BLOCK MAJORIZATION-MINIMIZATION ALGORITHMS FOR LOW-RANK CLUTTER SUBSPACE ESTIMATION

A. Breloy[1], *Y. Sun*[2], *P. Babu*[3], *D.P. Palomar*[2]

[1] LEME, Universite Paris Ouest, 50, rue de Sèvres, 92410 Ville d'Avray, France

[2] Department of Electronic and Computer Engineering, HKUST, Clear Water Bay, Kowloon, Hong Kong

[3] CARE, Indian Institute of Technology Delhi, New Delhi-110016, India

ABSTRACT

This paper addresses the problem of the clutter subspace projector estimation in the context of a disturbance composed of a low rank heterogeneous (Compound Gaussian) clutter and white Gaussian noise. We derive two algorithms based on the block majorization-minimization framework to reach the maximum likelihood estimator of the considered model. These algorithms are shown to be computationally faster than the state of the art, with guaranteed convergence. Finally, the performance of the related estimators is illustrated in terms of estimation accuracy and computation speed.

***Index Terms*—** Subspace estimation, Maximum Likelihood Estimator, Low Rank, Compound Gaussian, majorization-minimization.

1. INTRODUCTION

In array processing, many applications require the use of the covariance matrix (CM) of the noise [1–3]. In practice, the CM is unknown and has to be estimated from a set of M dimensional samples $\{\mathbf{z}_k\}$, which are K signal-free independent realizations of the noise. The CM estimate is then used to perform the so-called adaptive process. In radar systems, the noise is composed of a correlated noise, referred to as *clutter* (caused by the response of the environment to the emitted signal), and White Gaussian Noise (WGN, the thermal noise due to electronics). The total covariance of this disturbance is therefore

$$\mathbf{\Sigma}_{tot} = \mathbf{\Sigma} + \sigma^2 \mathbf{I}, \quad (1)$$

where $\mathbf{\Sigma}$ is the clutter CM and $\sigma^2\mathbf{I}$ is the CM of the WGN. In most cases, the clutter belongs to a subspace of limited dimension, meaning that the clutter CM $\mathbf{\Sigma}$ has rank $R < M$. Furthermore, when the clutter corresponds to a strong interference contained in a low dimensional subspace ($R \ll M$), one can use the following low rank (LR) approximation [4]:

$$\mathbf{\Sigma}_{tot}^{-1} \simeq \frac{1}{\sigma^2}\mathbf{\Pi}_c^{\perp} \simeq \frac{1}{\sigma^2}\left(\mathbf{I} - \mathbf{\Pi}_c\right),$$

where $\mathbf{\Pi}_c$, named clutter subspace projector (CSP), is the projector onto the subspace spanned by the R eigenvectors associated with the R largest eigenvalues of the matrix $\mathbf{\Sigma}$. This approximation allows developing an adaptive process that relies on a CSP estimate $\hat{\mathbf{\Pi}}_c$ rather than a total CM estimate $\hat{\mathbf{\Sigma}}_{tot}$. The practical use of the LR approximation is that an adaptive LR process requires less samples to reach the equivalent performance to the classical ones [5], which is valuable since the number of samples is often limited. Note that in this paper, we consider that the clutter rank R is known or fixed from a prior estimation step [6]. Moreover, in some applications, the clutter rank can be directly derived from the geometry of the system [7]).

Classically, the clutter is assumed to be Gaussian and the CSP is derived from the SVD of the Sample Covariance Matrix (SCM). However, it is now well-known that most modern radar clutter measurements are not Gaussian and behave heterogeneously. Therefore, the SCM may not provide the optimal solution since it is not an accurate estimator of the CM for heavy-tailed distributions. To account for the heterogeneity of the clutter, one can model it with a Compound Gaussian (CG) distribution [8]. The CG family covers a large panel of well-known and useful distributions, notably Weibull, K-distribution, t-distribution, etc. Moreover, this modeling presents good agreement with several real data sets [9, 10]. Eventually, the total disturbance will be modeled in this paper as an LR Compound Gaussian clutter plus a WGN of known variance (as done in [4, 11, 12]). For this context, we address the problem of obtaining the Maximum Likelihood Estimator (MLE) of the CSP [11, 13–15].

