by $I_{\max}$), the total complexity of the proposed CLMM algorithm is $I \cdot \mathcal{O}(M^3 + M^2N)$. As $N \ge M$ in all applications that we consider, the complexity of the proposed CLMM algorithm is $I \cdot \mathcal{O}(M^2N)$.



(a) $K = 1$. (b) $K = 2$. (c) $K = 4$.

Fig. 1: PER vs. SNR when $M = 6$, $N = 512$, and $L = 20$. The sparsity ratio K/M is $1/6$, $2/6$, and $4/6$ in subplots (a), (b), and (c) , respectively.



(a) $M = 16$. (b) $M = 20$. (c) $M = 24$.

Fig. 2: PER vs SNR when $K = 12$, $N = 512$, $L = 16$, and $M \in \{16, 20, 24\}$.

III. Simulation Results

This section compares the sparse signal recovery performance of the proposed CLMM algorithm to those of the SNIHT [14, Algorithm 1], SOMP [15, Algorithm 3.1], LWS-SBL [18, Algorithm 1], and LSOMP [16]. For the proposed CLMM algorithm, we set $\delta = 10^{-3}$ and $I_{\max} = 100$. Additionally, the indices of K largest elements of $\boldsymbol{\gamma}$ are taken as the signal support set $\mathcal{S}$. LWS-SBL and CLMM assume σ^2 is known. The dictionary $\mathbf{A} \in \mathbb{C}^{M\times N}$ is generated with i.i.d. entries from the set $\{\pm 1 \pm 1j\}/\sqrt{2}$ where the real and imaginary parts are independent Bernoulli random variables with equal probability. As is standard in compressed sensing, its columns are normalized to unit norm, which here corresponds to scaling by $1/\sqrt{M}$.

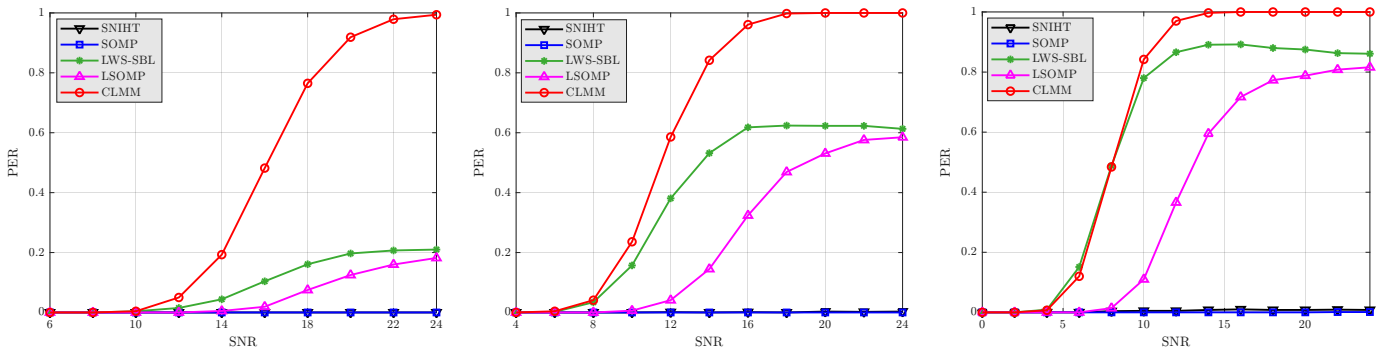
As performance measure we use the empirical *probability of exact recovery*,

$$\text{PER} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{I}\left(\hat{\mathcal{S}}^{(t)} = \mathcal{S}^{(t)}\right),$$

where $\mathrm{I}(\cdot)$ denotes the indicator function, and $\hat{\mathcal{S}}^{(t)}$ denotes the estimate of the true signal support $\mathcal{S}^{(t)}$ for t-th Monte Carlo (MC) trial. The number of MC trials is $T = 1000$ for all simulation examples. Unless otherwise stated, we assume that there are four source signal clusters. Let $\mathcal{S} = \{i_1, \ldots, i_K\}$ denote the true support set where K is chosen as $K = 4K_0$ with K_0 being the number of signals in each cluster. We define the power of the first source signal in the first cluster as $\sigma_1^2 = \gamma_{i_1}$. The powers of the second, third, and fourth signal clusters are set to be 1 dB, 2 dB, and 4 dB lower than that of the first cluster, respectively. In the figures, the signal-to-noise ratio (SNR) is defined as the SNR of the first signal cluster, i.e., $\text{SNR (dB)} = 10\log_{10}(\sigma_1^2/\sigma^2)$.

