Low-Complexity Algorithms for Low Rank Clutter Parameters Estimation in Radar Systems

Ying Sun, Arnaud Breloy, *Student Member, IEEE*, Prabhu Babu, Daniel P. Palomar, *Fellow, IEEE*, Frédéric Pascal, *Senior Member, IEEE*, and Guillaume Ginolhac, *Member, IEEE*

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. In such a context, adaptive processing based on an estimated orthogonal projector onto the clutter subspace (instead of an estimated covariance matrix) requires less samples than classical methods. The clutter subspace estimate is usually derived from the eigenvalue decomposition of a covariance matrix estimate. However, it has been previously shown that a direct maximum likelihood estimator of the clutter subspace projector can be obtained for the considered context. In this paper, we derive two algorithms based on the block majorization-minimization framework to reach this estimator. 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 on realistic Space Time Adaptive Processing for airborne radar simulations.

Index Terms—Covariance matrix and projector estimation, maximum likelihood estimator, low rank structure, compound Gaussian, majorization-minimization.

I. INTRODUCTION

IN array processing, many applications require the use of the covariance matrix (CM) of the noise; these include source localization techniques [1], [2], radar and sonar detection methods [3], [4], and filters [5]. In practice, the CM is unknown and has to be estimated from a set of samples $\mathbf{z}_k \in \mathbb{C}^M$, $k \in [\![1, K]\!]$, 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

$$\mathbf{\Sigma}_{\text{tot}} = \mathbf{\Sigma} + \sigma^2 \mathbf{I}, \tag{1}$$

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 $R < M$. Consequently, the total CM is structured as a Low Rank (LR) plus a scaled identity matrix.

When the clutter corresponds to a strong interference contained in a low dimensional subspace ($R \ll M$), one can use the following LR approximation [6], [7]:

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

where $\mathbf{\Pi}_c$, named clutter subspace projector, is the orthogonal projector onto the clutter subspace. This subspace is spanned by the R eigenvectors $\mathbf{v}_r$ associated with the R largest eigenvalues of the matrix $\mathbf{\Sigma}$, *i.e.*, $\mathbf{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H$. This approximation allows developing an adaptive process (*e.g.*, filter or detector) that relies on a clutter subspace projector estimate $\hat{\mathbf{\Pi}}_c$ rather than a total CM estimate $\hat{\mathbf{\Sigma}}_{\text{tot}}$. Note that in this paper, we consider that the clutter rank R is known or fixed from a prior estimation step. There are several methods of rank estimation present in the literature (*e.g.*, [8]–[10] and references therein). Moreover, in some applications, the clutter rank can be directly derived from the geometry of the system (such as in STAP [11] thanks to the Brennan rule [12]).

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, which is valuable since the number of available samples is often limited. For example, if we consider the case of a Gaussian distributed clutter, the optimal adaptive filter is built from the Sample Covariance Matrix (SCM), which is the Maximum Likelihood Estimator (MLE) in that scenario. In this case, $K = 2M$ samples are required to ensure a 3 dB loss of the output Signal to Noise Ratio (SNR) compared to the optimal non-adaptive filter [13]. If instead we use the MLE of the clutter subspace projector, which corresponds to the subspace spanned by the R largest eigenvectors of the SCM and can be obtained from its EigenValue Decomposition (EVD), one can build an adaptive LR filter that reaches the equivalent performance to the previous scheme with only $K = 2R \ll 2M$ samples [14].

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 or in the presence of outliers. To

Y. Sun, P. Babu, and D. P. Palomar are with the Hong Kong University of Science and Technology (HKUST), Hong Kong (e-mail: ysunac@ust.hk; eeprabhubabu@ust.hk; palomar@ust.hk).

A. Breloy is with the SATIE, ENS Cachan, CNRS, F-94230 Cachan, and also with SONDRA, Centrale-Supélec, F-91192 Gif-sur-Yvette Cedex, France (e-mail: abreloy@satie.ens-cachan.fr).

F. Pascal is with L2S, Centrale-Supélec, F-91192 Gif-sur-Yvette Cedex, France (e-mail: frederic.pascal@centralesupelec.fr).

G. Ginolhac is with the LISTIC, Université de Savoie Mont-Blanc, 74944 Annecy le Vieux Cedex, France (e-mail: guillaume.ginolhac@univ-savoie.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2512535

account for the heterogeneity of the clutter, one can model it with a Compound Gaussian distribution, which is a sub-family of the Complex Elliptically Symmetric (CES) distribution [15]. The Compound Gaussian family covers a large panel of well-known and useful distributions, notably Gaussian, Weibull, K-distribution, t-distribution, etc. Moreover, it has a well-founded physical interpretation and presents good agreement with several real data sets [16]–[19]. Eventually, the total disturbance will be modeled in this paper as an LR Compound Gaussian clutter plus a WGN (as done in [7], [20]–[22]).

We point out that the sum of a Compound Gaussian and a WGN cannot be represented as a simple Compound Gaussian vector with different distribution parameters (except for the trivial Gaussian case). Nevertheless, most of the related works consider the case of a full rank Compound Gaussian clutter and neglect the possible LR plus WGN structure. Under this framework, robust estimation[1] of the CM can be performed using the so-called M-estimators [23]–[25] that are seen as generalized MLEs for CES distributions. A detailed review of this framework can be found in [15]. These estimators have been studied in detail and successfully applied in modern detection/estimation literature due to their interest both from a theoretical and an application point of view [26]–[32]. To cope with an under-sampled scenario, regularization of these estimators has been studied intensively in the literature, and is still an on-going cutting-edge topic. For example, shrinkage estimators for the CES model have been considered in [33]–[36] that shrink the estimator to the identity matrix. Regularization by imposing constraints on the condition number of the covariance matrix for the compound Gaussian model has been studied in [37]. Another class of works tackle this problem by imposing structural constraints, including group symmetry, Toeplitz, Kronecker structures, on the covariance matrix, see [37]–[41] for example. Instead of the SCM, it is possible to derive clutter subspace projector estimators from the EVD of a (regularized or not) M-estimator. However, despite their robustness properties, M-estimators may not achieve the optimal clutter subspace estimation performance as they do not take into account the noise structure (1) that we consider in this work.

For this noise model, the seminal work [21] derived the MLE of the clutter subspace projector under the assumptions that the CM of the LR Compound Gaussian clutter has identical eigenvalues, and the Probability Density Function (PDF) of the texture is known. The assumption of known texture PDF has been relaxed in [42] by treating the texture as an unknown deterministic parameter. The expression of the MLE of the clutter CM parameters (eigenvectors and eigenvalues) was derived in [43], which does not have a simple closed-form and therefore requires an iterative algorithm to be reached. Initially, [43] focused on the clutter subspace projector MLE and proposed an ad-hoc algorithm to obtain it. This leads to an accurate clutter subspace projector estimator, but a poor recovery of the total CM. In [44], the high Clutter to Noise Ratio (CNR) assumption was made to derive a general "2-Step MLE" algorithm (with several possible adaptations) that was shown to provide accurate estimates of the CM. Most algorithms developed [43], [44] require the use of a gradient descent algorithm on the manifold proposed in [45], which can be difficult to tune and computationally costly.

In this paper, we propose to apply the block majorization-minimizaton algorithm framework to the problem of computing the considered MLE. We derive two new algorithms that enjoy the following properties:

- They do not rely on any heuristic (as in [43]) or high CNR assumption (as in [44]).
- They are computationally faster than the ones derived in [43] and [44], hence more suitable for implementation.
- Compared to the gradient descent algorithm [45], which requires many trials to find the descent step, the proposed algorithms only require basic matrix operations or Singular Value Decomposition (SVD).

The paper is organized as follows. Section II states the problem formulation and briefly reviews the block majorization-minimizaton methodology. Then, Sections III and IV derive the two new algorithms for computing the MLE of the considered problem. The difference between the two algorithms lies in the choice of the variables with respect to which the likelihood function is cyclically optimized. In Section V, we discuss possible methods for of estimating the clutter rank and the white Gaussian noise power, which are assumed to be known in this study. Section VI validates the performance of the proposed algorithms with simulations. Moreover, the proposed algorithms are applied to a realistic Space Time Adaptive Processing (STAP) for airborne radar simulation. Finally, Section VII draws the conclusions of this study.

Throughout the paper 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. $\mathrm{d}(\cdot)$ denotes the differential operator. $\mathcal{CN}(\mathbf{a}, \boldsymbol{\Sigma})$ is a complex-valued Gaussian distribution of mean $\mathbf{a}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{I}$ 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. $\{w_n\}_{n \in [\![1,N]\!]}$ denotes the set of n elements w_n with $n \in [\![1, N]\!]$, and will often be abbreviated to $\{w_n\}$ in the sequel. $\mathbf{e}_i$ is the ith vector of the canonical basis of appropriate dimension.

II. BACKGROUND

A. Problem Formulation

We assume that K samples $\{\mathbf{z}_k\}_{k \in [\![1,K]\!]}$ are available. Each of the data $\mathbf{z}_k \in \mathbb{C}^M$ corresponds to a realization of a proper circular LR Compound Gaussian process $\mathbf{c}_k$ plus an independent additive zero-mean complex WGN $\mathbf{n}_k$:

$$\mathbf{z}_k = \mathbf{n}_k + \mathbf{c}_k. \tag{2}$$

The WGN $\mathbf{n}_k$ follows the distribution

$$\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{3}$$

where the power of the WGN σ^2 is assumed to be known[2] and fixed to be $\sigma^2 = 1$ without loss of generality. The LR Compound Gaussian [15] $\mathbf{c}_k$ is an M-dimensional zero-mean complex Gaussian vector (the *speckle*) with CM $\boldsymbol{\Sigma}$, multiplied by

---

[1]*i.e.*, not highly sensitive to the underlying distribution

[2]This hypothesis is made for describing a valid theoretical framework. In practice, presented results can be applied with an estimate of σ^2 used as its actual value.

the square root of an independent positive random power factor (the *texture*) τ_k. Each $\mathbf{c}_k$ follows then, conditionally to τ_k:

$$(\mathbf{c}_k|\tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k\boldsymbol{\Sigma}), \tag{4}$$

where the Compound Gaussian clutter CM $\boldsymbol{\Sigma}$ has a Low Rank structure with $\mathrm{rank}(\boldsymbol{\Sigma}) = R < M$. The rank R is assumed to be known and a discussion on this assumption will be provided in the remark below. In this work, we do not assume the knowledge of the texture PDF, and treat each τ_k as an unknown deterministic variable. The likelihood function is then

$$f(\{\mathbf{z}_k\}|\boldsymbol{\Sigma}, \{\tau_k\}) = \prod_{k=1}^{K} \frac{e^{-\mathbf{z}_k^H \boldsymbol{\Sigma}_k^{-1} \mathbf{z}_k}}{\pi^M \det(\boldsymbol{\Sigma}_k)}, \tag{5}$$

with $\boldsymbol{\Sigma}_k = \tau_k\boldsymbol{\Sigma} + \mathbf{I}$.