In this paper, we propose to apply the block majorization-minimizaton (MM) algorithm framework to the problem of computing the considered MLE. We derive two new algorithms, which, compared to the state of the art on this problem, enjoy the following properties:

- They do not rely on approximations on the CM eigenvalues [11, 13], nor high clutter to noise ratio assumption [15].
- They are computationally faster than the ones derived in [14, 15] (that rely on the gradient descent developed in [16]). Hence, they are more suitable for implementation.

The following conventions are adopted: italic indicates a scalar quantity, lower case boldface indicates a vector quantity, and upper case boldface a matrix. ^{H} denotes the transpose conjugate operator or the simple conjugate operator for a scalar quantity. ^{T} denotes the transpose operator. $\mathcal{CN}(\mathbf{a}, \mathbf{\Sigma})$

is a complex-valued Gaussian distribution of mean $\mathbf{a}$ and covariance matrix $\mathbf{\Sigma}$. $\mathbf{I}$ is the identity matrix of appropriate dimension. $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$ stand for the determinant and the trace of a matrix, respectively. $\hat{d}$ is an estimate of the parameter d. $\{w_n\}$ denotes the set of n elements w_n with $n \in [\![1, N]\!]$.

2. BACKGROUND

2.1. Problem Formulation

We assume that K samples $\{\mathbf{z}_k\}$ are available. Each of the $\mathbf{z}_k \in \mathbb{C}^M$ is a realization of a proper circular LR CG $\mathbf{c}_k$ plus an independent additive zero-mean complex WGN $\mathbf{n}_k$, i.e. $\mathbf{z}_k = \mathbf{n}_k + \mathbf{c}_k$. The WGN $\mathbf{n}_k$ follows the distribution $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, where the variance σ^2 is assumed to be known and fixed to be $\sigma^2 = 1$ without loss of generality. The LR CG [8] clutter $\mathbf{c}_k$ is, conditionally to an unknown deterministic power factor τ_k (the *texture*), distributed as $(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\mathbf{\Sigma})$, where the CM $\mathbf{\Sigma}$ has a rank $R < M$ assumed to be known. The likelihood function is then

$$f(\{\mathbf{z}_k\}|\mathbf{\Sigma}, \{\tau_k\}) = \prod_{k=1}^{K} \frac{e^{-\mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k}}{\pi^M \det(\mathbf{\Sigma}_k)}, \tag{2}$$

with $\mathbf{\Sigma}_k = \tau_k \mathbf{\Sigma} + \mathbf{I}$. The MLE of the clutter CM $\mathbf{\Sigma}$ is therefore defined as the minimizer of the following problem (equivalent to the maximizer of the log-likelihood function):

$$\begin{aligned} \underset{\mathbf{\Sigma}_k, \{\tau_k\}, \mathbf{\Sigma} \succeq \mathbf{0}}{\text{minimize}} \quad & \sum_{k=1}^{K} \log\det(\mathbf{\Sigma}_k) + \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k \\ \text{subject to} \quad & \mathbf{\Sigma}_k = \tau_k \mathbf{\Sigma} + \mathbf{I} \\ & \tau_k \geq 0 \\ & \mathrm{rank}(\mathbf{\Sigma}) \leq R, \end{aligned} \tag{A}$$

where we denote the objective function by $L(\mathbf{\Sigma}, \{\tau_k\})$.

2.2. Block MM principle

To solve Problem (A), we adopt the block majorization-minimization (MM) algorithm framework, which is briefly stated below. Consider the following optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{3}$$

where the optimization variable $\mathbf{x}$ can be partitioned into m blocks as $\mathbf{x} = \left(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\right)$, with each n_i-dimensional block $\mathbf{x}^{(i)} \in \mathcal{X}_i$ and $\mathcal{X} = \prod_{i=1}^{m} \mathcal{X}_i$. At the $(t+1)$-th iteration, the i-th block $\mathbf{x}^{(i)}$ is updated by solving the following problem:

$$\begin{aligned} \underset{\mathbf{x}^{(i)}}{\text{minimize}} \quad & g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right) \\ \text{subject to} \quad & \mathbf{x}^{(i)} \in \mathcal{X}_i, \end{aligned} \tag{4}$$

with $i = (t \bmod m) + 1$ (so blocks are updated in cyclic order) and the continuous surrogate function $g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right)$ satisfying the following properties:

$$\begin{aligned} f(\mathbf{x}_t) &= g_i\left(\mathbf{x}_t^{(i)}|\mathbf{x}_t\right), \\ f\left(\mathbf{x}_t^{(1)}, \ldots, \mathbf{x}^{(i)}, \ldots, \mathbf{x}_t^{(m)}\right) &\leq g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right) \ \forall \mathbf{x}^{(i)} \in \mathcal{X}_i, \\ f'\left(\mathbf{x}_t; \mathbf{d}_i^0\right) &= g_i'\left(\mathbf{x}_t^{(i)}; \mathbf{d}_i|\mathbf{x}_t\right) \\ & \quad \forall \mathbf{x}_t^{(i)} + \mathbf{d}_i \in \mathcal{X}_i, \\ & \quad \mathbf{d}_i^0 \triangleq (\mathbf{0}; \ldots; \mathbf{d}_i; \ldots; \mathbf{0}), \end{aligned}$$

where $f'(\mathbf{x}; \mathbf{d})$ stands for the directional derivative at $\mathbf{x}$ along $\mathbf{d}$. In short, at each iteration, the block MM algorithm updates the variables in one block by minimizing a tight upperbound of the function while keeping the other blocks fixed. In the following sections, we present 2 MM algorithms (based on two diferent parametrization of the variables) to solve problem (A). Due to lack of space, the proofs of the propositions are left to the full publication of our results [17].

3. DIRECT BLOCK MAJORIZATION-MINIMIZATION (DBMM)

As $\mathrm{rank}(\mathbf{\Sigma}) \leq R$, the variable $\mathbf{\Sigma}$ can be reparameterized as $\mathbf{\Sigma} = \mathbf{W}\mathbf{W}^H$ with $\mathbf{W} \in \mathbb{C}^{M \times R}$. Problem (A) can then be written equivalently as

$$\begin{aligned} \underset{\mathbf{\Sigma}_k, \{\tau_k\}, \mathbf{W}}{\text{minimize}} \quad & \sum_{k=1}^{K} \log\det(\mathbf{\Sigma}_k) + \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{z}_k \\ \text{subject to} \quad & \mathbf{\Sigma}_k = \tau_k \mathbf{W}\mathbf{W}^H + \mathbf{I} \\ & \tau_k \geq 0. \end{aligned} \tag{B}$$

Following the block MM methodology, we partition the variables as $\{\{\tau_k\}, \mathbf{W}\}$ and derive an algorithm that updates the blocks in cyclic order. This algorithm is referred to as DBMM. Given a starting point $\left\{\{\tau_k\}^{t=0}, \mathbf{W}^{t=0}\right\}$, one iteratively:

- **updates $\{\tau_k\}^{t+1}$ for fixed $\mathbf{W} = \mathbf{W}^t$**

The objective function is separable in the τ_k's, and for each of them, the following problem should be solved:

$$\begin{aligned} \underset{\tau_k}{\text{minimize}} \quad & \log\det(\tau_k \mathbf{\Sigma} + \mathbf{I}) + \mathbf{z}_k^H (\tau_k \mathbf{\Sigma} + \mathbf{I})^{-1} \mathbf{z}_k \\ \text{subject to} \quad & \tau_k \geq 0. \end{aligned} \tag{B1}$$

Let the SVD of $\mathbf{\Sigma}$ be $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, with $\mathbf{\Lambda} = \mathrm{diag}(\{\lambda_m\})$ (with $\lambda_m = 0$ for $m > R$) and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$. The objective function of (B1) can be simplified to

$$\begin{aligned} L(\tau_k|\mathbf{W}) &= \sum_{m=1}^{M} \log(\tau_k\lambda_m + 1) + \sum_{m=1}^{M} s_{km}(\tau_k\lambda_m + 1)^{-1} \\ &= \sum_{m=1}^{R} \log(\tau_k\lambda_m + 1) + \sum_{m=1}^{R} s_{km}(\tau_k\lambda_m + 1)^{-1} + \text{const.}, \end{aligned} \tag{5}$$

where $s_{km} = \left\|\mathbf{z}_k^H \mathbf{u}_m\right\|^2$. This function has no closed-form minimizer. Therefore, the update of $\{\tau_k\}$ is obtained by minimizing surrogates functions, based on the following proposition:

Proposition 1 *The function $L\left(\tau_k|\mathbf{W}\right)$ in* (5) *can be upperbounded by the surrogate function $L\left(\tau_k|\tau_k^t,\mathbf{W}\right)$ defined as*

$$L\left(\tau_k|\tau_k^t,\mathbf{W}\right) = -\beta_k\log\tau_k + \text{const}+$$
$$\left(\sum_{m=1}^{R}\alpha_{km}\right)\log\left(\frac{\left(\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\alpha_{km}}\tau_k + \frac{\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}}{\sum_{m=1}^{R}\alpha_{km}}\right),$$

where

$$\alpha_{km} = s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1}+1 \quad \textit{and} \quad \beta_k = \sum_{m=1}^{R}s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1}.$$

The equality is achieved at $\tau_k=\tau_k^t$. The surrogate function $L\left(\tau_k|\tau_k^t,\mathbf{W}\right)$ is quasi-convex and has a unique minimizer given by

$$\tau_k^{t+1} = \frac{1}{R}\cdot\frac{\left(\sum_{m=1}^{R}s_{km}\frac{\tau_k^t\lambda_m}{\tau_k^t\lambda_m+1}\right)\cdot\left(\sum_{m=1}^{R}\frac{\alpha_{km}}{1+\lambda_m\tau_k^t}\right)}{\sum_{m=1}^{R}\frac{\alpha_{km}\lambda_m}{1+\lambda_m\tau_k^t}}. \quad (6)$$

• **updates $\mathbf{W}^{t+1}$ for fixed $\{\tau_k\}=\{\tau_k\}^{t+1}$**

To obtain the update of $\mathbf{W}$, one needs to solve

$$\begin{aligned}\underset{\boldsymbol{\Sigma}_k,\mathbf{W}}{\text{minimize}} \quad & \sum_{k=1}^{K}\log\det\left(\boldsymbol{\Sigma}_k\right) + \sum_{k=1}^{K}\mathbf{z}_k^H\boldsymbol{\Sigma}_k^{-1}\mathbf{z}_k \\ \text{subject to} \quad & \boldsymbol{\Sigma}_k = \tau_k\mathbf{W}\mathbf{W}^H+\mathbf{I}.\end{aligned} \quad \text{(B2)}$$

Denote $L\left(\mathbf{W}|\tau_k\right)$ the objective in (B2). Without loss of generality assume that $\tau_k>0$ (otherwise corresponding constant terms can be deleted in the objective). Problem (B2) has no closed-form minimizer. As in the previous part, the update of $\mathbf{W}$ is obtained by minimizing a surrogate function, according to the following proposition:

Proposition 2 *The objective function in* (B2) *$L\left(\mathbf{W}|\tau_k\right)$ can be upperbounded by the convex quadratic function*

$$L\left(\mathbf{W}|\tau_k,\mathbf{W}^t\right) = \text{Tr}\left(\mathbf{W}\mathbf{H}\mathbf{W}^H\right) - \text{Tr}\left(\mathbf{L}\mathbf{W}^H\right) - \text{Tr}\left(\mathbf{L}^H\mathbf{W}\right), \quad (7)$$

with equality achieved at $\mathbf{W}=\mathbf{W}^t$, where the matrices $\mathbf{H}$ and $\mathbf{L}$ are defined according to:

$$\mathbf{H}_k^t = \left(\tau_k^{-1}\mathbf{I}+\tilde{\boldsymbol{\Sigma}}^t\right)^{-1}\left(\mathbf{W}^t\right)^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}^t\left(\tau_k^{-1}\mathbf{I}+\tilde{\boldsymbol{\Sigma}}^t\right)^{-1},$$
$$\mathbf{L}_k^t = \mathbf{z}_k\mathbf{z}_k^H\mathbf{W}^t\left(\tau_k^{-1}\mathbf{I}+\tilde{\boldsymbol{\Sigma}}^t\right)^{-1},$$
$$\mathbf{H} = \textstyle\sum_{k=1}^{K}\left(\left(\tilde{\boldsymbol{\Sigma}}^t+\tau_k^{-1}\mathbf{I}\right)^{-1}+\mathbf{H}_k^t\right),$$
$$\mathbf{L} = \textstyle\sum_{k=1}^{K}\mathbf{L}_k^t,$$

