

Robust Covariance Matrix Estimation in Heterogeneous Low Rank Context

A. Breloy, *Member, IEEE*, G. Ginolhac, *Member, IEEE*, F. Pascal, *Senior Member, IEEE* and P. Forster, *Member, IEEE*.

Abstract—This paper addresses the problem of robust Covariance Matrix (CM) estimation in the context of a disturbance composed of a Low Rank (LR) heterogeneous clutter plus an additive white Gaussian noise (WGN). The LR clutter is modeled by a Spherically Invariant Random Vector (SIRV) with assumed high Clutter to Noise Ratio (CNR). In such a context, adaptive process should require less training samples than classical methods to reach equivalent performance as in a "full rank" clutter configuration. The main issue is that classical robust estimators of the CM can not be computed in the undersampled configuration. To overcome this issue, the current approach is based on regularization methods. Nevertheless, most of these solutions are enforcing the estimate to be well conditioned, while in our context, it should be LR structured. This paper therefore addresses this issue and derives an algorithm to compute the Maximum Likelihood Estimator (MLE) of the CM for the considered disturbance model. Several relaxations and robust generalizations of the result are discussed. Performance is finally illustrated on numerical simulations and on a Space Time Adaptive Processing (STAP) for airborne radar application.

Index Terms—Adaptive Signal Processing, Covariance Matrix and Subspace estimation, Robust estimation, ML-estimation, Low Rank, SIRV, STAP.

I. INTRODUCTION

COVARIANCE matrix (CM) estimation is a fundamental issue in adaptive signal processing [1–4]. In terms of application purposes, the accuracy of the CM estimate directly impacts the performance of algorithms and systems. The most common estimator of the CM is the traditional Sample Covariance Matrix (SCM), which is the Maximum Likelihood Estimator (MLE) of the CM in a Gaussian context. Nevertheless, when the samples are heavy-tailed distributed, and possibly corrupted by outliers, the SCM fails to provide an accurate estimate of the CM at finite sample support.

To overcome this issue, the robust estimation framework [5, 6] has recently attracted considerable interest [7] (and reference therein). Most of current work in CM estimation focuses on the family of Complex Elliptical Symmetric (CES) distributions [7] for modeling heterogeneous noises. Among the general CES class, this paper will focus on the SIRV's [8]

also referred to as Compound Gaussian (CG), which cover a large panel of well known distributions, notably heavy-tailed such as Weibull, Student's t -distribution, and K -distribution. Moreover, SIRV presents good agreement to several real data sets [9–12]. Under this framework, a robust estimation of the CM can be performed using the M -estimators [13–15]. A detailed review of the CES model, robust CM estimators and their properties is provided in [7]. This general framework has been extensively used in the modern detection/estimation literature [16–22] due to its interest, both from a theoretical and an application point of view.

In addition to the robustness issue, the problem of high dimensional data with low sample support occurs. Typically, when number of samples K is less than the dimension of the data M , the robust estimators are not defined anymore. To solve this problem, the current approach is based on regularization methods, such as shrinkage algorithms [23–29]. Another approach consists in introducing prior information on the CM structure through appropriate parameterization [30, 31] and eventual penalization [32] or relaxations [33] of the considered functional. In this paper, we will follow the path of introducing prior knowledge on the structure and consider the case of Low Rank (LR) structured CM. This work differs from previously cited works by two essential points. First, LR structured matrices do not define an affine (nor convex) set, so the addressed problem does not fall back on specific cases of previous references. Second and most importantly, the approach proposed in this paper is not derived from regularization/penalization of the robust estimators. Rather, the proposed CM estimator will be directly derived from a realistic model that accounts for heterogeneity of the noise.

Indeed, in some applications the heterogeneous disturbance (referred to as "clutter") lies in a lower dimensional subspace, therefore it has a LR CM of rank $R < M$. Taking into account the additive White Gaussian Noise (WGN), one can eventually model the samples as a LR SIRV (accounting for heterogeneity of the clutter) plus a WGN. This general model, already used in [34–39], leads to a total covariance matrix of form: LR plus scaled identity. As major hypothesis, we will assume that the rank R is known. The prior information on this rank can be obtained in some applications, for example, thanks to geometry of the considered system (as in STAP [1, 40]) or assuming prior information on the number of jammers (for denoising problem) or of sources (as in MUSIC-type applications [41]). Otherwise we will assume that it has been formerly estimated by a method from the state of the art (see e.g. [42] or more recent results based on random matrix theory

Arnaud Breloy is with LEME, University Paris West, F-92410 Ville d'Avray, France (e-mail: abreloy@u-paris10.fr). Guillaume Ginolhac is with LISTIC, Université de Savoie Mont-Blanc, F-74944 Annecy le Vieux Cedex, France (e-mail: guillaume.ginolhac@univ-smb.fr). Frédéric Pascal is with L2S, Centrale-Supélec, F-91192 Gif-sur-Yvette Cedex, France (e-mail: frederic.pascal@centralesupelec.fr). P. Forster is with SATIE, ENS Cachan, CNRS, F-94230 Cachan, France (e-mail: pforster@u-paris10.fr). Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

[43] or considering starved sample situation as [44, 45] and the references therein) and that we use the estimated value as the actual one.

Seminal work considering parameters estimation of the LR-SIRV plus WGN model has been proposed in [36, 46]. These previous works are focused on the Clutter Subspace Projector (CSP) (*i.e.* the subspace where the LR-SIRV clutter lies in) estimation and assume a clutter CM with identical eigenvalues, hence do not propose an estimate of the total CM. A relaxation over the hypothesis of identical eigenvalues of the clutter CM has been proposed in [39]. However, the estimator proposed in [39] suffers from a lack of tractability and resorts to an ad-hoc regularization of some side parameters. This may lead to a poor estimation of the total CM (as demonstrated in this paper), even if the CSP is well estimated. To solve this issue [47] proposed exact resolution algorithms based on Majorization-Minimization framework. Nevertheless these contributions focused only on subspace estimation and did not considered the problem of total covariance matrix recovery, *i.e.* neglecting the estimation accuracy of side parameters. Since the considered likelihood is a complex non-convex function, all the algorithms from [39, 47] inherently reach (or approach) a local maximum of the likelihood, so this previous state of the art may not be relevant from the point of view of CM estimation. This paper fills therefore the gaps by tackling the CM estimation problem and proposes new estimation process that are relevant to the purpose, using the assumption of high Clutter to Noise Ratio (CNR), which is valid and useful for most radar applications. Hence we propose a general algorithm to compute robust estimates of the CM according to the considered model (LR-SIRV plus WGN). This algorithm is developed according to the "2-Step MLE" approach, that relies on alternatively maximizing the likelihood over two parameters: the CSP and the LR clutter CM over this subspace. In addition, we also consider several modifications and adaptations of the proposed algorithm. Firstly, since the clutter Probability Density Function (PDF) is in practice unknown, we consider the use of robust estimators to address this issue. Secondly, we discuss several relaxations of the CSP MLE problem and their use in terms of computation. The proposed methodology offers eventually generalizations of previous results since it encompasses [36, 46].

This paper is organized as follows. Section II presents the statistical model, previous results and states the considered problematic. Then, we derive the proposed 2-Step MLE algorithm in section III. Section IV considers the use of robust estimators in this 2-step procedure. Moreover, we discuss several relaxations of the CSP MLE problem and their use in terms of computation in section V. Eventually, the proposed methodology offers a very general solution to compute LR structured robust CM estimators. In section VII, the proposed algorithms are compared with the state of art through validation and a real data STAP application. Section VIII draws some conclusions of this study.

Notations: The following convention is 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 is the transpose operator. $\mathcal{CN}(\mathbf{a}, \mathbf{\Sigma})$ is a complex Gaussian vector of mean \mathbf{a} and of covariance matrix $\mathbf{\Sigma}$. \mathbf{I}_M is the $M \times M$ identity matrix. $|\mathbf{\Sigma}|$ is the determinant of the matrix $\mathbf{\Sigma}$ and $\text{Tr}(\cdot)$ is the Trace operator. \hat{d} is an estimate of the parameter d . $\{w_n\}_{n \in \llbracket 1, N \rrbracket}$ denotes the set of n elements w_n with $n \in \llbracket 1, N \rrbracket$ and whose writing will often be contracted into $\{w_n\}$. $\text{diag}(a_n)$ is the $N \times N$ diagonal matrix with diagonal elements a_n . \mathfrak{U} is the concatenation operator, where $\mathbf{V} = \mathfrak{U}(\{\mathbf{v}_n\})$ is the matrix whose columns are the vectors \mathbf{v}_n for $n \in \llbracket 1, N \rrbracket$.