The MLE of the clutter CM $\boldsymbol{\Sigma}$ is therefore defined as the minimizer of the following problem (equivalent to the maximizer of the log-likelihood function):

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

where $\boldsymbol{\Sigma}$ is of size $M \times M$. Denote the objective function by $L(\boldsymbol{\Sigma}, \{\tau_k\})$.

B. 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:

$$\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\
\text{subject to} \quad & \mathbf{x} \in \mathcal{X},
\end{aligned} \tag{6}$$

where the optimization variable $\mathbf{x}$ can be partitioned into m blocks as $\mathbf{x} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)})$, 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:

$$\begin{aligned}
\underset{\mathbf{x}^{(i)}}{\text{minimize}} \quad & g_i\left(\mathbf{x}^{(i)}|\mathbf{x}_t\right) \\
\text{subject to} \quad & \mathbf{x}^{(i)} \in \mathcal{X}_i,
\end{aligned} \tag{7}$$

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

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

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 value of the other blocks fixed.

In practice, the surrogate functions are usually designed so that each sub-problem (7) can be solved easily, for example in closed-form.

III. DIRECT BLOCK MAJORIZATION-MINIMIZATION ALGORITHM

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

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

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 (note that variables $\boldsymbol{\Sigma}_k$ are implicitly optimized at every iteration while optimizing either τ_k or $\mathbf{W}$). In each iteration, an upperbound of the objective function is minimized, which guarantees a monotonic decrement of the objective value.

To be precise, given a starting point $\{\{\tau_k\}^{t=0}, \mathbf{W}^{t=0}\}$, one iteratively
- updates $\{\tau_k\}^{t+1}$ for fixed $\mathbf{W} = \mathbf{W}^t$ by minimizing a set of surrogate functions $L(\tau_k|\tau_k^t, \mathbf{W})$ for $k \in [\![1, K]\!]$,
- updates $\mathbf{W}^{t+1}$ for fixed $\{\tau_k\} = \{\tau_k\}^{t+1}$ by minimizing a surrogate function $L(\mathbf{W}|\{\tau_k\}, \mathbf{W}^t)$,

until convergence, which will produce a stationary point of Problem (B). This procedure is summed up in the box Algorithm 1.

A. Update $\{\tau_k\}$ With Fixed $\mathbf{W}$

Let $\mathbf{W} = \mathbf{W}^t$ and $\boldsymbol{\Sigma} = \mathbf{W}^t\mathbf{W}^{tH}$. To lighten notation, we omit the reference on t for these variables in this part. The objective function is separable in the τ_k's, and for each of them, the following problem should be solved:

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

Eigendecompose $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, with $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_R, 0, \dots, 0)$ and $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_M]$. The objective function of (B1) can be simplified to

$$\begin{aligned}
L(\tau_k|\mathbf{W}) &= \sum_{m=1}^{M} \log (\tau_k\lambda_m + 1) + \sum_{m=1}^{M} s_{km} (\tau_k\lambda_m + 1)^{-1} \\
&= \sum_{m=1}^{R} \log (\tau_k\lambda_m + 1) + \sum_{m=1}^{R} s_{km} (\tau_k\lambda_m + 1)^{-1} \\
&\quad + \text{const.},
\end{aligned} \tag{8}$$

where $s_{km} = \|\mathbf{z}_k^H \mathbf{u}_m\|^2$. This function is the sum of quasi-convex functions, which is not necessarily quasi-convex, and has no closed-form minimizer.

Applying the block MM algorithm, we find an upperbound (surrogate function) of $L(\tau_k|\mathbf{W})$, with equality achieved at $\tau_k = \tau_k^t$, and minimize this surrogate function. The update of τ_k is derived based on the following two propositions:

Proposition 1: The function $L(\tau_k|\mathbf{W})$ in (8) can be upperbounded by the surrogate function $L(\tau_k|\tau_k^t, \mathbf{W})$ defined as

$$L\left(\tau_k|\tau_k^t, \mathbf{W}\right) = -\beta_k \log \tau_k$$
$$+ \left(\sum_{m=1}^{R} \alpha_{km}\right) \log \left(\frac{\left(\sum_{m=1}^{R} \frac{\alpha_{km}\lambda_m}{1+\lambda_m \tau_k^t}\right)}{\sum_{m=1}^{R} \alpha_{km}} \tau_k + \frac{\sum_{m=1}^{R} \frac{\alpha_{km}}{1+\lambda_m \tau_k^t}}{\sum_{m=1}^{R} \alpha_{km}}\right)$$
$$+ \text{const.}, \tag{9}$$

where

$$\alpha_{km} = s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1} + 1 \tag{10}$$

and

$$\beta_k = \sum_{m=1}^{R} s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1}. \tag{11}$$

The equality is achieved at $\tau_k = \tau_k^t$.

Proof: See Appendix A. ∎

Proposition 2: The surrogate function $L(\tau_k|\tau_k^t, \mathbf{W})$ is quasi-convex and has a unique minimizer given by

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{m=1}^{R} s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1}\right) \cdot \left(\sum_{m=1}^{R} \frac{\alpha_{km}}{1+\lambda_m \tau_k^t}\right)}{\sum_{m=1}^{R} \frac{\alpha_{km}\lambda_m}{1+\lambda_m \tau_k^t}}. \tag{12}$$

Proof: See Appendix B. ∎

B. Update $\mathbf{W}$ With Fixed $\{\tau_k\}$

Let now fix $\{\tau_k\} = \{\tau_k\}^{t+1}$. To lighten the notation, we omit the reference on t for this set of variables in this part. To obtain the update of $\mathbf{W}$, we need to solve

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

Without loss of generality assume that $\tau_k > 0$, otherwise the terms with $\tau_k = 0$ can be deleted from the summation in the objective function as they are constants.

Problem (B2) has no closed-form minimizer. As in the previous part, we are going to derive a tight upperbound of the objective function $L(\mathbf{W}|\tau_k)$, with equality achieved at $\mathbf{W} = \mathbf{W}^t$.

Proposition 3: The function

$$L(\mathbf{W}|\tau_k) = \sum_{k=1}^{K} \log \det(\boldsymbol{\Sigma}_k) + \sum_{k=1}^{K} \mathbf{z}_k^H \boldsymbol{\Sigma}_k^{-1} \mathbf{z}_k, \tag{13}$$

where $\boldsymbol{\Sigma}_k = \tau_k \mathbf{W}\mathbf{W}^H + \mathbf{I}$, can be upperbounded by the convex quadratic function

$$L\left(\mathbf{W}|\tau_k, \mathbf{W}^t\right) = \text{Tr}\left(\mathbf{W}\mathbf{H}\mathbf{W}^H\right) - \text{Tr}\left(\mathbf{L}\mathbf{W}^H\right)$$
$$- \text{Tr}\left(\mathbf{L}^H \mathbf{W}\right), \tag{14}$$

with equality achieved at $\mathbf{W} = \mathbf{W}^t$, where the matrices $\mathbf{H}$ and $\mathbf{L}$ are defined according to (41), (42), (46), and (47).

Proof: See Appendix C. ∎

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

$$\mathbf{W}^{t+1} = \mathbf{L}\mathbf{H}^{-1}.$$

To efficiently compute this solution, let the SVD of $\mathbf{W}^t$ be $\mathbf{W}^t = \mathbf{U}\mathbf{S}\mathbf{V}^H$. Then

$$\mathbf{L} = \sum_{k=1}^{K} \mathbf{L}_k^t = \left(\sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1}\right) \mathbf{V}^H, \tag{15}$$

and

$$\mathbf{H} = \sum_{k=1}^{K} \left(\left(\mathbf{W}^t\right)^H \mathbf{W}^t + \tau_k^{-1}\mathbf{I}\right)^{-1} + \sum_{k=1}^{K} \mathbf{H}_k^t$$
$$= \sum_{k=1}^{K} \mathbf{V}\left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1} \mathbf{S}^H \mathbf{U}^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1} \mathbf{V}^H$$
$$+ \sum_{k=1}^{K} \mathbf{V}\left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1} \mathbf{V}^H, \tag{16}$$

where $\mathbf{S}^2 = \mathbf{S}^H\mathbf{S}$ (since $\mathbf{S}$ is a real diagonal matrix). $\mathbf{W}^{t+1}$ can therefore be computed as

$$\mathbf{W}^{t+1} = \left(\sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}\mathbf{S}\left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1}\right)$$
$$\times \left(\sum_{k=1}^{K} \left(\left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1}\right.\right.$$
$$+ \left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1} \mathbf{S}^H \mathbf{U}^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}\mathbf{S}$$
$$\left.\left.\times \left(\tau_k^{-1}\mathbf{I} + \mathbf{S}^2\right)^{-1}\right)\right)^{-1} \mathbf{V}^H. \tag{17}$$

Algorithm 1: Direct Block Majorization-Minimization Algorithm

---

1: Compute the sample covariance matrix
   $\hat{\boldsymbol{\Sigma}}^{\text{SCM}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$.
2: Eigendecompose $\hat{\boldsymbol{\Sigma}}^{\text{SCM}}$ as $\hat{\boldsymbol{\Sigma}}^{\text{SCM}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_M)$ with $\lambda_1 \geq \cdots \geq \lambda_M$.
3: Initialize $\mathbf{W}$ to be $\mathbf{W} = \mathbf{U}\text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_R}, 0, \ldots, 0)$, and $\{\tau_k\}_{k=1}^{K}$ to be arbitrary positive real numbers.
4: **repeat**
5:   Update $\{\tau_k\}_{k=1}^{K}$ with (10) and (12).
6:   Update $\mathbf{W}$ with (17).
7: **until** Some convergence criterion is met
8: Eigendecompose $\mathbf{W}\mathbf{W}^H$ as $\mathbf{V}\mathbf{D}\mathbf{V}^H$.
9: $\hat{\boldsymbol{\Pi}}_c = \mathbf{V}\mathbf{V}^H$.