The matrix $\mathbf{H}$ is positive definite by definition, hence the upperbound $L\left(\mathbf{W}|\tau_k,\mathbf{W}^t\right)$ has a unique minimizer given by

$$\mathbf{W}^{t+1} = \mathbf{L}\mathbf{H}^{-1}. \quad (8)$$

4. EIGENSPACE BLOCK MAJORIZATION-MINIMIZATION ALGORITHM (EBMM)

As $\text{rank}\left(\boldsymbol{\Sigma}\right)\le R$, the variable $\boldsymbol{\Sigma}$ can be reparameterized by its eigendecomposition $\boldsymbol{\Sigma}=\sum_{r=1}^{R}c_r\mathbf{v}_r\mathbf{v}_r^H$. Problem (A) can then be rewritten as

$$\begin{aligned}\underset{\{\tau_k\},\{c_r\},\{\mathbf{v}_r\}}{\text{minimize}} \quad & -\sum_{k=1}^{K}\sum_{r=1}^{R}\frac{\tau_k c_r}{\tau_k c_r+1}\mathbf{z}_k^H\mathbf{v}_r\mathbf{v}_r^H\mathbf{z}_k \\ & +\sum_{k=1}^{K}\sum_{r=1}^{R}\log\left(1+\tau_k c_r\right) \\ \text{subject to} \quad & \tau_k\ge 0,\ c_r\ge 0,\ \text{orthonormal } \mathbf{v}_r\text{'s}.\end{aligned} \quad \text{(C)}$$

Following the same methodology as in the previous section, we partition the variables as $\left\{\{\tau_k\},\{c_r\},\{\mathbf{v}_r\}\right\}$ and derive an algorithm that updates the blocks in cyclic order by minimizing an upperbound of the objective. This algorithm is referred to as EBMM. Given a starting point $\left\{\{\tau_k\}^{t=0},\{c_r\}^{t=0},\{\mathbf{v}_r\}^{t=0}\right\}$, one iteratively:

• **updates $\{\tau_k\}^{t+1}$ for fixed $\{c_r\}=\{c_r\}^t$, $\{\mathbf{v}_r\}=\{\mathbf{v}_r\}^t$**

The objective function is separable in the τ_k's, and for each of them the following problem should be solved:

$$\begin{aligned}\underset{\tau_k}{\text{minimize}} \quad & \sum_{r=1}^{R}\log\left(1+\tau_k c_r\right)+\sum_{r=1}^{R}\frac{s_{kr}}{1+\tau_k c_r} \\ \text{subject to} \quad & \tau_k\ge 0,\end{aligned} \quad \text{(C1)}$$

with $s_{kr}=\left\|\mathbf{z}_k^H\mathbf{v}_r\right\|^2$. Notice that the objective function of (C1) has the same form as (5), therefore we can apply Proposition 1 to obtain the τ_k's updates as

$$\tau_k^{t+1} = \frac{1}{R}\cdot\frac{\left(\sum_{r=1}^{R}s_{kr}\frac{\tau_k^t c_r}{\tau_k^t c_r+1}\right)\cdot\left(\sum_{r=1}^{R}\frac{\alpha_{kr}}{1+c_r\tau_k^t}\right)}{\sum_{r=1}^{R}\frac{\alpha_{kr}c_r}{1+c_r\tau_k^t}}, \quad (9)$$

where $\alpha_{kr}=s_{kr}\frac{\tau_k^t c_r}{\tau_k^t c_r+1}+1$.

• **updates $\{c_r\}^{t+1}$ for fixed $\{\tau_k\}=\{\tau_k\}^{t+1}$, $\{\mathbf{v}_r\}=\{\mathbf{v}_r\}^t$**

The objective function is separable in the c_r's, and for each of them the following problem should be solved:

$$\begin{aligned}\underset{c_r}{\text{minimize}} \quad & \sum_{k=1}^{K}\log\left(1+\tau_k c_r\right)+\sum_{k=1}^{K}\frac{s_{kr}}{1+\tau_k c_r} \\ \text{subject to} \quad & c_r\ge 0,\end{aligned} \quad \text{(C2)}$$

with $s_{kr}=\left\|\mathbf{z}_k^H\mathbf{v}_r\right\|^2$. Notice that the c_r's in (C2) play a similar role as the τ_k's in (C1). Similar to the update of τ_k, we can apply Proposition 1 to obtain the c_r's updates as

$$c_r^{t+1} = \frac{1}{K}\cdot\frac{\left(\sum_{k=1}^{K}s_{kr}\frac{\tau_k c_r^t}{\tau_k c_r^t+1}\right)\cdot\left(\sum_{k=1}^{K}\frac{\alpha_{kr}}{1+c_r^t\tau_k}\right)}{\sum_{k=1}^{K}\frac{\alpha_{kr}\tau_k}{1+c_r^t\tau_k}}, \quad (10)$$

where $\alpha_{kr}=s_{kr}\frac{\tau_k c_r^t}{\tau_k c_r^t+1}+1$.