II. MODEL AND MAXIMUM LIKELIHOOD ESTIMATOR

A. Model

We assume that K secondary data $\{\mathbf{z}_k\}_{k \in \llbracket 1, K \rrbracket}$ are available. Each of these data $\mathbf{z}_k \in \mathbb{C}^M$ corresponds to a realization of a proper circular LR-SIRV process \mathbf{c}_k plus an independent additive zero-mean complex WGN \mathbf{n}_k .

$$\mathbf{z}_k = \mathbf{n}_k + \mathbf{c}_k . \quad (1)$$

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

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

where the power of the WGN σ^2 is assumed to be known¹. The LR-SIRV [7, 8] \mathbf{c}_k is a M -dimensional zero-mean complex Gaussian vector (the *speckle*) with CM $\mathbf{\Sigma}_c$, multiplied by the square root of a positive random power factor (the *texture*) τ_k of PDF f_τ . For all the paper, we assume that $\mathbb{E}(\tau) < \infty$ so the CM of the considered SIRV exists. Each realization \mathbf{c}_k follows then, conditionally to τ_k :

$$(\mathbf{c}_k | \tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{\Sigma}_c) , \quad (3)$$

where the LR-SIRV clutter CM $\mathbf{\Sigma}_c$ is defined by its rank $R < M$ (assumed to be known), its unknown orthonormal eigenvectors $\{\mathbf{v}_r\}$ and unknown corresponding eigenvalues $\{c_r\}$ for $r \in \llbracket 1, R \rrbracket$:

$$\mathbf{\Sigma}_c = \sum_{r=1}^R c_r \mathbf{v}_r \mathbf{v}_r^H = \mathbf{V}_c \mathbf{C}_c \mathbf{V}_c^H , \quad (4)$$

with $\mathbf{C}_c = \text{diag}(c_r)_{r \in \llbracket 1, R \rrbracket}$ the $R \times R$ diagonal matrix of the eigenvalues and $\mathbf{V}_c = \mathfrak{U}(\{\mathbf{v}_r\})$ the $M \times R$ concatenation of the orthonormal eigenvectors \mathbf{v}_r . We also define the Clutter Subspace orthogonal Projector (CSP):

$$\mathbf{\Pi}_c = \sum_{r=1}^R \mathbf{v}_r \mathbf{v}_r^H = \mathbf{V}_c \mathbf{V}_c^H , \quad (5)$$

and its complementary subspace projector $\mathbf{\Pi}_c^\perp = \mathbf{I}_M - \mathbf{\Pi}_c$. The orthonormal completion of the basis \mathbf{V}_c is a set of $M - R$ vectors $\{\mathbf{v}_r\}_{r \in \llbracket R+1, M \rrbracket}$, concatenated in the matrix denoted \mathbf{V}_c^\perp and verifies $\mathbf{\Pi}_c^\perp = \mathbf{V}_c^\perp \mathbf{V}_c^{\perp H}$.

Considering the presented model, each data \mathbf{z}_k can be described, conditionally to τ_k , by:

$$(\mathbf{z}_k | \tau_k) \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k) , \quad (6)$$

¹This hypothesis is made for describing a valid theoretical framework. In practice, presented results could be applied with an estimate of σ^2 used as its actual value, see e.g. [48].

with:

$$\boldsymbol{\Sigma}_k = \tau_k \boldsymbol{\Sigma}_c + \sigma^2 \mathbf{I}_M . \quad (7)$$

The likelihood of the data set is defined using f_τ , the PDF of τ :

$$f(\{\mathbf{z}_k\}|\boldsymbol{\Sigma}_c) = \prod_{k=1}^K \int_0^{+\infty} \frac{e^{-\mathbf{z}_k^H \boldsymbol{\Sigma}_c^{-1} \mathbf{z}_k}}{\pi^M |\boldsymbol{\Sigma}_k|} f_\tau(\tau_k) d\tau_k , \quad (8)$$

and the total CM of the LR-SIRV plus WGN is:

$$\boldsymbol{\Sigma}_{tot} = \mathbb{E}(\boldsymbol{\Sigma}_k) = \mathbb{E}(\tau) \boldsymbol{\Sigma}_c + \sigma^2 \mathbf{I}_M , \quad (9)$$

that does not depend on k since the τ_k are i.i.d. according to f_τ , i.e. $\mathbb{E}(\tau_k) = \mathbb{E}(\tau) \forall k$. We will initially consider that no information about the PDF f_τ is available, except that $\mathbb{E}(\tau)$ exists and is finite. We therefore treat each τ_k as unknown deterministic variables. The likelihood is then:

$$f(\{\mathbf{z}_k\}|\boldsymbol{\Sigma}_c, \{\tau_k\}) = \prod_{k=1}^K \frac{e^{-\mathbf{z}_k^H \boldsymbol{\Sigma}_k^{-1} \mathbf{z}_k}}{\pi^M |\boldsymbol{\Sigma}_k|} . \quad (10)$$

B. MLE and addressed problem

Following the derivation made in [39], one can express the MLE of the CM parameters as the solution of the problem:

$$\begin{aligned} \underset{\{\mathbf{v}_r\}}{\text{maximize}} \quad & f_0(\{\mathbf{v}_r\}) = \sum_{r=1}^R \mathbf{v}_r^H \hat{\mathbf{M}}_r \mathbf{v}_r , \\ \text{subject to} \quad & \mathbf{v}_r^H \mathbf{v}_r = 1 , \quad r \in \llbracket 1, R \rrbracket , \\ & \mathbf{v}_r^H \mathbf{v}_j = 0 , \quad r, j \in \llbracket 1, R \rrbracket , \quad r \neq j . \end{aligned}$$

where the matrices $\hat{\mathbf{M}}_r$ are defined by:

$$\hat{\mathbf{M}}_r = \sum_{k=1}^K \frac{\hat{c}_r \hat{\tau}_k}{\sigma^2 + \hat{c}_r \hat{\tau}_k} \mathbf{z}_k \mathbf{z}_k^H , \quad (11)$$

with \hat{c}_r and $\hat{\tau}_k$ the MLE of respectively c_r and τ_k .

Nevertheless, this formulation does not directly provide a solution because MLE of the parameters c_r and τ_k are not tractable. Moreover, they are depending on parameters $\{\mathbf{v}_r\}$ which implies a need for iterative algorithms to reach the MLE. The proposed solution in [39] is to derive the MLE of side parameters $d_r^k = c_r \tau_k$ then apply an ad-hoc regularization over those estimates to recover estimators of c_r and τ_k . This eventually leads to an accurate estimator of the CSP $\boldsymbol{\Pi}_c$. However, this solution is not reliable when one wants to reconstruct the total CM (as demonstrated in the simulation section of this paper).

The contributions of this paper are therefore the following: in Section III, we derive a new algorithm using the high CNR assumption to efficiently address this issue. Moreover, the considered framework allows several generalizations and adaptations that will be developed in Section IV and V. The overall estimation algorithm, accounting for possible variations, is summed up in the box "Algorithm 1".

III. ML ESTIMATION OF THE CLUTTER CM

The main issue to compute the MLE of the CM is that the parameters \mathbf{C}_c and \mathbf{V}_c can only be estimated conditionally to each other. We propose therefore to use a general 2-step MLE algorithm that alternatively maximizes the likelihood over the sets of parameters $\{c_r\}, \{\tau_k\}$ and $\{\mathbf{v}_r\}$. This algorithm ensures convergence to a local maximum. Following sections will derive the two ML estimation procedure for each step.