C. Convergence Analysis

Proposition 4: Any limit point of the pair $\{\{\tau_k^t\}, \mathbf{W}^t\}$ generated by Algorithm 1 is a stationary point of Problem (B).

Proof: We have proved the quasi-convexity and the uniqueness of the minimizer of the surrogate functions $L(\tau_k|\tau_k^t, \mathbf{W}^t)$ and $L(\mathbf{W}|\tau_k, \mathbf{W}^t)$ in the previous subsections. The algorithm convergence is a direct application of Theorem 2 (a) in [46]. ∎

IV. EIGENSPACE BLOCK MAJORIZATION-MINIMIZATION ALGORITHM

As $\mathrm{rank}(\mathbf{\Sigma}) \leq R$, the variable $\mathbf{\Sigma}$ can be reparameterized by its eigendecomposition:

$$\mathbf{\Sigma} = \sum_{r=1}^{R} c_r \mathbf{v}_r \mathbf{v}_r^H.$$

Problem (A) can then be rewritten as

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

Following the same methodology as in the previous section, we partition the variables as $\{\{\tau_k\},\{c_r\},\{\mathbf{v}_r\}\}$ and derive an algorithm that updates the blocks in cyclic order by minimizing an upperbound of the objective.

Given a starting point $\{\{\tau_k\}^{t=0},\{c_r\}^{t=0},\{\mathbf{v}_r\}^{t=0}\}$, one iteratively
- updates $\{\tau_k\}^{t+1}$ for fixed $\{c_r\} = \{c_r\}^t$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$ by minimizing a set of surrogates functions $L(\tau_k | \tau_k^t, \{c_r\},\{\mathbf{v}_r\})$ for $k \in [\![1, K]\!]$,
- updates $\{c_r\}^{t+1}$ for fixed $\{\tau_k\} = \{\tau_k\}^{t+1}$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$ by minimizing a set of surrogates functions $L(c_k | \{\tau_k\}, c_r^t, \{\mathbf{v}_r\})$ for $r \in [\![1, R]\!]$,
- updates $\{\mathbf{v}_r\}^{t+1}$ for fixed $\{\tau_k\} = \{\tau_k\}^{t+1}$ and $\{c_r\} = \{c_r\}^{t+1}$ by minimizing the objective w.r.t. $\{\mathbf{v}_r\}$ under orthonormality constraints. This step is done by iteratively minimizing surrogate functions $L(\{\mathbf{v}_r\} | \{\tau_k\}\{c_r\},\{\mathbf{v}_r\}^t)$ (so there is an inner-loop in the block MM iterations).

This procedure is summed up in the box Algorithm 2.

A. Update $\{\tau_k\}$ with Fixed $\{c_r\}$ and $\{\mathbf{v}_r\}$

Let $\{c_r\} = \{c_r\}^t$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$. To lighten notation, we omit the reference on t for these variables in this part. The objective function is separable in the τ_k's, and for each of them the following problem (noticing that $\mathbf{\Sigma}$ has $M - R$ zero eigenvalues) should be solved:

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

with $s_{kr} = \|\mathbf{z}_k^H \mathbf{v}_r\|^2$. Notice that the objective function of (C1) has the same form as (8), therefore we can apply Propositions 1 and 2 to obtain the τ_k's updates as

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{r=1}^{R} s_{kr} \frac{\tau_k^t c_r}{\tau_k^t c_r + 1}\right) \cdot \left(\sum_{r=1}^{R} \frac{\alpha_{kr}}{1 + c_r \tau_k^t}\right)}{\sum_{r=1}^{R} \frac{\alpha_{kr} c_r}{1 + c_r \tau_k^t}}, \quad (18)$$

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

B. Update $\{c_r\}$ With Fixed $\{\tau_k\}$ and $\{\mathbf{v}_r\}$

Let $\{\tau_k\} = \{\tau_k\}^{t+1}$ and $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$. As before we omit the reference on t for these variables in this part. The objective

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

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

with $s_{kr} = \|\mathbf{z}_k^H \mathbf{v}_r\|^2$. 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 Propositions 1 and 2 to obtain the c_r's updates as

$$c_r^{t+1} = \frac{1}{K} \cdot \frac{\left(\sum_{k=1}^{K} s_{kr} \frac{\tau_k c_r^t}{\tau_k c_r^t + 1}\right) \cdot \left(\sum_{k=1}^{K} \frac{\alpha_{kr}}{1 + c_r^t \tau_k}\right)}{\sum_{k=1}^{K} \frac{\alpha_{kr} \tau_k}{1 + c_r^t \tau_k}}, \quad (19)$$

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

C. Update $\{\mathbf{v}_r\}$ With Fixed $\{\tau_k\}$ and $\{c_r\}$

With $\{\tau_k\}$ and $\{c_r\}$ fixed, minimizing the objective w.r.t. $\{\mathbf{v}_r\}$ is equivalent to solving the problem

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

where

$$\mathbf{M}_r = \sum_{k=1}^{K} \frac{\tau_k c_r}{\tau_k c_r + 1} \mathbf{z}_k \mathbf{z}_k^H. \quad (20)$$

We start with the following proposition.

Proposition 5: The function

$$L\left(\{\mathbf{v}_r\} \,|\, \{\tau_k\},\{c_r\}\right) = \sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r \quad (21)$$

can be lowerbounded by the surrogate function

$$
\begin{aligned}
& L\left(\{\mathbf{v}_r\} \,|\, \{\tau_k\},\{c_r\},\{\mathbf{v}_r\}^t\right) \\
& = \sum_{r=1}^{R} \left[\left(\mathbf{v}_r^t\right)^H \mathbf{M}_r \mathbf{v}_r + \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r^t\right] + \text{const.} \quad (22)
\end{aligned}
$$

with equality achieved at $\{\mathbf{v}_r\} = \{\mathbf{v}_r\}^t$.

Proof: As the matrices $\mathbf{M}_r$ are Hermitian positive semidefinite, the objective function is convex, and therefore is minorized by its first order Taylor expansion at $\mathbf{v}_r^t$, which is the considered surrogate function (22). ∎

Maximizing $L(\{\mathbf{v}_r\} | \{\tau_k\},\{c_r\},\{\mathbf{v}_r\}^t)$, under orthonormality constraints on the $\{\mathbf{v}_r\}$, is equivalent to solving

$$
\begin{aligned}
\underset{\mathbf{V}}{\text{maximize}} \quad & \mathrm{Tr}\left(\mathbf{A}^H \mathbf{V}\right) + \mathrm{Tr}\left(\mathbf{V}^H \mathbf{A}\right) \\
\text{subject to} \quad & \mathbf{V}^H \mathbf{V} = \mathbf{I}, \quad \text{(23)}
\end{aligned}
$$

where

$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_R] \quad (24)$$

and

$$\mathbf{A} = \left[\mathbf{M}_1 \mathbf{v}_1^t; \ldots; \mathbf{M}_R \mathbf{v}_R^t\right], \quad (25)$$

i.e., $\mathbf{A}$ is constructed by stacking the column vectors $\mathbf{M}_r \mathbf{v}_r^t$. Problem (23) is equivalent to

$$\underset{\mathbf{V}}{\text{minimize}} \quad \|\mathbf{A} - \mathbf{V}\|_F^2$$
$$\text{subject to} \quad \mathbf{V}^H \mathbf{V} = \mathbf{I}. \quad (26)$$

Let the thin SVD of $\mathbf{A}$ be $\mathbf{A} = \mathbf{V}_A \mathbf{D}_A \mathbf{U}_A^H$, the optimal $\mathbf{V}$ is given by [45]

$$\mathbf{V}^{t+1} = \mathbf{V}_A \mathbf{U}_A^H. \quad (27)$$

Note that in this step, we maximize the objective function of (C3) by iteratively maximizing a series of surrogates (steps 7–11 in Algorithm 2). These iterations lead to a local solution of (C3). 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 this inner loop provides a faster decreasing rate of the objective value than updating the $\mathbf{v}_r$'s only once.

Algorithm 2: Eigenspace Block Majorization-Minimization Algorithm

---

1: Compute the sample covariance matrix $\hat{\mathbf{\Sigma}}^{\text{SCM}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$.
2: Initialize $\{c_r\}_{r=1}^{R}$ to be the first R leading eigenvalues of $\hat{\mathbf{\Sigma}}^{\text{SCM}}$, $\{\mathbf{v}_r\}_{r=1}^{R}$ to be the corresponding eigenvectors, and $\{\tau_k\}_{k=1}^{K}$ to be arbitrary positive real numbers.
3: **repeat**
4:     Update τ_k with (18).
5:     Update c_r with (19).
6:     Compute $\mathbf{M}_r$ with (20).
7:     **repeat** (*optional inner loop*)
8:         Compute $\mathbf{A}$ with (25).
9:         Decompose $\mathbf{A}$ as $\mathbf{A} = \mathbf{V}_A \mathbf{D}_A \mathbf{U}_A^H$ (Thin SVD).
10:        Update $\mathbf{V}$ as $\mathbf{V} = \mathbf{V}_A \mathbf{U}_A^H$.
11:     **until** Some convergence criterion is met
12: **until** Some convergence criterion is met
13: $\hat{\mathbf{\Pi}}_c = \mathbf{V} \mathbf{V}^H$.