• **updates $\{\mathbf{v}_r\}^{t+1}$ for fixed $\{\tau_k\}=\{\tau_k\}^{t+1}$, $\{c_r\}=\{c_r\}^{t+1}$**

Minimizing the objective $w.r.t.$ $\{\mathbf{v}_r\}$ is equivalent to solving the problem

$$\begin{aligned}\underset{\{\mathbf{v}_r\}}{\text{maximize}} \quad & \sum_{r=1}^{R}\mathbf{v}_r^H\mathbf{M}_r\mathbf{v}_r \\ \text{subject to} \quad & \text{orthonormal } \mathbf{v}_r\text{'s},\end{aligned} \quad \text{(C3)}$$

where $\mathbf{M}_r=\sum_{k=1}^{K}(\tau_k c_r)/(\tau_k c_r+1)\mathbf{z}_k\mathbf{z}_k^H$. The update of $\{\mathbf{v}_r\}$ is obtained thanks to the following proposition:

Proposition 3 *The objective function in* (C3) *can be lower-bounded by the surrogate function*

$$L\left(\{\mathbf{v}_r\} \,|\, \{\tau_k\}, \{c_r\}, \{\mathbf{v}_r\}^t\right) = \sum_{r=1}^{R} \left[\left(\mathbf{v}_r^t\right)^H \mathbf{M}_r \mathbf{v}_r + \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r^t\right] + \text{const.} \quad (11)$$

with equality achieved at $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$. *Maximizing this surrogate, under orthonormality constraints on the* $\{\mathbf{v}_r\}$, *is equivalent to solving*

$$\begin{array}{ll} \underset{\mathbf{V}}{\text{minimize}} & \|\mathbf{A} - \mathbf{V}\|_F^2 \\ \text{subject to} & \mathbf{V}^H\mathbf{V} = \mathbf{I}. \end{array} \quad (12)$$

where

$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_R] \quad \textit{and} \quad \mathbf{A} = [\mathbf{M}_1\mathbf{v}_1^t; \ldots; \mathbf{M}_R\mathbf{v}_R^t]$$

Let the thin SVD of $\mathbf{A}$ *be* $\mathbf{A} = \mathbf{V}_A\mathbf{D}_A\mathbf{U}_A^H$. *The update of* $\{\mathbf{v}_r\}$ *is the solution of* (12), *given by [16] as:*

$$\mathbf{V}^{t+1} = \mathbf{V}_A\mathbf{U}_A^H. \quad (13)$$

Optional loop: While the block MM framework indicates a cyclic update of the variables $\{\tau_k\}, \{c_r\}, \{\mathbf{v}_r\}$, our numerical tests reveal that including an inner loop on this last step provides a faster decreasing rate of the objective value than updating the $\mathbf{v}_r$'s only once.

4.1. Convergence Analysis

DBMM Algorithm: Any limit point of the pair $\{\{\tau_k^t\}, \mathbf{W}^t\}$ generated by Algorithm DBMM is a stationary point of Problem (B). Indeed, we have proved the quasi-convexity and the uniqueness of the minimizer of the surrogate functions $L\left(\tau_k|\tau_k^t, \mathbf{W}^t\right)$ and $L\left(\mathbf{W}|\tau_k, \mathbf{W}^t\right)$ in the previous subsections. The algorithm convergence is a direct application of Theorem 2 (a) in [18].

EBMM Algorithm: Since the constraint set of Problem (C) is non-convex, the convergence result provided in [18] cannot be applied for EBMM. To the best of our knowledge, there is no convergence result of general block descent type algorithms with a non-convex constraint set. The sequence of objective values generated by Algorithm 2 will converge because of monotonicity, but the convergence of the points $\left(\{\tau_k\}^t, \{c_r\}^t, \{\mathbf{v}_r\}^t\right)$ remains unknown. Nevertheless, section 5 will show that the numerical performance of Algorithm 2 is satisfactory.