A. Step-1 : Estimation of the core matrix \mathbf{C}_c and textures τ_k for fixed clutter subspace

From now on, the CNR will be considered high, which is a realistic assumption for most applications. High CNR means that the WGN is negligible compared to the SIRV noise over the clutter subspace. Otherwise stated, we consider that $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_k)$ with:

$$\boldsymbol{\Sigma}_k \simeq \tau_k \boldsymbol{\Sigma}_c + \sigma^2 \boldsymbol{\Pi}_c^\perp . \quad (12)$$

We will assume that an arbitrary orthonormal basis of the clutter subspace \mathbf{U}_c is given from the previous step, that is $\boldsymbol{\Pi}_c = \mathbf{U}_c \mathbf{U}_c^H$. Let \mathbf{U}_c^\perp be the completion of this basis, so that $\boldsymbol{\Pi}_c^\perp = \mathbf{U}_c^\perp \mathbf{U}_c^{\perp H}$. The CM of the data \mathbf{z}_k , conditionally to τ_k , is expressed with $\mathbf{U} = [\mathbf{U}_c \ \mathbf{U}_c^\perp]$ as:

$$\boldsymbol{\Sigma}_k \simeq \mathbf{U} \begin{pmatrix} \tau_k \mathbf{M}_c & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{M-R} \end{pmatrix} \mathbf{U}^H . \quad (13)$$

Remark 1: Note that we consider an arbitrary basis \mathbf{U}_c rather than the eigenvectors of the clutter CM \mathbf{V}_c . This is done since the estimation of the CSP $\boldsymbol{\Pi}_c$ in step 2 does not provide the exact eigenvectors basis but only a basis that spans the desired subspace. Here, we assume that \mathbf{U}_c spans the clutter subspace so it is equal to \mathbf{V}_c up to a rotation:

$$\mathbf{U}_c = \mathbf{V}_c \mathbf{Q} , \quad (14)$$

with \mathbf{Q} a $R \times R$ unitary matrix. Therefore, the core matrix \mathbf{M}_c is not necessarily diagonal.

One has then the following result:

Theorem 1 For a given basis \mathbf{U}_c the MLE of the matrix \mathbf{M}_c is defined as the unique fixed point:

$$\hat{\mathbf{M}}_c = \frac{R}{K} \sum_{k=1}^K \frac{\mathbf{U}_c^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}_c}{\mathbf{z}_k^H \mathbf{U}_c \hat{\mathbf{M}}_c^{-1} \mathbf{U}_c^H \mathbf{z}_k} , \quad (15)$$

that can be computed through the following iterations :

$$\mathbf{M}_{(n+1)} = \frac{R}{K} \sum_{k=1}^K \frac{\mathbf{U}_c^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}_c}{\mathbf{z}_k^H \mathbf{U}_c \mathbf{M}_{(n)}^{-1} \mathbf{U}_c^H \mathbf{z}_k} , \quad (16)$$

that converges to $\hat{\mathbf{M}}_c$ under the condition that $K > R$ and that there is no $\mathbf{U}_c^H \mathbf{z}_k$ that lies on the origin.

Proof: Consider the independent and dimension reduced variables \mathbf{z}_c and \mathbf{z}_g

$$\tilde{\mathbf{z}} = \mathbf{V}^H \mathbf{z} = \begin{bmatrix} \mathbf{V}_c^H \mathbf{z} \\ \mathbf{V}_c^{\perp H} \mathbf{z} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_c \\ \mathbf{z}_g \end{bmatrix} . \quad (17)$$

Each one has the distribution:

$$\begin{cases} (\mathbf{z}_c | \tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{C}_c), \\ \mathbf{z}_g \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{M-R}), \end{cases} \quad (18)$$

The result directly follows from the fact that $\mathbf{U}_c^H \mathbf{z}$ is the vector \mathbf{z}_c projected in a basis that spans the same subspace as \mathbf{V}_c . Hence $\mathbf{U}_c^H \mathbf{z}$ is a R -dimensional random vector distributed as a SIRV with unknown texture PDF and its CM MLE is well known from the CES theory [7, 15]. ■

The MLE of the core CM corresponds therefore to the FPE [14, 15] applied on the projected data. This estimation procedure also provides the estimates of the unknown deterministic textures by:

$$\hat{\tau}_k = \frac{\mathbf{z}_k^H \mathbf{U}_c \hat{\mathbf{M}}_c^{-1} \mathbf{U}_c^H \mathbf{z}_k}{R}. \quad (19)$$

For a given clutter subspace basis, the LR-SIRV CM is then obtained by:

$$\left(\hat{\Sigma}_c | \mathbf{U}_c \right) = \mathbf{U}_c \hat{\mathbf{M}}_c \mathbf{U}_c^H = \hat{\mathbf{V}}_c \hat{\mathbf{C}}_c \hat{\mathbf{V}}_c^H. \quad (20)$$

The parameters $\left(\{ \hat{\mathbf{V}}_c, \hat{\mathbf{C}}_c \} | \mathbf{U}_c \right)$ are then obtained through the SVD of this matrix.

Remark 2: To reconstruct the total CM, one has to remove a scaling ambiguity, *i.e.* estimate the CNR. The full rank CM of the LR-SIRV plus WGN is Σ_{tot} defined in Eq.(9). In the case where f_τ is known, the MLE of this matrix is obtained as:

$$\hat{\Sigma} = \mathbb{E}(\tau) \hat{\Sigma}_c + \sigma^2 (\mathbf{I}_M - \hat{\Pi}_c). \quad (21)$$

However, if the textures are considered as unknown deterministic parameters the MLE of this matrix is evaluated through:

$$\hat{\Sigma} = \hat{\mathbb{E}}(\tau) \hat{\Sigma}_c + \sigma^2 (\mathbf{I}_M - \hat{\Pi}_c), \quad (22)$$

where $\hat{\mathbb{E}}(\tau)$ is the empirical mean of the textures estimates:

$$\hat{\mathbb{E}}(\tau) = \frac{1}{K} \sum_{k=1}^K \hat{\tau}_k. \quad (23)$$

Step-1 summed up: The first step is the ML estimation of the core matrix $\hat{\mathbf{C}}_c^{(n)}$ and the textures $\{ \hat{\tau}_k \}^{(n)}$. For a given orthonormal clutter subspace basis $\hat{\mathbf{U}}_c^{(n)}$, the iterations (16) provide the updated LR clutter CM $\hat{\Sigma}_c^{(n)}$ and the updated textures $\{ \hat{\tau}_k \}^{(n)}$ with Eq.(19). We denote:

$$\hat{\mathbf{U}}_c^{(n)} \xrightarrow{\text{MLE Step-1}} \begin{cases} \hat{\Sigma}_c^{(n)} = \mathbf{U}_c^{(n)} \hat{\mathbf{M}}_c^{(n)} \mathbf{U}_c^{(n)H} \\ \quad \quad \quad = \hat{\mathbf{V}}_c^{(n)} \hat{\mathbf{C}}_c^{(n)} \hat{\mathbf{V}}_c^{(n)H} \\ \{ \hat{\tau}_k \}^{(n)} \end{cases}$$

B. Step-2: Clutter subspace basis \mathbf{V}_c estimation

We assume that the set of texture $\{ \tau_k \}$ and the core matrix \mathbf{C}_c are known and fixed from the previous step. The log of the likelihood function Eq.(10) is expressed:

$$\log f(\{ \mathbf{z}_k \} | \mathbf{V}_c) = - \sum_{k=1}^K \mathbf{z}_k^H \Sigma_k^{-1} \mathbf{z}_k + \gamma, \quad (24)$$

where γ is a constant not depending on $\{ \mathbf{v}_r \}$, since the determinants terms $|\Sigma_k|$ are only depending on \mathbf{C}_c . Expanding this expression, taking the high CNR assumption in Eq.(12) into account, gives:

$$\begin{aligned} \log f &= - \sum_{k=1}^K \frac{1}{\tau_k} \mathbf{z}_k^H \mathbf{V}_c \mathbf{C}_c^{-1} \mathbf{V}_c^H \mathbf{z}_k \\ &\quad - \sum_{k=1}^K \frac{1}{\sigma^2} \mathbf{z}_k^H \mathbf{V}_c^\perp \mathbf{V}_c^{\perp H} \mathbf{z}_k + \gamma. \end{aligned} \quad (25)$$

To inherently constrain orthonormality between \mathbf{V}_c and \mathbf{V}_c^\perp , one use the following change of variables: $\mathbf{V}_c^\perp \mathbf{V}_c^{\perp H} = \mathbf{I}_m - \mathbf{V}_c \mathbf{V}_c^H$, that leads to:

$$\begin{aligned} \log f &= - \sum_{k=1}^K \frac{1}{\tau_k} \mathbf{z}_k^H \mathbf{V}_c \mathbf{C}_c^{-1} \mathbf{V}_c^H \mathbf{z}_k \\ &\quad - \sum_{k=1}^K \frac{1}{\sigma^2} \mathbf{z}_k^H \mathbf{I}_m \mathbf{z}_k + \sum_{k=1}^K \frac{1}{\sigma^2} \mathbf{z}_k^H \mathbf{V}_c \mathbf{V}_c^H \mathbf{z}_k + \gamma. \end{aligned} \quad (26)$$