D. Convergence Analysis

Since the constraint set of Problem (C) is non-convex, the convergence result provided in [46] cannot be applied here. 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 $(\{\tau_k\}^t, \{c_r\}^t, \{\mathbf{v}_r\}^t)$ remains unknown. Nevertheless, Section VI will show that the numerical performance of Algorithm 2 is satisfactory.

V. HYPER-PARAMETERS ESTIMATION

The clutter subspace estimation algorithms proposed in this paper assume the prior knowledge of two hyper-parameters: the clutter rank R and the white Gaussian noise power σ^2. These assumptions are made for deriving a valid and clear theoretical framework, focusing on the core of the subspace estimation problem. In practice, the proposed algorithms can be called using plug-in estimates $\hat{R}$ and $\hat{\sigma}^2$.

To estimate the clutter rank, there are several methods present in the literature (*e.g.*, [8]–[10] and references therein). Moreover, in some applications, the clutter rank can be derived from the geometry of the system (such as in STAP [11] thanks to the Brennan rule [12]).

Regarding the WGN power, the algorithms derived in this paper can also be extended to estimate it by introducing one more block for the parameter σ^2. The update of this block σ^2 for $\{\mathbf{\Sigma}_k\}$ fixed can be obtained by majorizing the objective function *w.r.t.* σ^2 in a way similar to the parameters $\{\tau_k\}$. To account for possible knowledge on the lower and upper bounds of the WGN power [47], [48] (denoted by σ_l^2 and σ_u^2, respectively), one can further impose the constraint that $\sigma_l^2 \leq \sigma^2 \leq \sigma_u^2$. For the sake of simplicity and presentation clarity, we focused on the part of estimating $\{\mathbf{\Sigma}_k\}$, which is the bottleneck of the considered problem. Estimating σ^2 jointly is therefore left as a potential extension of this work.

VI. NUMERICAL RESULTS

This section is devoted to numerical simulations to illustrate the performance of different CM estimators in the considered context. In particular, we will study the following clutter subspace projector estimators:

- $\hat{\mathbf{\Pi}}_{\text{SCM}}$: the clutter subspace projector estimator derived from the EVD of the SCM defined as $\hat{\mathbf{\Sigma}}_{\text{SCM}} = \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H / K$.
- $\hat{\mathbf{\Pi}}_{SFPE}$: the clutter subspace projector estimator derived from the EVD of the Shrinkage-FPE (SFPE) [34]–[36], which is a regularized Tyler's estimator [24] defined as the unique solution of the following fixed-point equation:

$$\hat{\mathbf{\Sigma}}_{S-FPE}(\beta) = \frac{(1-\beta)M}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_{S-FPE}^{-1}(\beta) \mathbf{z}_k} + \beta \mathbf{I},$$

  for $\beta \in (\max(0, 1 - K/M), 1]$. This estimator can be computed with the simple fixed point iterations provided in [34]–[36]. Since there is no rule to adaptively select the optimal shrinkage parameter β for the considered 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 $\epsilon = 10^{-2}$.
- $\hat{\mathbf{\Pi}}_{MLE-MM1}$: the clutter subspace projector MLE computed with the direct block-MM Algorithm in Section III.
- $\hat{\mathbf{\Pi}}_{MLE-MM2}$: the clutter subspace projector MLE computed with the Eigenspace block-MM Algorithm in Section IV.
- $\hat{\mathbf{\Pi}}_{MLE}$: the clutter subspace projector MLE under high CNR assumption, computed with Algorithm 1 provided in [44]. This algorithm will be referred to as "Algorithm 3" in the rest of the paper.
- $\hat{\mathbf{\Pi}}_{A-MLE}$: the clutter subspace projector Approached MLE under high CNR assumption, computed with Algorithm 2 provided in [44]. This algorithm will be referred to as "Algorithm 4" in the rest of the paper.

A. Validation Simulations and Algorithm Complexity

Simulation Parameters: Samples $\mathbf{z}_k$ are generated according to the LR Compound Gaussian plus WGN model described in
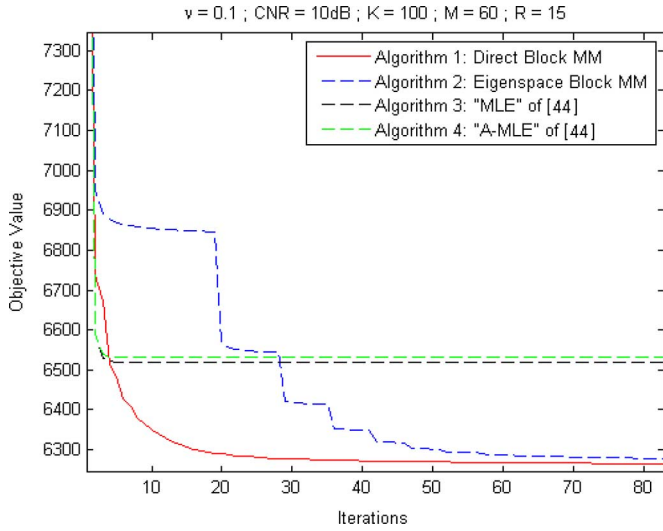
Fig. 1. Objective value versus number of iterations for different MLE algorithms: Direct Block-MM (red), Eigenspace Block-MM (blue), "MLE" algorithm of [44] (black), "A-MLE" algorithm of [44] (green). $K = 100$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$ dB.

Fig. 2. Objective value versus computation time for different MLE algorithms: Direct Block-MM (red), Eigenspace Block-MM (blue), "MLE" algorithm of [44] (black), "A-MLE" algorithm of [44] (green). $K = 100$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$ dB.

Section II: $\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k$. 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 CNR, defined as CNR $= \mathbb{E}(\tau) \mathrm{Tr}(\mathbf{\Sigma})/(R\sigma^2)$, to a given value.

Fig. 1 displays a typical realization of the objective value versus the number of iterations of different algorithms. The objective value at each inner loop is also displayed for the proposed algorithms. One can observe that the MLE and A-MLE algorithms from [44] 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). The MLE Algorithm provides a slightly better objective value than the A-MLE algorithm, thanks to the use of the modified gradient descent algorithm [45] instead of an EVD relaxation for updating the subspace estimate. Contrary to these algorithms, the Block MM algorithm (Algorithm 1) converges to a critical point with a smaller objective value. We also notice that Algorithm 2 converges in practice to the same point as Algorithm 1, i.e., a critical point.

Fig. 2 displays a typical another realization of the objective value versus the time of computation. It illustrates that despite a fast convergence, Algorithm 3 has a slow computation due to the use of the modified gradient descent [45]. It also shows that Algorithm 2 is less computationally intensive than Algorithm 1 since it requires less time to converge. This can be explained by the fact that constructing the update of $\mathbf{W}$ (step 6) in Algorithm 1 has a complexity that grows linearly with the sample size. On the contrary, the inner loops (steps 7–11) in Algorithm 2 involve
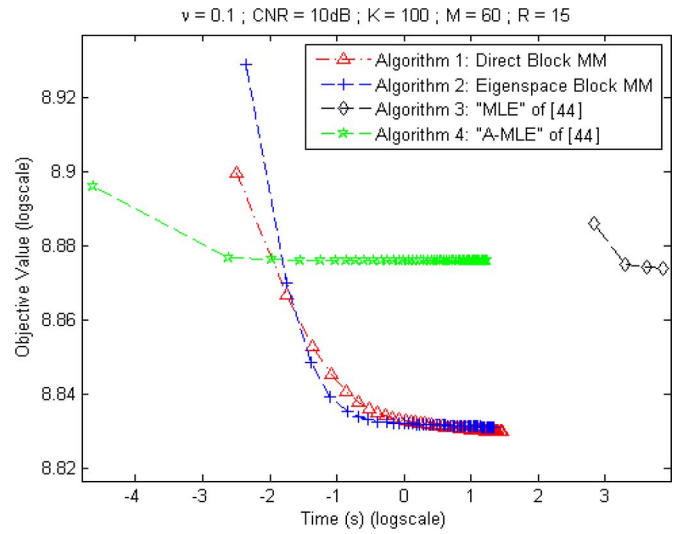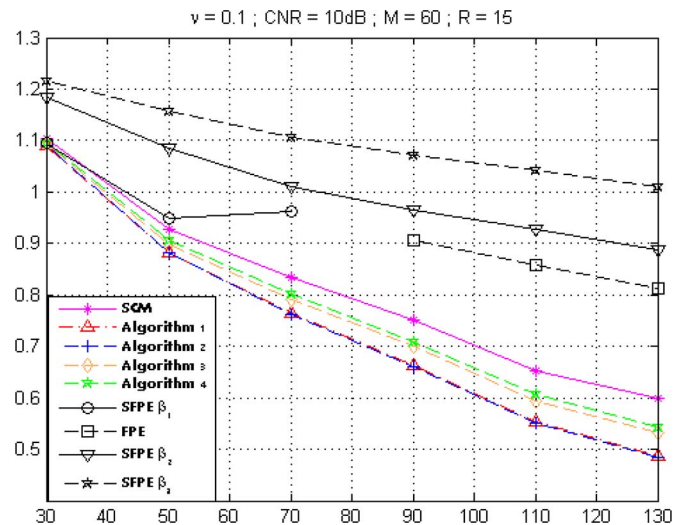

Fig. 3. Mean NMSE of the estimators $\hat{\mathbf{\Pi}}_{\mathrm{SCM}}$ (magenta), $\hat{\mathbf{\Pi}}_{MLE}$ (orange), $\hat{\mathbf{\Pi}}_{A-MLE}$ (green), $\hat{\mathbf{\Pi}}_{MLE-MM1}$ (red), $\hat{\mathbf{\Pi}}_{MLE-MM2}$ (blue), $\hat{\mathbf{\Pi}}_{SFPE}$ for different β (black), versus the number of samples K. $M = 60$, $R = 15$, $\nu = 0.1$, $\rho = 0.9$, CNR $= 10$ dB.

the SVD of a matrix of fixed dimension, which is not costly in comparison.