5. NUMERICAL RESULTS

Simulation parameters Samples $\mathbf{z}_k$ are generated according to the LR Compound Gaussian plus WGN model described in section 2: $\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k$. The WGN is distributed as $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\sigma^2 = 1$. The LR Compound Gaussian clutter is distributed as $(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\mathbf{\Sigma})$, with a random texture τ_k, $i.i.d.$ generated for each sample. The texture PDF is a Gamma distribution (leading to a K-distributed clutter) of shape parameter ν and scale parameter $1/\nu$, denoted $\tau \sim \Gamma(\nu, 1/\nu)$, which satisfies $\mathbb{E}(\tau) = 1$. The rank R clutter CM $\mathbf{\Sigma}_c$ is constructed with the largest R eigenvalues and the corresponding eigenvectors of a Toeplitz matrix of correlation parameter $\rho \in [0, 1]$. This matrix is then scaled to set the clutter to noise ratio, defined as CNR $= \mathbb{E}(\tau)\text{Tr}(\mathbf{\Sigma})/(R\sigma^2)$.

Studied Estimators $\hat{\mathbf{\Pi}}_{SCM}$ denotes the CSP estimator derived from the EVD of the SCM, i.e. $\hat{\mathbf{\Sigma}}_{SCM} = \sum_{k=1}^{K} \mathbf{z}_k\mathbf{z}_k^H/K$. $\hat{\mathbf{\Pi}}_{SFPE}$ denotes the CSP estimator derived from the EVD of the Shrinkage-FPE (SFPE) [19, 20]. Since there is no rule to adaptively select the optimal shrinkage parameter β for the CSP estimation problem, we test the following values: $\beta_1 = \max(1 - K/M + \epsilon\,,\,0)$, which is the lowest β allowed for the under-sampled cases ($K \leq M$) and coincides with Tyler's estimator (referred to as FPE) for the over-sampled cases ($K > M$); $\beta_2 = (\beta_1 + \beta_3)/2$; and $\beta_3 = 1 - \epsilon$. We set $\epsilon = 10^{-2}$. $\hat{\mathbf{\Pi}}_{DBMM}$ denotes the CSP MLE computed with the DBMM Algorithm in Section 3. $\hat{\mathbf{\Pi}}_{EBMM}$ denotes the clutter subspace projector MLE computed with the EBMM Algorithm in Section 4. $\hat{\mathbf{\Pi}}_{MLE}$ denotes the CSP MLE under high CNR assumption, computed with the algorithm from [15]. This algorithm will be referred to as "Algorithm 3" in the rest of the paper. $\hat{\mathbf{\Pi}}_{A-MLE}$ denotes the clutter subspace projector approached MLE under high CNR assumption, computed with the second algorithm from [15]. This algorithm will be referred to as "Algorithm 4" in the rest of the paper.

Results Fig.1 displays the mean NMSE criterion ($\mathbb{E}(||\hat{\mathbf{\Pi}}_c - \mathbf{\Pi}_c||^2)/R$ for a given estimator $\hat{\mathbf{\Pi}}_c$) versus K of the studied estimators for a given configuration (computed over 1000 Monte-Carlo simulations). It illustrates the performance of the proposed methods: block MM Algorithms 1 and 2 reach the identically lowest NMSE, and algorithms from [15] lead to slightly higher NMSE, yet better than the SFPE with various β, and the SCM. Fig.2 displays a typical realization of the objective value versus the time of computation. One can observe that the MLE and A-MLE algorithms from [15] converge to a sub-optimal point of the problem. This was to be expected since these algorithms are optimizing a modified likelihood (assuming High CNR, the WGN is ignored over the clutter subspace). Contrary to these algorithms, DBMM converges to a critical point with a smaller objective value. We also notice that EBMM converges in practice to the same point as DBMM, *i.e.*, a critical point. Fig.2 also illustrates that despite a fast convergence, Algorithm 3 has a slow computation due to the use of the modified gradient descent [16]. It also shows that EBMM is less computationally intensive than DBMM since it requires less time to converge. This can be explained by the fact that constructing the update of $\mathbf{W}$ in DBMM has a complexity that grows linearly with the sample size. On the contrary, the update of $\{\mathbf{v}_r\}$ in EBMM involves the SVD of a matrix of fixed dimension. To conclude, both