If this constraint is not taken into account, one obtains a different relaxed estimator, as derived in [49]. The previous expression, once developed, gives:

$$\log f = \sum_{k=1}^K \sum_{r=1}^R \left(\frac{1}{\sigma^2} - \frac{1}{c_r \tau_k} \right) \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k + \gamma, \quad (27)$$

where γ has absorbed the constant $\sum_{k=1}^K \|\mathbf{z}_k\|^2 / \sigma^2$ which does not depend on $\{ \mathbf{v}_r \}$. Defining the matrices \mathbf{M}_r as

$$\mathbf{M}_r = \sum_{k=1}^K \left(\frac{1}{\sigma^2} - \frac{1}{c_r \tau_k} \right) \mathbf{z}_k \mathbf{z}_k^H. \quad (28)$$

The log-likelihood is finally:

$$\log f = \sum_{r=1}^R \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r + \gamma, \quad (29)$$

which is to be maximized w.r.t. the set $\{ \mathbf{v}_r \}$ under the unitary constraints $\mathbf{V}_c^H \mathbf{V}_c = \mathbf{I}_R$. The MLE of the CSP basis is therefore the solution of the following constrained optimization problem:

$$\begin{aligned} \underset{\{ \mathbf{v}_r \}}{\text{maximize}} \quad & f_0(\mathbf{V}) = f_0(\{ \mathbf{v}_r \}) = \sum_{r=1}^R \mathbf{v}_r^H \mathbf{M}_r \mathbf{v}_r, \\ \text{subject to} \quad & \mathbf{v}_r^H \mathbf{v}_r = 1, \quad r \in \llbracket 1, R \rrbracket, \\ & \mathbf{v}_r^H \mathbf{v}_j = 0, \quad r, j \in \llbracket 1, R \rrbracket, \quad r \neq j. \end{aligned} \quad (\text{A})$$

To the best of our knowledge, there is no closed form solution for this problem. Therefore, one has to resort to iterative algorithms to reach local maxima of f_0 , (or appropriate relaxations, as discussed Section V). To compute the solution, we will apply algorithm 15 from [50]: "Modified Steepest descent on Stiefel manifold". We also refer to other possible algorithms [51–53]. This algorithm ensures to increase the likelihood over the Stiefel manifold, so each update is satisfying the unitary constraints. We note that there is no uniqueness of the solution and that the algorithm could be trapped in local extrema. However, one could expect that the

previous estimation step provides a starting point near to the global optimum.

Remark 3: We derive the MLE of the clutter CM eigenvectors $\hat{\mathbf{V}}_c$ that span the MLE of the subspace of interest $\hat{\mathbf{\Pi}}_c = \hat{\mathbf{V}}_c \hat{\mathbf{V}}_c^H$. As discussed before, the Step-1 only requires to have the MLE of the clutter subspace, *i.e.* $\hat{\mathbf{\Pi}}_c$ or an arbitrary basis $\hat{\mathbf{U}}_c$ that spans this same subspace (which is equal to $\hat{\mathbf{V}}_c$ up to a rotation). However, we point out that such a relaxed MLE is not derivable in practice since the objective function is not rotation invariant: $f_0(\mathbf{V}) \neq f_0(\mathbf{V}\mathbf{Q})$ for an arbitrary rotation \mathbf{Q} . Section V discusses several relaxations of the problem so that an arbitrary basis $\hat{\mathbf{U}}_c$ may be obtained as a solution.

Remark 4: When it comes to the clutter subspace estimation, one may notice that the involved matrices $\hat{\mathbf{M}}_r$ are SCMs of the data scaled by estimated factors α_r^k that are:

$$\alpha_r^k = \left(\frac{1}{\sigma^2} - \frac{1}{c_r \tau_k} \right) \simeq \frac{1}{\sigma^2} \left(\frac{\tau_k c_r}{\sigma^2 + \tau_k c_r} \right) \propto \frac{\omega_r^k}{1 + \omega_r^k},$$

where $\omega_r^k = c_r \tau_k / \sigma^2$. The α_r^k factors are thus corresponding to the "power proportion" of the clutter over the rank one subspace defined by $\mathbf{v}_r \mathbf{v}_r^H$. Same type of factor appears in [36, 39, 46] but their MLE expression differs due to the different model hypothesis. Yet, the CSP estimation strategy is similar: $\hat{\alpha}_r^k$ will increase as the texture increases, which means that realizations \mathbf{z}_k that contain more power in the subspace of interest are given more significance in its estimation process. Note that on the contrary, robust estimates of the CM are scaling the data by factors inversely proportional to the texture. The CSP MLE approach can be seen as robust from the point of view of estimating the CSP (not the CM) where the "outliers" are data without clutter. The approach may be useful in some specific cases, where the clutter is not well represented over the secondary data set [36, 54]. This scenario also corresponds to a low/average CNR and/or very impulsive clutter.

Step-2 summed up: The second step is the ML estimation of the clutter subspace basis. $\hat{\mathbf{C}}_c^{(n)}$ and $\{\hat{\tau}_k^{(n)}\}$ are used to compute the set of matrices $\{\hat{\mathbf{M}}_r^{(n)}\}$. Solving the constrained optimization (A) gives the updated clutter subspace basis $\hat{\mathbf{U}}_c^{(n+1)}$. We denote:

$$\hat{\mathbf{C}}_c^{(n)}, \{\hat{\tau}_k^{(n)}\} \xrightarrow{\text{MLE Step 2}} \hat{\mathbf{U}}_c^{(n+1)}.$$

IV. GENERALIZATION TO M -ESTIMATORS

It is worth replacing the core CM \mathbf{C}_c estimation in the broader context of M -estimation. We initially assumed that no prior information was available on the texture PDF f_τ . If f_τ is known, the MLE takes the form:

$$\hat{\mathbf{M}}_c = \frac{1}{K} \sum_{k=1}^K \psi(\mathbf{z}_k^H \mathbf{U}_c \hat{\mathbf{M}}_c^{-1} \mathbf{U}_c^H \mathbf{z}_k) \mathbf{U}_c^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}_c, \quad (30)$$

with $\psi(t) = -h'_r(t)/h_r(t)$ (where ' stands for the derivative *w.r.t.* t) and

$$h_r(t) = \int_0^{+\infty} \exp(-t/\tau) \tau^R f_\tau(\tau) d\tau. \quad (31)$$

This solution is known to exist and to be unique for $K > R$ if ψ satisfies some conditions stated in Theorem 6 of [7]. It can be evaluated with the following fixed point equations:

$$\hat{\mathbf{M}}_{(n+1)} = \frac{1}{K} \sum_{k=1}^K \psi(\mathbf{z}_k^H \mathbf{U}_c \hat{\mathbf{M}}_{(n)}^{-1} \mathbf{U}_c^H \mathbf{z}_k) \mathbf{U}_c^H \mathbf{z}_k \mathbf{z}_k^H \mathbf{U}_c, \quad (32)$$

that are known to converge to the solution $\hat{\mathbf{M}}_c$. The MLE of the textures is also obtained with:

$$\hat{\tau}_k = \psi(\mathbf{z}_k^H \mathbf{U}_c \hat{\mathbf{M}}_c^{-1} \mathbf{U}_c^H \mathbf{z}_k)^{-1}. \quad (33)$$

In practice, the pdf of the texture f_τ is unknown. In that case one can resort to robust M -estimation [5–7] and use a functional ψ that is not necessarily related to the texture pdf f_τ . For example the FPE is a special case where $\psi(t) = R/t$. M -estimators can therefore be seen as generalized MLE for the CES class. The choice of the functional can be either linked to some prior on the texture pdf, either designed to ensure a compromise between performance and robustness to various distributions. Generalization of the proposed 2-Step algorithm to M -estimators can therefore be done by replacing iterations Eq.(16) by Eq.(32) and the texture estimation Eq.(19) by Eq.(33).

V. DERIVING FAST APPROACHED CSP MLE

To compute the MLE of the CSP basis \mathbf{V}_c , one has to solve Problem (A). Since there is no closed form solution, one has to resort to iterative algorithms to reach the solution, which can be computationally costly. This section aims at proposing two relaxations of the problem (A) to obtain approached CSP MLE that are easily computable.

In the special case where all the matrices \mathbf{M}_r are equals, *i.e.* $\mathbf{M}_r = \mathbf{M} \forall r$, the ML estimation problem (A) can be equivalently reformulated:

$$\begin{aligned} & \underset{\{\mathbf{V}_c\}}{\text{maximize}} && \text{Tr}\{\mathbf{V}_c^H \mathbf{M} \mathbf{V}_c\} \\ & \text{subject to} && \mathbf{V}_c^H \mathbf{V}_c = \mathbf{I}_R \end{aligned}, \quad (B)$$

which has a well known solution that is $\hat{\mathbf{V}}_c$ defined as the R strongest eigenvectors of the matrix \mathbf{M} , or any basis $\hat{\mathbf{U}}_c$ that spans the same subspace. This solution is obtainable through the SVD of the matrix \mathbf{M} .