Fig. 3 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 estimators $\hat{\mathbf{\Pi}}_{\mathrm{SCM}}$, $\hat{\mathbf{\Pi}}_{SFPE}$ (with β_1, β_2 and β_3), $\hat{\mathbf{\Pi}}_{MLE-MM1}$, $\hat{\mathbf{\Pi}}_{MLE-MM2}$, $\hat{\mathbf{\Pi}}_{MLE}$ and $\hat{\mathbf{\Pi}}_{A-MLE}$ for a given configuration (computed over 100 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 [44] lead to slightly higher NMSE, yet better than the SFPE with various β, and the SCM.

Fig. 4 displays the mean computation time (over 100 Monte Carlo simulations) of the algorithms for the same set of parameters as in previous figures. The stop criterion for each algorithm is when the minimum achievable objective value is reached up
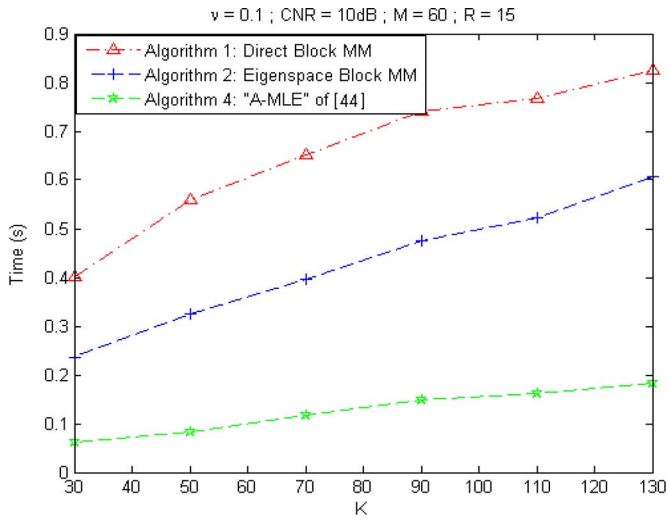
Fig. 4. Mean Computation time of each MLE algorithm over 100 Monte-Carlo Simulation, versus K. Same parameters as Fig. 3. Algorithms stop criterion: minimum achievable objective value is reached up to 1%. Stopping criterion for Algorithm 2 inner loop: $\mathbf{V}$: $\frac{\|\mathbf{v}^{t+1} - \mathbf{v}^t\|_F^2}{M \times R} \leq 10^{-6}$. Computer specifics: Intel(R) Core(TM) i5-3230M CPU @2.6 Ghz.

to 1% (The limit value is obtained by a prior run of the considered algorithm for a sufficiently long time until the objective value stabilizes). This criterion is chosen to provide a fair comparison in terms of computational time for algorithms with different convergence rates. The computation time of MLE Algorithm 3 is not displayed since this algorithm (relying on a costly gradient descent) always reaches a maximum time stop criterion, greatly larger than the presented values (as seen in Fig. 2). One can observe that Algorithm 2 and 4 are the "fastest" algorithms, which is due to their low computational cost at each step. Algorithm 4 is the less computationally intensive since it has a fast convergence, as observed in Figs. 1 and 2. Nevertheless, Algorithm 1 and 2 offer better performance (see Fig. 3). We also point out that, although Algorithm 2 is computationally faster than Algorithm 1, the latter has more theoretical guarantees (convergence to a critical point). Both Figs. 3 and 4 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.

B. Application to LR-STAP Filtering

STAP is a technique used in airborne phased array radar to detect a moving target embedded in an interference background such as jamming or strong clutter [11]. The radar receiver consists of an array of Q antenna elements processing P pulses in a coherent processing interval. The received signal is $\mathbf{z} = \alpha\mathbf{p} + \mathbf{c} + \mathbf{n}$, where α is the target power and $\mathbf{p}$ is the so-called STAP steering vector, $\mathbf{c}$ is the heterogeneous ground clutter and $\mathbf{n}$ is the thermal noise. It is important to notice that the application fits the model considered in this paper since in side looking STAP, the clutter CM is known to be LR. Moreover, the rank of the clutter CM can be evaluated thanks to the Brennan Rule [12]. Since the clutter could behave heterogeneously, the total interference can be modeled as LR Compound Gaussian (the ground clutter) plus WGN (the thermal noise).

The theoretical optimal STAP filter is $\mathbf{w}_{\text{opt}} = \boldsymbol{\Sigma}_{\text{tot}}^{-1}\mathbf{p}$ [11]. In the context of an LR clutter, it is well-known that a correct

sub-optimal filter [6], [14] is $\mathbf{w}_{lr} = \boldsymbol{\Pi}_c^{\perp}\mathbf{p} = (\mathbf{I} - \boldsymbol{\Pi}_c)\mathbf{p}$. In practice, $\boldsymbol{\Pi}_c^{\perp}$ is unknown and has to be estimated using the samples $\{\mathbf{z}_k\}$ to perform adaptive filtering. The adaptive LR filter is then $\hat{\mathbf{w}}_{lr} = \hat{\boldsymbol{\Pi}}_c^{\perp}\mathbf{d} = (\mathbf{I} - \hat{\boldsymbol{\Pi}}_c)\mathbf{p}$, with $\hat{\boldsymbol{\Pi}}_c$ being an estimate of the clutter subspace projector. Consequently, the performance of the LR filters directly relies on the estimation accuracy of $\boldsymbol{\Pi}_c$. To evaluate the performance, we use the SINR-Loss criterion [11], which is the expected ratio between the $SINR_{out}$ computed for $\hat{\mathbf{w}}_{lr}$, and $SINR_{max}$ computed for the optimal filter $\mathbf{w} = \boldsymbol{\Sigma}_{\text{tot}}^{-1}\mathbf{d}$. For an estimate of the clutter subspace $\hat{\boldsymbol{\Pi}}_c$, the SINR-Loss expression is given by

$$
\begin{aligned}
\rho_{\hat{\boldsymbol{\Pi}}_c} &= \frac{SINR_{out}}{SINR_{max}} \\
&= \mathbb{E}\left(\frac{(\mathbf{p}^H\hat{\boldsymbol{\Pi}}_c^{\perp}\mathbf{p})^2}{\left(\mathbf{p}^H\hat{\boldsymbol{\Pi}}_c^{\perp}\boldsymbol{\Sigma}_{\text{tot}}\hat{\boldsymbol{\Pi}}_c^{\perp}\mathbf{p}\right)\left(\mathbf{p}^H\boldsymbol{\Sigma}_{\text{tot}}^{-1}\mathbf{p}\right)}\right). \quad (28)
\end{aligned}
$$

We consider the following STAP configuration. The number Q of sensors is 8 and the number P of coherent pulses is also 8. The center frequency and the bandwidth are equal to $f_0 = 450$ MHz and $B = 4$ MHz, respectively. The radar velocity is 100 m/s. The inter-element spacing is $d = \frac{c}{2f_0}$ (c is the celerity of light) and the pulse repetition frequency is $f_r = 600$ Hz. The clutter CM is computed according to the model described in [11]. The rank of the clutter CM $\boldsymbol{\Sigma}$ is evaluated from the Brennan rule and is equal to $R = 15 \ll 64$, therefore, the LR assumption is valid. The texture PDF is a Gamma distribution of shape parameter $\nu = 0.1$ and scale parameter $1/\nu$, so the clutter follows a K-distribution. The target $\mathbf{p}$ has a celerity of $V = 35$ m/s and is at $+10°$ Azimuth. CNR is defined as CNR $= \mathbb{E}(\tau)\text{Tr}(\boldsymbol{\Sigma}_c)/(\sigma^2 R)$, and we set $\sigma^2 = 1$.

Since the two proposed algorithms have been shown to reach identical performance in Fig. 3, we only display the results for the projector estimate computed with the fastest algorithm, namely Algorithm 2 (Eigenspace Block MM), referred to as $\hat{\boldsymbol{\Pi}}_{EBMM}$ in the following. Algorithm 3 and Algorithm 4 (respectively "MLE" and "A-MLE" of [44]) have also been shown to reach identical performance, therefore we only display results for the projector estimate computed with Algorithm 3, referred to as $\hat{\boldsymbol{\Pi}}_{A-MLE}$.