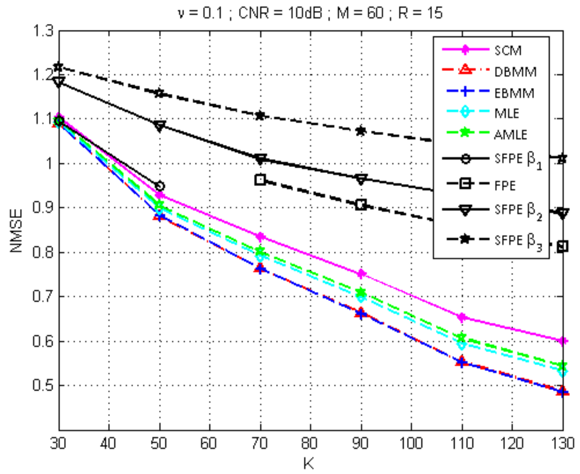

Fig. 1. Mean NMSE of the different estimators versus the number of samples K. $M = 60$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$ dB.

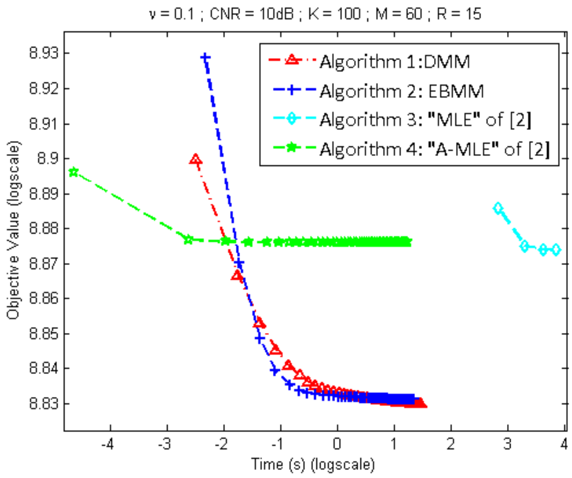

Fig. 2. Objective value versus computation time for different MLE algorithms: Direct Block-MM (red), Eigenspace Block-MM (blue), "MLE" algorithm of [15] (black), "A-MLE" algorithm of [15] (green). $K = 100$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$ dB.

figures illustrate the applicative interest of the two proposed computation methods. Indeed, they reach the best estimation performance for the considered model at a low computational cost. They could therefore be suitable for applications that involve the estimation of signal subspace or Low Rank structured Covariance Matrices.

6. REFERENCES

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 3, pp. 276–280, March 1986.

[2] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 1–16, january 2001.

[3] S. Haykin, *Adaptive filter theory*, ser. Prentice-Hall information and system sciences series. Prentice Hall, 2002.

[4] M. Rangaswamy, F. Lin, and K. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, pp. 1653 – 1665, 2004.

[5] A. Haimovich, "Asymptotic distribution of the conditional signal-to-noise ratio in an eigenanalysis-based adaptive array," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, pp. 988 – 997, 1997.

[6] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, July 2004.

[7] L. E. Brennan and F. Staudaher, "Subclutter visibility demonstration," RL-TR-92-21, Adaptive Sensors Incorporated, Tech. Rep., March 1992.

[8] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[9] M. Greco, F. Gini, and M. Rangaswamy, "Statistical analysis of measured polarimetric clutter data at different range resolutions," *Radar, Sonar and Navigation, IEE Proceedings -*, vol. 153, no. 6, pp. 473–481, December 2006.

[10] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 876–879, Dec 2012.

[11] R. Raghavan, "Statistical interpretation of a data adaptive clutter subspace estimation algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1370 – 1384, 2012.

[12] G. Ginolhac, P. Forster, F. Pascal, and J.-P. Ovarlez, "Performance of two low-rank STAP filters in a heterogeneous noise," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 57–61, 2013.

[13] A. Breloy, L. L. Magoarou, G. Ginolhac, F. Pascal, and P. Forster, "Maximum likelihood estimation of clutter subspace in non homogeneous noise context," in *Proceedings of EUSIPCO*, Marrakech, Morocco, September 2013.

[14] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Clutter subspace estimation in low rank heterogeneous noise context," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2173–2182, May 2015.

[15] ——, "Robust covariance matrix estimation in heterogeneous low-rank context," *IEEE Transactions on Signal Processing, submitted.*

[16] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 635–650, Mar 2002.

[17] Y. Sun, A. Breloy, P. Babu, D. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *Signal Processing, IEEE Transactions on, to appear*.

[18] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[19] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5640–5651, Nov 2014.

[20] Y. Sun, P. Babu, and D. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5143–5156, Oct 2014.