Fig. 1 shows the PER for $K = 1$ (Fig. 1a), $K = 2$ (Fig. 1b), and $K = 4$ (Fig. 1c) when $M = 6$, $N = 512$, and $L = 20$. In Fig. 1a (sparsity ratio $K/M = 1/6$), it is observed that all the methods tested perform similarly. In Fig. 1b (sparsity ratio $K/M = 2/6$), the SNIHT is the worst performing method while the proposed CLMM method is the best performing. The accumulation of errors and performance saturation, which are well-known behaviors of greedy-based approaches when the sparsity ratio is large, are already visible here for SNIHT, SOMP, LWS-SBL, and LSOMP. In other words, increasing SNR beyond a threshold point does not improve their performance and they may fail to achieve perfect signal recovery. In Fig. 1c (sparsity ratio $K/M = 4/6$), the proposed CLMM method substantially outperforms other methods tested at mid and high SNR regimes.

Fig. 2 shows the PER for $M = 16$ (Fig. 2a), $M = 20$



(a) $L = 8$. (b) $L = 16$. (c) $L = 32$.

Fig. 3: PER vs. SNR when $K = 12$, $N = 512$, $M = 16$, and $L \in \{8, 16, 32\}$.

(Fig. 2b), and $M = 24$ (Fig. 2c) when $K = 12$, $N = 512$, and $L = 16$. In Fig. 2a, the SNIHT and SOMP methods completely collapse. The proposed CLMM method is the best performing method here. In Fig. 2b, the SNIHT method is the worst performing method while the LWS-SBL and CLMM method have equivalent performance at low SNR but CLMM outperforms LWS-SBL in higher SNR. In Fig. 2c, the LWS-SBL, LSOMP and CLMM are the best performing while the SNIHT method is the worst.

Fig. 3 shows the PER for $L = 8$ (Fig. 3a), $L = 16$ (Fig. 3b), and $L = 32$ (Fig. 3c) when $K = 12$, $N = 512$, and $M = 16$. In Figs. 3a and 3b, the proposed CLMM method significantly outperforms other methods tested. In Fig. 3c, the proposed CLMM is the best performing method.

IV. Conclusion

In this paper, we have developed a covariance learning-based method for SSR within the MM framework. The proposed algorithm, called CLMM, is built by first constructing a surrogate optimization problem through suitable majorizers for both the convex and concave components of the original covariance learning objective function. Using this surrogate, CLMM employs an efficient iterative procedure to estimate the sparse signal powers and the associated support set, without requiring explicit estimation of the source signal matrix. Extensive simulation results show that CLMM achieves superior recovery performance compared to existing approaches, with gains being particularly visible as the sparsity ratio K/M increases and approaches 1. These findings highlight the potential of MM-based covariance learning as a powerful and flexible tool for SSR in a broad range of signal processing applications.

References

[1] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[2] Z. Qin, J. Fan, Y. Liu, Y. Gao, and G. Y. Li, "Sparse representation for wireless communications: A compressive sensing approach," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 40–58, 2018.

[3] M. Esfandiari, P. Pulkkinen, S. A. Vorobyov, and V. Koivunen, "AdaBoost-based channel estimation in one-bit millimeter-Wave MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Hyderabad, 2025, pp. 1–5.

[4] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Med. imag.*, vol. 28, no. 1, pp. 106–121, 2008.

[5] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[6] M. Esfandiari, P. Pulkkinen, S. A. Vorobyov, and V. Koivunen, "CS-AdaBoost: One-bit compressed sensing for direction-of-arrival estimation," in *33rd Europ. Signal Process. Conf. (EUSIPCO)*, Palermo, 2025, pp. 2097–2101.

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

[8] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," vol. 1, no. Jun, pp. 211–244, 2001.

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

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

[11] M. Al-Shoukairi, P. Schniter, and B. D. Rao, "A GAMP-based low complexity sparse Bayesian learning algorithm," *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 294–308, 2017.

[12] A. Hashemi, C. Cai, Y. Gao, S. Ghosh, K.-R. Müller, S. S. Nagarajan, and S. Haufe, "Joint learning of full-structure noise in hierarchical bayesian regression models," *IEEE Trans. Med. Imag*, vol. 43, no. 2, pp. 610–624, 2022.

[13] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2016.

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

[15] 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.

[16] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer Science & Business Media, 2010.

[17] E. Ollila, "Matching pursuit covariance learning," in *2024 32nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2024, pp. 2447–2451.

[18] R. R. Pote and B. D. Rao, "Theory and practice of light-weight sequential SBL algorithm: An alternative to OMP," *IEEE Trans. Signal Process.*, 2025.

[19] Y. Sun, P. Babu, and D. P. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3576–3590, 2016.