Fig. 5 displays the SINR-Loss versus K for various clutter configurations: from average to high CNR (10 dB, 20 dB, and 30 dB) and for mildly ($\nu = 1$) and highly ($\nu = 0.1$) heterogeneous clutter. First, one can state that $\hat{\boldsymbol{\Pi}}_{EBMM}$ reaches the best SINR-Loss for all configurations. Under standard conditions (20 dB and 30 dB of CNR and $\nu = 1$), $\hat{\boldsymbol{\Pi}}_{SCM}$ and $\hat{\boldsymbol{\Pi}}_{A-MLE}$ provide a clutter subspace projector estimate that is close to the $\hat{\boldsymbol{\Pi}}_{EBMM}$, hence they have similar performances in terms of SINR-Loss. One can observe that under these conditions, the classical -3 dB SINR-Loss of the filter built from $\hat{\boldsymbol{\Pi}}_{SCM}$ is reached with $K \simeq 2R$ samples, as theoretically expected from [14], [22]. $\hat{\boldsymbol{\Pi}}_{FPE}$ also reaches a performance close to $\hat{\boldsymbol{\Pi}}_{MLE}$ but requires $K > M$ samples to be computed. $\hat{\boldsymbol{\Pi}}_{SFPE}$ can be computed with a smaller K, but its SINR-Loss decreases as β increases (which is expected since a larger β implies a higher bias). When standard conditions are not met (low/average CNR or highly heterogeneous clutter), one can see that the performance of $\hat{\boldsymbol{\Pi}}_{FPE}$ and $\hat{\boldsymbol{\Pi}}_{SFPE}$ greatly drops compared to other
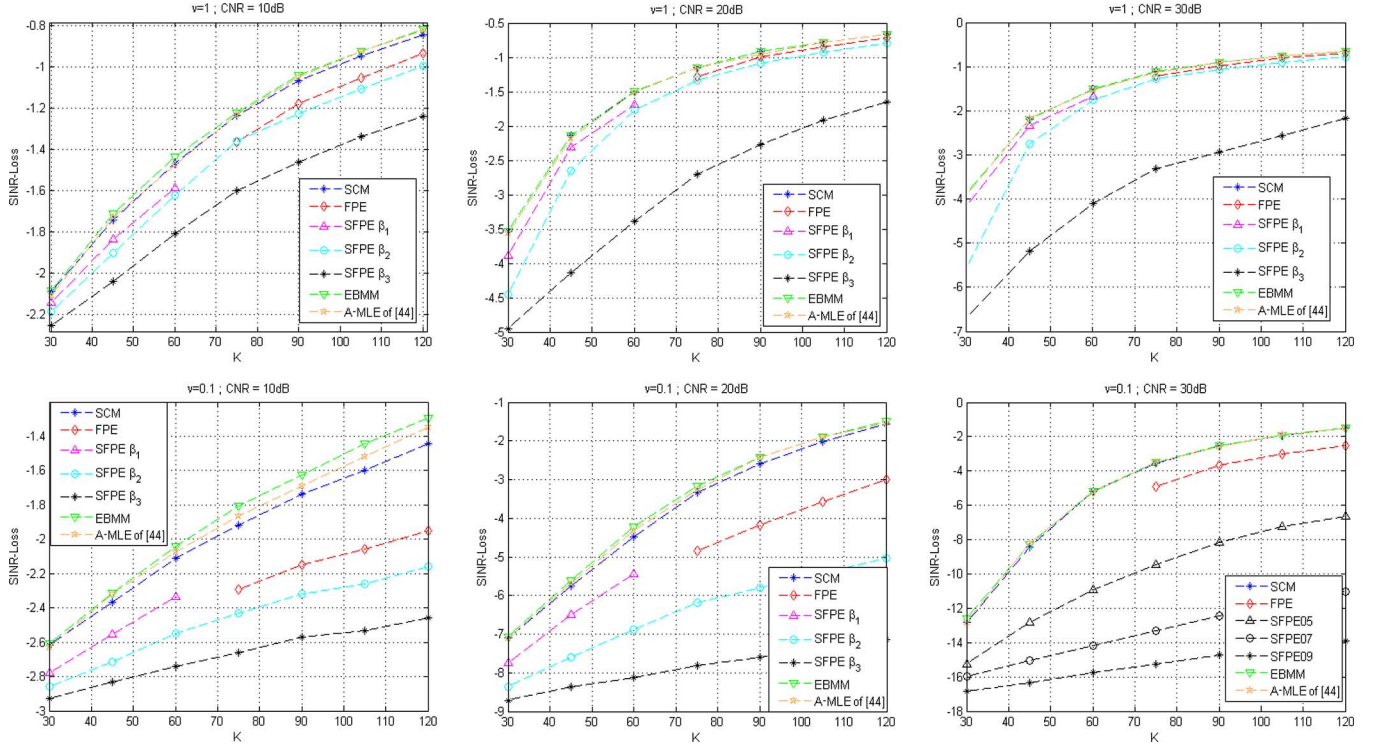
Fig. 5. Mean SINR-Loss of the adaptive LR filters built from various clutter subspace projector estimators: $\hat{\mathbf{\Pi}}_{\mathrm{SCM}}$ (blue), $\hat{\mathbf{\Pi}}_{FPE}$ (red), $\hat{\mathbf{\Pi}}_{SFPE}$ for β_1 (magenta), β_2 (cyan) and β_3 (black), $\hat{\mathbf{\Pi}}_{EBMM}$ computed with Algorithm 2: EBMM (green), $\hat{\mathbf{\Pi}}_{A-MLE}$ computed with algorithm "A-MLE" from [44] (orange). The SINR-Loss is presented for various clutter configurations. Columns from the left to the right: CNR = 10 dB, CNR = 20 dB, CNR = 30 dB. Up row: $\nu = 1$ (mildly heterogeneous clutter), Down row: $\nu = 0.1$ (highly heterogeneous clutter).

estimators. On the contrary, $\hat{\mathbf{\Pi}}_{EBMM}$ offers a better resistance to these conditions and also outperforms $\hat{\mathbf{\Pi}}_{\mathrm{SCM}}$ and $\hat{\mathbf{\Pi}}_{A-MLE}$.

C. Note on the Robustness to a Wrong Rank Evaluation

One can not directly study the estimation accuracy (NMSE) on $\mathbf{\Pi}_c$ for estimators with various ranks since projectors on subspaces of different dimension are not comparable objects. Such a comparison in NMSE can be done, for example, in terms of reconstructed total CM. For this criterion, [44] showed that the proposed approach is robust to a misevaluated rank.

The robustness to a wrong rank evaluation can also be relevantly studied through a criterion linked to the considered application. For example, for the LR-filtering problem:

- If the rank is under-evaluated, then a portion of the interference may not be completely canceled by the filter, leading to poor performances in terms of output SINR.
- If the rank is over-evaluated, the interference may be completely canceled, but the filter can also cancel a part of the target response, also leading to a lower output SINR.

However, these phenomena depend highly on the clutter and target parameters, which is why a corresponding study goes beyond the scope of this paper, as it is more a problem related to a process design rather than a subspace estimation accuracy.

VII. CONCLUSION

In this paper, we derived two algorithms based on the block MM framework for computing the MLE of the CM parameter when the samples are modeled as the sum of an LR Compound Gaussian (with known rank) and a WGN. This complex problem

was initially considered in [43] and [44], where several algorithms were proposed. The new algorithms proposed in this paper enjoy two major advantages: firstly, they do not rely on any heuristic (as in [43]) or high CNR assumption (as in [44]), thus they reach the exact MLE (at least locally) of the considered problem with no approximation; secondly, they are computationally faster and easier to implement than the ones derived in [43] and [44], as they do not require the use of the modified gradient descent algorithm [45], which can be computationally expensive. The performance of the related estimators was illustrated on Space Time Adaptive Processing filtering for airborne radar simulations.

As a side benefit of this study, we emphasize that the proposed algorithms allow a full estimation of the CM parameters, while presented simulations only focus on the clutter subspace projector estimation. They could therefore be suitable for other applications that involve the estimation of Low Rank structured Covariance Matrices.

APPENDIX A
PROOF OF PROPOSITION 1

Before going to the formal proof of Proposition 1, we need to first state the two following lemmas:

Lemma 6 ((5) of [49]): For $x_i > 0$ and $\alpha_i \in \mathbb{R}$, the function $\prod_{i=1}^{n} x_i^{\alpha_i}$ can be lower bounded by

$$\prod_{i=1}^{n} x_i^{\alpha_i} \geq \prod_{j=1}^{n} \left(x_j^t \right)^{\alpha_j} \left(1 + \sum_{i=1}^{n} \alpha_i \log x_i - \sum_{i=1}^{n} \alpha_i \log x_i^t \right),$$

where x_i^t is some arbitrary positive real number. The equality is achieved at $x_i = x_i^t$. More specifically, in the uni-dimensional case $n = 1$ and for $\alpha_1 = 1$, one has

$$x \geq x^t \left(1 + \log(x) - \log(x^t)\right) \tag{29}$$

Lemma 7: The function $\sum_{i=1}^{n} \alpha_i \log(1 + c_i x)$ with $\alpha_i > 0$ can be upperbounded by

$$\sum_{i=1}^{n} \alpha_i \log \left(1 + c_i x\right) \leq \sum_{i=1}^{n} \alpha_i \log \left(1 + c_i x^t\right)$$
$$+ \left(\sum_{i=1}^{n} \alpha_i\right) \log \left(\frac{\sum_{i=1}^{n} \alpha_i \cdot \frac{1+c_i x}{1+c_i x^t}}{\sum_{i=1}^{n} \alpha_i}\right), \quad (30)$$

where x^t is some arbitrary real number such that $1 + c_i x^t > 0$, $\forall i$. The equality is achieved at $x = x^t$

Proof: The $\log(\cdot)$ function is concave, and Jensen's inequality states that

$$\log \left(\frac{\sum_{i=1}^{n} \alpha_i z_i}{\sum_{i=1}^{n} \alpha_i}\right) \geq \frac{\sum_{i=1}^{n} \alpha_i \log z_i}{\sum_{i=1}^{n} \alpha_i} \tag{31}$$

for $\alpha_i > 0$ and $z_i > 0$, and equality is achieved if the z_i's are equal. Let $z_i = \frac{1+c_i x}{1+c_i x^t}$, we have

$$\frac{\sum_{i=1}^{n} \alpha_i \log \left(\frac{1+c_i x}{1+c_i x^t}\right)}{\sum_{i=1}^{n} \alpha_i}$$
$$= \frac{\sum_{i=1}^{n} \alpha_i \log \left(1 + c_i x\right) - \sum_{i=1}^{n} \alpha_i \log \left(1 + c_i x^t\right)}{\sum_{i=1}^{n} \alpha_i}$$
$$\leq \log \left(\frac{\sum_{i=1}^{n} \alpha_i \cdot \frac{1+c_i x}{1+c_i x^t}}{\sum_{i=1}^{n} \alpha_i}\right). \tag{32}$$

Rearranging the terms leads to the inequality (30), with equality achieved at $x = x^t$, which corresponds to $z_i = 1$, $\forall i = 1, \ldots, n$. ∎

We can now turn to the proof of Proposition 1:

Proof: Ignoring the constant term, $L(\tau_k | \mathbf{W})$ can be written as:

$$L\left(\tau_k | \mathbf{W}\right)$$
$$= \sum_{m=1}^{R} \log \left(\tau_k \lambda_m + 1\right) + \sum_{m=1}^{R} s_{km} \left(1 - \frac{\tau_k \lambda_m}{\tau_k \lambda_m + 1}\right)$$
$$= \sum_{m=1}^{R} \log \left(\tau_k \lambda_m + 1\right) - \sum_{m=1}^{R} s_{km} \frac{\tau_k \lambda_m}{\tau_k \lambda_m + 1} + \text{const.} \tag{33}$$

First, applying (29) of Lemma 6 to the second term of (33) (with x parameterized as $\frac{\tau_k \lambda_m}{\tau_k \lambda_m + 1}$) yields

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

$$\leq \sum_{m=1}^{R} \left(s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1} + 1\right) \log \left(1 + \lambda_m \tau_k\right)$$
$$- \left(\sum_{m=1}^{R} s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1}\right) \log \tau_k + \text{const.} \tag{34}$$

where the equality is achieved at $\tau_k = \tau_k^t$. This upperbound is convex in $\tilde{\tau}_k \triangleq \log \tau_k$, but the minimizer cannot be computed in closed-form. Denote

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

We now apply (30) of Lemma 7 to the first term of the upperbound in (34), which leads to:

$$L\left(\tau_k | \mathbf{W}\right)$$
$$\leq \sum_{m=1}^{R} \alpha_{km} \log \left(1 + \lambda_m \tau_k\right) - \beta_k \log \tau_k + \text{const.}$$
$$\leq \left(\sum_{m=1}^{R} \alpha_{km}\right) \log \left(\frac{\sum_{m=1}^{R} \alpha_{km} \frac{1+\lambda_m \tau_k}{1+\lambda_m \tau_k^t}}{\sum_{m=1}^{R} \alpha_{km}}\right) - \beta_k \log \tau_k + \text{const.}$$
$$= \left(\sum_{m=1}^{R} \alpha_{km}\right) \log \left(\frac{\left(\sum_{m=1}^{R} \frac{\alpha_{km} \lambda_m}{1+\lambda_m \tau_k^t}\right)}{\sum_{m=1}^{R} \alpha_{km}} \tau_k + \frac{\sum_{m=1}^{R} \frac{\alpha_{km}}{1+\lambda_m \tau_k^t}}{\sum_{m=1}^{R} \alpha_{km}}\right)$$
$$- \beta_k \log \tau_k + \text{const.} \tag{35}$$

Ignoring the constant term, we arrive at the surrogate function $L(\tau_k | \tau_k^t, \mathbf{W})$ defined in Proposition 1. Note that the equality still holds at $\tau_k = \tau_k^t$. ∎

APPENDIX B
PROOF OF PROPOSITION 2

Proof: The surrogate function, ignoring the constant term, has the form

$$L\left(\tau_k | \tau_k^t, \mathbf{W}^t\right) \triangleq a \log \left(b \tau_k + c\right) - \beta_k \log \tau_k.$$

The gradient of this surrogate function is

$$\frac{\partial L\left(\tau_k | \tau_k^t, \mathbf{W}\right)}{\partial \tau_k} = \frac{ab}{b \tau_k + c} - \beta_k \frac{1}{\tau_k},$$

and a zero of the gradient can be solved in closed-form as

$$\tau_k^{t+1} = \frac{\beta_k c}{(a - \beta_k) b}.$$

Since

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

and

$$\beta_k = \sum_{m=1}^{R} s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1},$$

we have $a - \beta_k = R$. Therefore,

$$\tau_k^{t+1} = \frac{1}{R} \cdot \frac{\left(\sum_{m=1}^{R} s_{km} \frac{\tau_k^t \lambda_m}{\tau_k^t \lambda_m + 1}\right) \cdot \left(\sum_{m=1}^{R} \frac{\alpha_{km}}{1+\lambda_m \tau_k^t}\right)}{\sum_{m=1}^{R} \frac{\alpha_{km} \lambda_m}{1+\lambda_m \tau_k^t}} \geq 0.$$

It remains to be shown that τ_k^{t+1} is the unique minimizer of $L(\tau_k|\tau_k^t, \mathbf{W}^t)$ on $\mathbb{R}_+$.

First assume that $\beta_k \neq 0$. Since c is positive, $L(\tau_k|\tau_k^t, \mathbf{W}^t) \to +\infty$ as $\tau_k \to 0$. Besides, as $a > \beta_k$, $L(\tau_k|\tau_k^t, \mathbf{W}^t) \to +\infty$ as $\tau_k \to +\infty$. Since $L(\tau_k^{t+1}|\tau_k^t, \mathbf{W}^t)$ is finite, τ_k^{t+1} is the unique minimizer by the continuity of $L(\tau_k|\tau_k^t, \mathbf{W}^t)$. If $\beta_k = 0$, then $\tau_k^{t+1} = 0$. Function $a\log(b\tau_k + c)$ is monotonically increasing on $\mathbb{R}_+$, therefore $\tau_k^{t+1} = 0$ is the unique minimizer. Moreover, as $L(\tau_k|\tau_k^t, \mathbf{W}^t)$ is a one-dimensional function, it has to be quasi-convex [50]. ∎

APPENDIX C
PROOF OF PROPOSITION 3

Proof: By the matrix inversion lemma we have the following equality:

$$(\tau_k \boldsymbol{\Sigma} + \mathbf{I})^{-1} = \mathbf{I} - \mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \mathbf{W}^H\mathbf{W}\right)^{-1}\mathbf{W}^H$$
$$= \mathbf{I} - \mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H. \quad (36)$$

where $\tilde{\boldsymbol{\Sigma}} \triangleq \mathbf{W}^H\mathbf{W}$. Therefore, the second term of $L(\mathbf{W}|\tau_k)$ can be written as

$$\mathbf{z}_k^H(\tau_k\boldsymbol{\Sigma} + \mathbf{I})^{-1}\mathbf{z}_k$$
$$= \mathbf{z}_k^H\mathbf{z}_k - \mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k, \quad (37)$$

which is jointly concave in $\{\mathbf{W}, \tilde{\boldsymbol{\Sigma}}\}$. The differential of (37) can be computed according to (38).

$$\mathrm{d}\left(\mathbf{z}_k^H(\tau_k\boldsymbol{\Sigma} + \mathbf{I})^{-1}\mathbf{z}_k\right)$$
$$= -\mathrm{d}\left(\mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k\right)$$
$$= -\left(\mathbf{z}_k^H\mathrm{d}\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k\right.$$
$$+ \mathbf{z}_k^H\mathbf{W}\mathrm{d}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k$$
$$\left. + \mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathrm{d}\mathbf{W}^H\mathbf{z}_k\right)$$
$$= -\left(\mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k\mathbf{z}_k^H\mathrm{d}\mathbf{W}\right)\right.$$
$$+ \mathrm{Tr}\left(\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathrm{d}\mathbf{W}^H\right)$$
$$\left. -\mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathrm{d}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k\right)$$
$$= -\left(\mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k\mathbf{z}_k^H\mathrm{d}\mathbf{W}\right)\right.$$
$$+ \mathrm{Tr}\left(\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathrm{d}\mathbf{W}^H\right)$$
$$\left. -\mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathbf{W}^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\mathrm{d}\tilde{\boldsymbol{\Sigma}}\right)\right)$$
$$(38)$$

The concavity of (37) implies that it can be upperbounded by its first order Taylor expansion around the pair $\left\{\mathbf{W}^t, \tilde{\boldsymbol{\Sigma}}^t\right\}$, i.e.,

$$\mathbf{z}_k^H(\tau_k\boldsymbol{\Sigma} + \mathbf{I})^{-1}\mathbf{z}_k$$
$$\leq -\mathrm{Tr}\left(\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}^t\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}\mathbf{W}^H\right)$$
$$- \mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}(\mathbf{W}^t)^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}\right)$$
$$+ \mathrm{Tr}\left(\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}(\mathbf{W}^t)^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}^t\right.$$
$$\left. \times \left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}\tilde{\boldsymbol{\Sigma}}\right)$$
$$+ \mathrm{const.} \quad (39)$$

Substituting $\tilde{\boldsymbol{\Sigma}} = \mathbf{W}^H\mathbf{W}$ into (39) leads to the following quadratic upperbound for $\mathbf{z}_k^H(\tau_k\boldsymbol{\Sigma} + \mathbf{I})^{-1}\mathbf{z}_k$:

$$\mathrm{Tr}\left(\mathbf{W}\mathbf{H}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left((\mathbf{L}_k^t)^H\mathbf{W}\right), \quad (40)$$

where

$$\mathbf{H}_k^t = \left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}(\mathbf{W}^t)^H\mathbf{z}_k\mathbf{z}_k^H\mathbf{W}^t\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}$$
$$(41)$$

and

$$\mathbf{L}_k^t = \mathbf{z}_k\mathbf{z}_k^H\mathbf{W}^t\left(\tau_k^{-1}\mathbf{I} + \tilde{\boldsymbol{\Sigma}}^t\right)^{-1}. \quad (42)$$

As for the $\log\det(\cdot)$ term in $L(\mathbf{W}|\tau_k)$, we can apply the Sylvester's determinant theorem

$$\log\det\left(\tau_k\mathbf{W}\mathbf{W}^H + \mathbf{I}\right) = \log\det\left(\tau_k\mathbf{W}^H\mathbf{W} + \mathbf{I}\right) \quad (43)$$

and get the upperbound

$$\mathrm{Tr}\left(\tau_k\left(\tau_k\tilde{\boldsymbol{\Sigma}}^t + \mathbf{I}\right)^{-1}\mathbf{W}^H\mathbf{W}\right) + \mathrm{const.} \quad (44)$$

by linearization over $\tilde{\boldsymbol{\Sigma}}$.