1) 0 order approximation of the CSP MLE:

Under the high CNR assumption (12) the matrices \mathbf{M}_r can be approached by:

$$\hat{\mathbf{M}}_r = \sum_{k=1}^K \left(\frac{1}{\sigma^2} - \frac{1}{c_r \tau_k} \right) \mathbf{z}_k \mathbf{z}_k^H \simeq \left(\frac{1}{\sigma^2} \right) \sum_{k=1}^K \mathbf{z}_k \mathbf{z}_k^H, \quad (34)$$

so they are all approximately equal to:

$$\hat{\mathbf{M}}_r \simeq \frac{K}{\sigma^2} \hat{\mathbf{M}}_{SCM} \quad \forall r \in [1, R], \quad (35)$$

where \mathbf{M}_{SCM} is the Sample Covariance Matrix:

$$\hat{\mathbf{M}}_{SCM} = \frac{1}{K} \sum_{k=1}^K \mathbf{z}_k \mathbf{z}_k^H. \quad (36)$$

The ML estimation problem of the CSP basis follows then the expression (B). The solution to find the 0 order approximation of the clutter subspace basis MLE can therefore be found as $\hat{\mathbf{V}}_c$ defined as the R strongest eigenvectors of the SCM.

2) Average Relaxation of the CSP MLE or A-MLE:

The proposed relaxation consists in averaging the factors $c_r \tau_k$ in the expression of the matrices \mathbf{M}_r . This relaxation can also have a ML interpretation by considering the average power over the clutter subspace as parameter of interest, as done in [36], or simply under the hypothesis of equal eigenvalues of the clutter CM, as done in [46]. In that case, one can relax the problem into form 2 without ignoring the possible impact of the textures (as done with the SCM in the 0 order approximation). Let us define the averaged textures:

$$\tilde{\tau}_k = \frac{1}{R} \sum_{r=1}^R c_r \tau_k, \quad (37)$$

and corresponding averaged power proportion factors:

$$\tilde{\alpha}_k = \frac{\tilde{\tau}_k}{\sigma^2 + \tilde{\tau}_k}. \quad (38)$$

With these definitions, one can then approach the matrices \mathbf{M}_r by:

$$\mathbf{M}_r \simeq \tilde{\mathbf{M}} = \frac{1}{\sigma^2} \sum_{k=1}^K \tilde{\alpha}_k \mathbf{z}_k \mathbf{z}_k^H \quad \forall r \in \llbracket 1, R \rrbracket. \quad (39)$$

Since $\tilde{\mathbf{M}}$ does not depend of the index r , the MLE of the CSP basis can be relaxed in the form of (B) as done previously. The CSP Approached-MLE (A-MLE) basis is then $\hat{\mathbf{V}}_c$ defined as the R stronger eigenvectors of the matrix $\tilde{\mathbf{M}}$.

VI. CONVERGENCE ANALYSIS

In this section, we discuss the convergence properties of the three algorithms presented in previous sections.

MLE: For a number of samples $K > R + 1$, Step-1 reaches the unique existing maximum (up to a non affecting scaling factor) of the objective function for fixed parameter \mathbf{U}_c , as guaranteed by [15]. The convergence of Step-2 (updating \mathbf{U}_c for other fixed parameters) to a local maximum is ensured by [50]. Hence, the sequence of objective values generated by MLE algorithm will converge because of monotonicity. However, the convergence of the algorithm in terms of variables remains unknown because of the non convexity of the constraints involved in Step-2. To the best of our knowledge, there is no convergence result of general block descent type algorithms with a non-convex constraint set (e.g. generalization of [55]), and deriving such a specific proof is beyond the scope of this paper. Nevertheless, we note that convergence in terms of variables is also observed in practice.

Algorithm 1 Robust CM estimation in LR clutter context

- 1: Initialize \mathbf{V}_c and \mathbf{U}_c with the R strongest eigenvectors of the SCM
- 2: **for** $n = 0 \rightarrow N$ **do**
- 3: * Step-1 for fixed \mathbf{U}_c :
- 4: Compute \mathbf{M}_c with iterations (16) (or (32) if ψ known)
- 5: $\hat{\Sigma}_c = \mathbf{U}_c \mathbf{M}_c \mathbf{U}_c^H \xrightarrow{\text{SVD}} \mathbf{C}_c = \text{diag}(\{c_r\})$ and \mathbf{V}_c
- 6: Update $\{\tau_k\}$ with (19) (or (33) if ψ known)
- 7: * Step-2 for fixed $\{\tau_k\}$ and $\{c_r\}$
- 8: Switch case on algorithm (0-MLE, A-MLE or MLE):
- 9: * 0-MLE (0 order approximation):
- 10: **end algorithm**
- 11: * A-MLE (Average relaxation):
- 12: Compute $\tilde{\mathbf{M}}$ with (37), (38) and (39)
- 13: $\tilde{\mathbf{M}} \xrightarrow{\text{SVD}}$ updated \mathbf{U}_c
- 14: * MLE (Exact resolution):
- 15: Compute $\{\mathbf{M}_r\}$ with (28)
- 16: Update \mathbf{U}_c by solving (A) (algorithm 15 of [50])
- 17: **end for**
- 18: CSP estimate: $\hat{\Pi}_c^{MLE} = \mathbf{V}_c \mathbf{V}_c^H$
- 19: $\hat{\mathbb{E}}(\tau) = \text{mean}(\{\tau_k\})$
- 20: CM estimate: $\hat{\Sigma}^{MLE} = \hat{\mathbb{E}}(\tau) \hat{\Sigma}_c + \sigma^2 (\mathbf{I}_M - \hat{\Pi}_c^{MLE})$

A-MLE: Both steps of this algorithm are ensured to provide a unique optimal solution: by [15] for Step-1 and by the uniqueness of SVD for Step-2. Nevertheless, since the overall algorithm alternately optimizes different objective functions (which is due to the relaxation in Step-2), its theoretical convergence remains unknown. The convergence of A-MLE algorithm is observed in practice, as illustrated in [47]. Moreover, this algorithm also leads to good performance, as shown in Section VII.

0-MLE: The convergence of this algorithm is trivial since it consists of performing Step-1 only.

VII. SIMULATIONS AND APPLICATION TO STAP

This section is devoted to numerical simulations to illustrate the performance of different CM estimators in the considered context (for criterions linked to CSP estimation that provide different results, see [47, 56]). We will study the following CM estimators:

- $\hat{\Sigma}_{SCM}$ is the classical SCM defined in Eq.(36).
- $\hat{\Sigma}_{S-FPE}$ is the Shrinkage-FPE (S-FPE) [24–29], which is a regularization of Tyler’s estimator [14] that can be computed for $K < M$, and is defined and unique for $\beta \in (\max(0, 1 - K/M), 1]$ by:

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

In the presented simulations, the S-FPE is computed with the iterative fixed point algorithm described in [27–29]. For the validations simulations, β is set to 0 (and S-FPE is referred to as FPE) in order to compare only unbiased estimators. For the detection simulations, different values of coefficient $\beta > 1 - K/M$ are tested.

- $\hat{\Sigma}_{RC-ML}$ is the estimate of the CM proposed in [57], namely RC-ML (for Rank Constrained ML). It consists in the SCM with thresholded eigenvalues so that the estimator satisfies the LR structure constraint. Let the SVD of the SCM be $\hat{\mathbf{U}}_{SCM}$ and $\mathbf{\Lambda}_{SCM} = \text{diag}(\{\lambda_m^{SCM}\})$. The following threshold:

$$\hat{\lambda}_m^{RC-ML} = \begin{cases} \max(\hat{\lambda}_m^{SCM}, \sigma^2) & \text{for } m \in \llbracket 1, R \rrbracket \\ \sigma^2 & \text{for } m \in \llbracket R+1, M \rrbracket \end{cases}$$

defines RC-ML, with:

$$\hat{\Sigma}_{RC-ML} = \hat{\mathbf{U}}_{SCM} \text{diag}(\{\hat{\lambda}_m^{RC-ML}\}) \hat{\mathbf{U}}_{SCM}^H \quad (40)$$

- $\hat{\Sigma}_{D-MLE}$ is the estimator derived from the algorithm proposed in [39], where $\{c_\tau\}$ and $\{\tau_k\}$ are estimated via regularized side parameters (as discussed in Remark 2 Section III).
- $\hat{\Sigma}_{MM1-MLE}$ and $\hat{\Sigma}_{MM2-MLE}$ are the estimator derived from the majorization-minimization algorithm proposed in [47]
- $\hat{\Sigma}_{0-MLE}$ is the estimator proposed in box *Algorithm 1* using the 0 order approximation for CSP estimation.
- $\hat{\Sigma}_{A-MLE}$ is the estimator proposed in box *Algorithm 1* using the average relaxation for CSP estimation.
- $\hat{\Sigma}_{MLE}$ is the estimator proposed in box *Algorithm 1* with exact resolution for CSP estimation.