Together with (40) we have an upperbound (ignoring the constant terms) for $L(\mathbf{W}|\tau_k)$ at $\mathbf{W}^t$ as

$$\mathrm{Tr}\left(\mathbf{W}\left(\sum_{k=1}^K\left(\tilde{\boldsymbol{\Sigma}}^t + \tau_k^{-1}\mathbf{I}\right)^{-1}\right)\mathbf{W}^H\right)$$
$$+ \sum_{k=1}^K\left(\mathrm{Tr}\left(\mathbf{W}\mathbf{H}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}_k^t\mathbf{W}^H\right) - \mathrm{Tr}\left((\mathbf{L}_k^t)^H\mathbf{W}\right)\right)$$
$$= \mathrm{Tr}\left(\mathbf{W}\mathbf{H}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}\mathbf{W}^H\right) - \mathrm{Tr}\left(\mathbf{L}^H\mathbf{W}\right), \quad (45)$$

where

$$\mathbf{H} = \sum_{k=1}^K\left(\left(\tilde{\boldsymbol{\Sigma}}^t + \tau_k^{-1}\mathbf{I}\right)^{-1} + \mathbf{H}_k^t\right) \quad (46)$$

and

$$\mathbf{L} = \sum_{k=1}^K\mathbf{L}_k^t. \quad (47)$$

∎

REFERENCES

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[2] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariant techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[3] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 2, pp. 208–216, 1992.

[4] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[5] S. Haykin, *Adaptive Filter Theory*, ser. Prentice-Hall information and system sciences series. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[6] I. Kirsteins and D. Tufts, "Adaptive detection using a low rank approximation to a data matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 1, pp. 55–67, 1994.

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

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

[9] P. Perry and P. Wolfe, "Minimax rank estimation for subspace tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 504–513, Jun. 2010.

[10] M. Gavish and D. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 5040–5053, 2014.

[11] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., MIT, Lexington, MA, USA, Tech. Rep., Dec. 1994.

[12] L. E. Brennan and F. Staudaher, "Subclutter visibility demonstration," Adaptive Sensors Inc., Santa Monica, CA, USA, Tech. Rep. RL-TR-92-21, Mar. 1992.

[13] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 6, pp. 853–863, Nov. 1974.

[14] A. Haimovich, "Asymptotic distribution of the conditional signal-to-noise ratio in an eigenanalysis-based adaptive array," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 3, pp. 988–997, 1997.

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

[16] J. Billingsley, "Ground clutter measurements for surface-sited radar," MIT, Cambridge, MA, USA, Tech. Rep. 780, Feb. 1993.

[17] E. Conte, A. De Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 903–918, July 2004.

[18] M. Greco, F. Gini, and M. Rangaswamy, "Statistical analysis of measured polarimetric clutter data at different range resolutions," *Proc. Inst. Electr. Eng.—Radar, Sonar, Navig.*, vol. 153, no. 6, pp. 473–481, Dec. 2006.

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

[20] M. Rangaswamy, I. Kirsteins, B. Freburger, and D. Tufts, "Signal detection in strong low rank compound-Gaussian interference," in *Proc. IEEE Sensor Array Multichannel Signal Proces. Workshop*, 2000, pp. 144–148.

[21] R. Raghavan, "Statistical interpretation of a data adaptive clutter subspace estimation algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1370–1384, 2012.

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

[23] R. A. Maronna, "Robust *M*-estimators of multivariate location and scatter," *Ann. Statist.*, vol. 4, no. 1, pp. 51–67, Jan. 1976.

[24] D. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, 1987.

[25] F. Pascal, Y. Chitour, J. Ovarlez, P. Forster, and P. Larzabal, "Existence and characterization of the covariance matrix maximum likelihood estimate in spherically invariant random processes," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[26] E. Conte, A. D. Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1908–1915, Aug. 2002.

[27] F. Gini and A. Farina, "Vector subspace detection in compound-Gaussian clutter. Part I: Survey and new results," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1295–1311, 2002.

[28] F. Gini, A. Farina, and M. Montanari, "Vector subspace detection in compound-Gaussian clutter. Part II: Performance analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1312–1323, 2002.

[29] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process., Special Section on SP With Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[30] E. Ollila and D. Tyler, "Distribution-free detection under complex elliptically symmetric clutter distribution," in *Proc. IEEE 7th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2012, pp. 413–416.

[31] O. Besson and Y. Abramovich, "Adaptive detection in elliptically distributed noise and under-sampled scenario," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1531–1535, Dec. 2014.

[32] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched condition," *Signal Process. Elsevier*, vol. 104, pp. 381–386, Nov. 2014.

[33] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, 2011.

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

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

[36] E. Ollila and D. Tyler, "Regularized *M*-estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov. 2014.

[37] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 29–38, 2012.

[38] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6182–6189, Dec. 2012.

[39] I. Soloveychik and A. Wiesel, "Group symmetric robust covariance estimation," 2014, arXiv Preprint arXiv:1412.2345.

[40] I. Soloveychik and A. Wiesel, "Tyler's covariance matrix estimator in elliptical models with convex structure," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5251–5259, Oct. 2014.

[41] Y. Sun, P. Babu, and D. Palomar, "Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions," *IEEE Trans. Signal Process.*, submitted for publication.

[42] A. Breloy, L. L. Magoarou, G. Ginolhac, F. Pascal, and P. Forster, "Maximum likelihood estimation of clutter subspace in non homogeneous noise context," presented at the EUSIPCO, Marrakech, Morocco, Sep. 2013.

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

[44] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Robust covariance matrix estimation in heterogeneous low-rank context," *IEEE Trans. Signal Process.*, submitted for publication.

[45] J. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.

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

[47] M. Steiner and K. Gerlach, "Fast converging adaptive processor or a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, 2000.

[48] A. Aubry, A. De Maio, and V. Carotenuto, "Optimality claims for the FML covariance estimator with respect to two matrix norms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 3, pp. 2055–2057, Jul. 2013.

[49] K. Lange and H. Zhou, "MM algorithms for geometric and signomial programming," *Math. Programm.*, vol. 143, no. 1–2, pp. 339–356, 2014.

[50] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
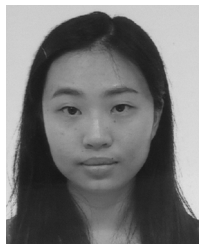
Ying Sun received the B.E. degree in electronic information from the Huazhong University of Science and Technology, Wuhan, China, in 2011. She is currently pursuing the Ph.D. degree in electronic and computer engineering at The Hong Kong University of Science and Technology.

Her research interests include statistical signal processing, optimization algorithms and machine learning.

Arnaud Breloy (S'14) graduated from Ecole Centrale Marseille and received a Master's degree of Signal and Image Processing from university of Aix-Marseille in 2012–2013. Then, he received the Ph.D. degree from University of Paris Saclay in signal processing in 2015. The thesis was done in collaboration between SATIE (ENS Cachan) and SONDRA (Centrale-Supelec) laboratories. He is currently temporary assistant professor at University Institute of Technology of Ville d'Avray and researcher at the LEME laboratory. His research interests focus on statistical signal processing, array and radar signal processing, robust estimation methods and low rank methods.

Prabhu Babu received the Ph.D. degree in electrical engineering from the Uppsala University, Sweden, in 2012. He is currently a Research Associate with The Hong Kong University of Science and Technology.

Daniel P. Palomar (S'99–M'03–SM'08–F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He is a Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. Since 2013 he is a Fellow of the Institute for Advance Study (IAS) at HKUST. He had previously held several research appointments, namely, at King's College London (KCL), London, UK; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza", Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems, big data systems, and communication systems.

Dr. Palomar is a recipient of a 2004/06 Fulbright Research Fellowship, the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society, the 2002/03 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC), the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation.

He is a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 2016 Special Issue on "Financial Signal Processing and Machine Learning for Electronic Trading" and has been Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY and of IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on "Convex Optimization for Signal Processing," the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on "Game Theory in Communication Systems," and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on "Optimization of MIMO Transceivers for Realistic Communication Networks."

Frédéric Pascal (S'04–M'06–SM'14) received the Master's degree ("Probabilities, Statistics and Applications: Signal, Image et Networks") with merit, in applied statistics from University Paris VII—Jussieu, Paris, France, in 2003. Then, he received the Ph.D. degree of Signal Processing, from University Paris X—Nanterre in 2006. From November 2006 to February 2008, he made a post-doctoral position in the Signal Processing and Information team of the laboratory SATIE, CNRS, ENS Cachan, France. Between March 2008 and December 2011 (respectively, January 2012–December 2013), he was an Assistant Professor (respectively, Associate Professor) in SONDRA at CentraleSupelec. In 2012, he obtained a Research Directorship Habilitation (HDR) thesis in Signal Processing from the University of Paris-Sud. Between August 2013 and Aug. 2014, he was a Visiting Associate Professor in the ECE department at the National University of Singapore. From January 2014, Frederic Pascal is a full Professor in the L2S laboratory at CentraleSupelec. From September 2015, Frederic Pascal is the Chair of the EURASIP SAT in Theoretical and Methodological Trends in Signal Processing (TMTSP) and he is a member of the IEEE Signal Processing Society SAM technical committee (January 2015–present). Frederic Pascal serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2015–present) and for the *EURASIP Journal on Advances in Signal Processing* (2015–present). His research interests contain estimation, detection and classification for statistical signal processing and applications in radar and image processing.

Guillaume Ginolhac (M'14) was born in France in 1974. He received the degree of electrical engineer in 1997 and a Ph.D. degree in signal processing in 2001 both from the Grenoble-INP University. From 2002 to 2012, he was associate professor of electrical engineering at Université Paris X. During this period, he was also a researcher of the Laboratory of Systems and Applications of Information and Energy Technologies (SATIE). He is currently a professor of electrical engineering at Polytech Annecy-Chambéry at the Université Savoie Mont-Blanc, France, and a member of the LISTIC Lab. His research interests are in estimation, and detection theory with applications to array processing, radar, and sonar.