For all the simulations, estimators that require iterative algorithm have been computed with 20 iterations.

A. CM estimation accuracy

In the following simulations we will study the accuracy of the different estimators regarding to the CM. The studied criterion is the Normalized Mean Square Error (NMSE) defined for a given estimator $\hat{\Sigma}$:

$$\text{NMSE}_{\hat{\Sigma}} = \frac{\|\hat{\Sigma}/\text{Tr}(\hat{\Sigma}) - \Sigma/\text{Tr}(\Sigma)\|^2}{\|\Sigma/\text{Tr}(\Sigma)\|^2}, \quad (41)$$

Note that we use the trace normalization to avoid any scaling ambiguity. We are therefore interested in the estimation of the shape of the total CM.

Simulation parameters: Secondary data have been generated according to the LR-SIRV plus WGN model described in Section II. To recall it quickly: one has $\mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k$. The WGN is distributed as $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ and $\sigma^2 = 1$. The LR-SIRV clutter is distributed as $(\mathbf{c}_k | \tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k \Sigma_c)$, 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 Σ_c is constructed with the first R eigenvectors and eigenvalues of a Toeplitz matrix of correlation parameter $\rho \in [0, 1]$ (obtained through SVD). In these simulations we fixed $\rho = 0.9$. The obtained matrix is then re-scaled to set the CNR thanks to the definition $\text{CNR} = \mathbb{E}(\tau) \text{Tr}(\Sigma_c) / (R \sigma^2)$.

Fig. 1 shows the evolution of the NMSE versus the number of samples K . The CNR is set to 30dB, so the high CNR assumption is satisfied. Results are displayed for moderately

($\nu = 1$) and highly ($\nu = 0.1$) heterogeneous clutter. First, one can see that all the MLE algorithms proposed in this paper provide the most accurate estimates of the CM for the considered context. Note that the MLE algorithms from previous works reach performance different from the ones proposed in this paper, which highlights the non-convexity of the likelihood and the possibility of getting trapped into a local maximum. This also demonstrates the interest of considering the high CNR assumption, as proposed in this work. We also note that these three estimators perform almost the same, meaning that the 0 order and average relaxation proposed are reliable to estimate the CSP. Nevertheless, we point out that, though it appears that the choice of the second step does not impact mainly the performance in terms of error on the estimation of the shape matrix, the exact resolution of this step may improve the performance in terms of CSP estimation as shown in [39]. Moreover this step can also provide an additional robustness to corruption of the samples by outliers, as shown in [58]. Also notice that when few samples are available, the approached solutions may provide lower NMSE (which is also observed in Fig. 2). This is due to the fact that the exact MLE relies on the estimation of more parameters, which may induce a slightly higher error in very under-sampled scenario. However, when more samples are available the exact MLE ultimately reaches the lowest NMSE. The SCM provides the less accurate estimator of the CM, which was expected since the noise is heterogeneous. RC-ML reaches the same performance as the SCM. Since this estimator is based on the SCM with a threshold on the eigenvalues, it seems that this threshold does not compensate the error induced by the use of the SCM in an heterogeneous context. D-MLE and MM1-MLE also reach the same performance as the SCM while MM2-MLE performs better. D-MLE was originally designed as a CSP estimator so it seems that this method leads to poor estimates of the eigenvalues, even if it is an accurate estimator of the CSP. The FPE (S-FPE with $\beta = 0$) reaches lower NMSE than SCM, RCML and D-MLE. However, FPE cannot be computed with $K < M$ and still does not provide as accurate estimation as the proposed methods for the considered context.

Fig. 2 shows the evolution of the NMSE for the same configuration, however the CNR is lowered to 10dB. The same general conclusions as previous can be drawn. One notes that D-MLE provides a slight improvement compared to SCM and RC-ML at low sample support. However, these estimators do not perform as well as the proposed MLE algorithms. In a low CNR configuration and moderately heterogeneous clutter ($\nu = 1$) all MLE based algorithms appear to reach close performance. We also notice that in some low CNR and highly heterogeneous ($\nu = 0.1$) clutter configurations, the FPE, MM1-MLE and MM2-MLE can outperform the proposed MLE algorithms, which could have been inferred as our proposed scheme relies on the high CNR assumption. However the proposed algorithms are still valid alternatives for these low CNR contexts, as shown in the next section.

Fig. 3 illustrates the robustness of the proposed approach to a wrong rank evaluation. It shows the evolution of the NMSE versus K for $A - MLE$ algorithms (since different

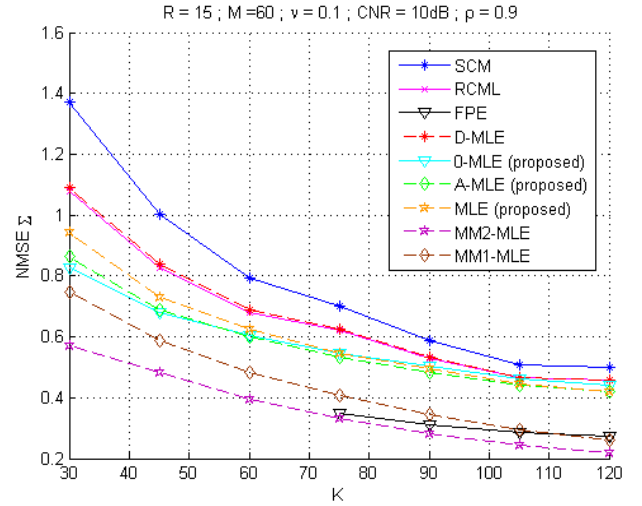
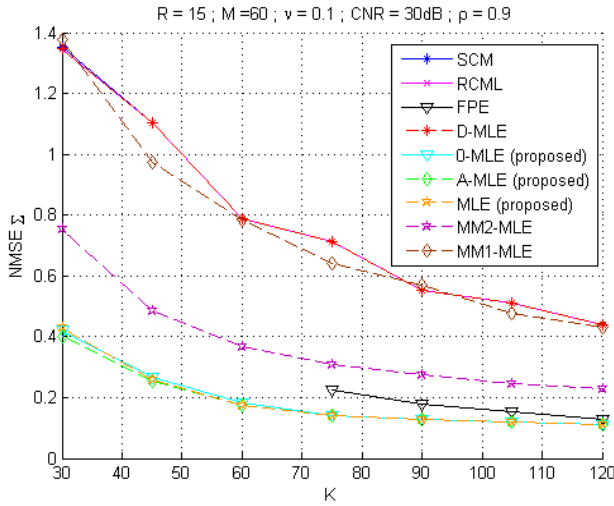
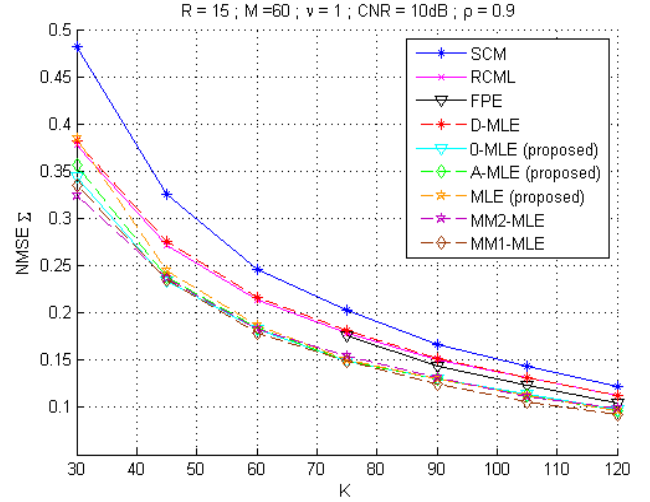
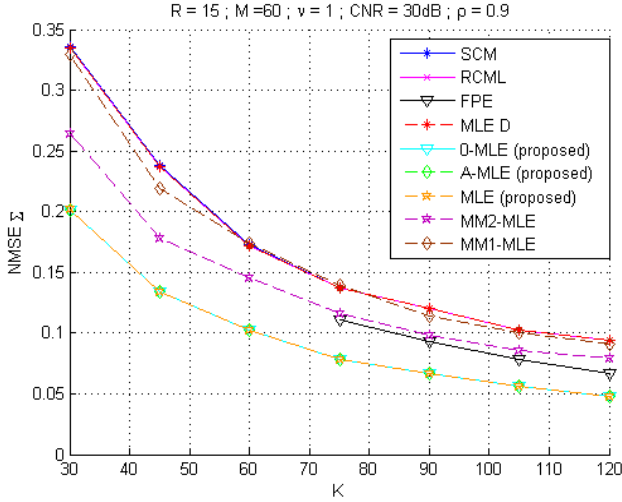


Fig. 1. NMSE versus K . $M = 60$, $R = 15$, $\text{CNR} = 30\text{dB}$. Top $\nu = 1$, bottom $\nu = 0.1$

Fig. 2. NMSE versus K . $M = 60$, $R = 15$, $\text{CNR} = 10\text{dB}$. Top $\nu = 1$, bottom $\nu = 0.1$

MLE schemes have very close performance) computed with different ranks \hat{R} around its actual value (here $R = 15$), we tested $\hat{R} = 15 \pm 5$. First, one can see that when the sample support is high enough (here $K = 40$) the A-MLE algorithm with true rank proposes the lower NMSE. Then one can notice that the performance of the proposed method is not strongly impacted by a wrong rank evaluation. For low sample support ($K = 30$), a lower rank \hat{R} involves less degree of freedom and could provide lower NMSE, however the estimator is biased. For higher sample support, one can see that over-evaluating the rank \hat{R} impacts less the estimation process than under-estimating it.

B. Application to STAP detection

STAP is a technique used in airborne phased array radar to detect moving target embedded in an interference background such as jamming or strong clutter [1]. The radar receiver consists in an array of Q antenna elements processing P pulses in a coherent processing interval. The received signal \mathbf{z} is composed of the sum of:

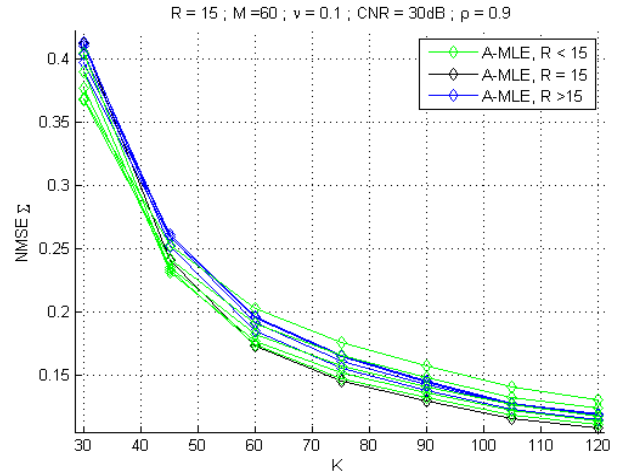


Fig. 3. NMSE versus K for A-MLE algorithm computed with different $\hat{R} \in [10, 20]$. Same configuration as in Fig. 1 (bottom): $M = 60$, $R = 15$, $\nu = 0.1$, $\text{CNR} = 30\text{dB}$.

- a possible target \mathbf{p} , the so-called STAP steering vector.

- ground clutter \mathbf{c} .
- thermal noise \mathbf{n} .

It is important to notice that 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 [40]. Hence, since this clutter could behave heterogeneously, the total interference can be modeled as LR-SIRV (the ground clutter) plus WGN (the thermal noise).

This section is devoted to the performance analysis of the different CM estimators on both synthetic and real STAP data set. This performance is analyzed through the target detection problem, formalized as a binary hypothesis test

$$\begin{cases} H_0 : \mathbf{z}_0 = \mathbf{c}_0 + \mathbf{n}_0 & ; \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \forall k \in \llbracket 1, K \rrbracket \\ H_1 : \mathbf{z}_0 = \mathbf{p} + \mathbf{c}_0 + \mathbf{n}_0 & ; \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \forall k \in \llbracket 1, K \rrbracket \end{cases}$$

where \mathbf{p} is the target and the $\mathbf{c}_k + \mathbf{n}_k \forall k \in \llbracket 0, K \rrbracket$ are the clutter plus noise realizations following the model considered in this paper. We use the Adaptive Normalized Matched Filter (ANMF) [59, 60], which is defined as:

$$\hat{\Lambda}(\hat{\Sigma}) = \frac{|\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{z}_0|^2}{|\mathbf{p}^H \hat{\Sigma}^{-1} \mathbf{p}| |\mathbf{z}_0^H \hat{\Sigma}^{-1} \mathbf{z}_0|} \underset{H_0}{\overset{H_1}{\geq}} \delta_{\hat{\Sigma}}, \quad (42)$$

for a given estimator of the CM $\hat{\Sigma}$, computed from the data $\{\mathbf{z}_k\}$ excluding \mathbf{z}_0 . For the rest of the paper, the adaptive detectors $\hat{\Lambda}$ will be denoted with the same index as the CM estimate they are build from, e.g. $\hat{\Lambda}_{SCM}$ stands for $\hat{\Lambda}(\hat{\Sigma}_{SCM})$, the ANMF build from the SCM.

1) Simulations:

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 respectively equal to $f_0 = 450$ MHz and $B = 4$ MHz. 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 Σ_c is computed with the model from [1], and its rank is computed from the Brennan rule [40] and is equal to $r = 15 \ll 64$, therefore, the low rank assumption is valid. The texture PDF is a Gamma law of shape parameter $\nu = 0.1$ and scale parameter $1/\nu$ so the clutter is distributed as a K-distribution. The target \mathbf{p} has a celerity of $V = 35$ m/s and is at $+10^\circ$ Azimuth. Clutter to Noise Ratio is defined as $\text{CNR} = \mathbb{E}(\tau) \text{Tr}(\Sigma_c) / \sigma^2$ and Signal to noise Ratio as $\text{SNR} = \text{norm}(\mathbf{p}) / \sigma^2$, and we set $\sigma^2 = 1$.

In these simulations, we first computed the False Alarm Probability in function of the threshold for 10^6 Monte Carlo Simulations under H_0 i.e. \mathbf{z}_0 is target free. This step allows to set proper threshold so that the False Alarm Probability is 10^{-3} for every detector in every considered configuration. Then, we compute and display the Detection Probability of the different detectors under H_1 in function of the SNR of the target \mathbf{p} , with 10^3 Monte Carlo Simulations. This process is realized for several configurations of the parameters (various ν , and CNR),

the number of secondary data is fixed to $K = 3R = 45$ (under-sampled configuration). We test the following detectors:

- $\hat{\Lambda}_{RC-ML}$: based on the CM estimate proposed in [57],
- $\hat{\Lambda}_{S-FPE}$: with different values of $\beta \in [0.5, 0.7, 0.9]$ since there is currently no rule for optimal adaptive selection of value of the β for the considered data model.
- $\hat{\Lambda}_{A-MLE}$: We do not display results for $\hat{\Lambda}_{0-MLE}$ and $\hat{\Lambda}_{MLE}$ since they are exactly the same as for $\hat{\Lambda}_{A-MLE}$. We do not display results for $\hat{\Lambda}_{0-MLE}$ and $\hat{\Lambda}_{MLE}$ since the previous section already showed that the estimators reach equivalent performance.
- $\hat{\Lambda}_{MM2-MLE}$ where the power of the thermal noise σ^2 is the same as for RC-ML. $\hat{\Lambda}_{MM1-MLE}$ is not displayed since the MM2 algorithm was shown to perform better in terms of CM estimation accuracy in the previous section.

Fig. 4 displays the probability of detection versus SNR for high clutter to noise ratio (CNR= 30dB), for slightly heterogeneous ($\nu = 1$) and very heterogeneous ($\nu = 0.1$) clutter. Fig. 5 displays the same results for an average clutter to noise ratio (CNR= 10dB). Result shows that for almost every configuration, the proposed CM estimator provides the best performance in terms of detection since its associated ANMF detector reaches the highest detection probability at different SNRs. The estimator A-MLE outperforms MM2-MLE at high CNR context. Interestingly, A-MLE also reaches performance equivalent to MM2-MLE at lower CNR even if it relies on high CNR assumption. Hence, the proposed estimation method is the most beneficial for the considered detection application. In some configurations, the estimator S-FPE can also be very close to A-MLE and even outperform it for some small value of the SNR (as in Fig. 4, bottom). However this is conditional to a good selection of the shrinkage parameter β . We emphasis that an "optimal" method for adaptive selection of this parameter is, to the best of our knowledge not derived yet for the configuration we consider (Low Rank heterogeneous clutter). Therefore, applying existent rules is not necessarily relevant. For example, we tested the oracle schemes proposed in [24] and [29]. These methods provided adaptive β almost always higher than 0.9, which is not the most efficient in the considered context. We also observe that the RC-ML does not lead to good performance for under-sampled configuration in presence of heterogeneous clutter, compared to the state of the art.

2) Real data:

The STAP real Data are provided by the French agency DGA/MI: the clutter is real but the targets are synthetic. The number of sensors is $Q = 4$ and the number of coherent pulses is $P = 64$, the size of the data is then $M = QP = 256$. The center frequency and the bandwidth are respectively equal to $f_0 = 10$ GHz and the bandwidth $B = 5$ MHz. The radar celerity is $V = 100$ m/s. The inter-element spacing is $d = 0,3$ m and the pulse repetition frequency is $f_r = 1$ kHz. The clutter rank, computed from Brennan Rule [40], is $R = 45$ and the CNR is evaluated around 20dB. The maximum number of target free samples available for all scenarios is $N = 408$. We

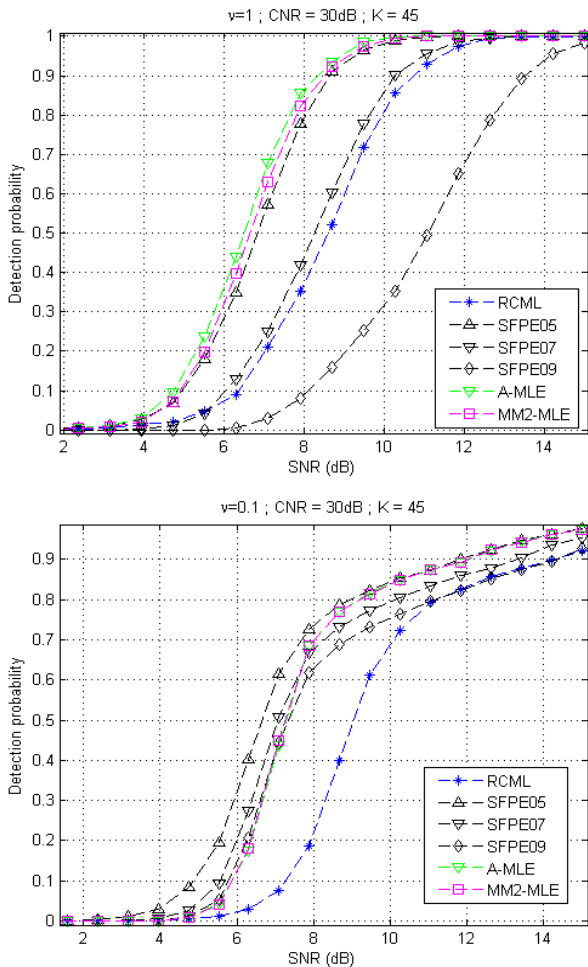


Fig. 4. Detection Probability versus SNR of the target for false alarm probability of the different detectors set to 10^{-3} . $M = 64$, $R = 15$, $\text{CNR} = 30\text{dB}$. Top $\nu = 1$, bottom $\nu = 0.1$

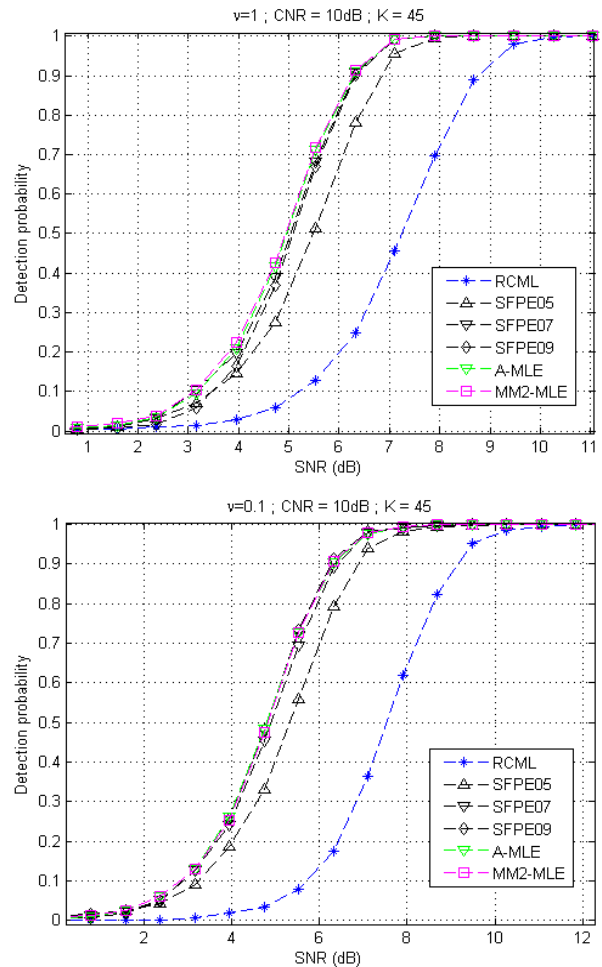


Fig. 5. Detection Probability versus SNR of the target for false alarm probability of the different detectors set to 10^{-3} . $M = 64$, $R = 15$, $\text{CNR} = 10\text{dB}$. Top $\nu = 1$, bottom $\nu = 0.1$

consider a test cell under H_1 with a target at (4 m/s, 0 deg) and with a Signal to Clutter Ratio (SCR) of -5dB . In the set, four guard cells are removed around the considered range cell. The covariance matrix estimator built on the set of secondary data, i.e. data that are assumed to be signal-free and i.i.d.

We test the following detectors:

- $\hat{\Lambda}_{SCM}$: Note that if $K < M$, this estimator can not be inverted. For this case, we used the following shrinkage:

$$\hat{\Sigma}_{S-SCM} = (1 - \beta)\hat{\Sigma}_{SCM} + \beta\mathbf{I}_M, \quad (43)$$

and used $\hat{\Lambda}_{S-SCM}$ instead of $\hat{\Lambda}_{SCM}$. As shown in [27], the value of β for S-SCM does not changes the obtained results for the considered application on the data set. We arbitrarily took $\beta = 0.75$.

- $\hat{\Lambda}_{RC-ML}$: with the power of the thermal noise σ^2 is pre-estimated with the whole data cube as the mean value of the last $M - R$ eigenvalues of the SCM [48] and this estimated value is used as its actual one.
- $\hat{\Lambda}_{S-FPE}$ with fixed $\beta = 0.75$. After several tests (also shown in [27]), the optimal value for this detection problem was found to be around $0.7 \sim 0.8$ depending on K .

- $\hat{\Lambda}_{A-MLE}$ where the power of the thermal noise σ^2 is the same as for RC-ML.

Fig. 6 and Fig. 7 present several detection output, i.e. the detection test performed over a grid of \mathbf{p} corresponding to different target angle and velocity. Fig. 6 shows detector output for an undersampled configuration with $K = 100$ (still $K > 2R$). $\hat{\Lambda}_{S-SCM}$ does not whiten the clutter, so the target present at (4 m/s, 0 deg) is not distinguishable. $\hat{\Lambda}_{RC-ML}$ and $\hat{\Lambda}_{S-FPE}$ allows to perform target detection but $\hat{\Lambda}_{A-MLE}$ provides the better interference rejection. One can also observe that both detectors build from structured estimators ($\hat{\Lambda}_{RC-ML}$ and $\hat{\Lambda}_{A-MLE}$) provide a better rejection of the clutter ridge, which illustrates the interest of using structure prior. Fig. 7 shows detector output for an oversampled configuration with $K = 300$. the same conclusions as previous can be drawn: the $\hat{\Lambda}_{SCM}$ fails to provide a reliable detector. $\hat{\Lambda}_{RC-ML}$ and $\hat{\Lambda}_{A-MLE}$ provide a better rejection of the clutter ridge. We remark that $\hat{\Lambda}_{RC-ML}$ and $\hat{\Lambda}_{A-MLE}$ have close performance here. This is to be expected since the clutter in the data set is not very heterogeneous, and therefore almost Gaussian.

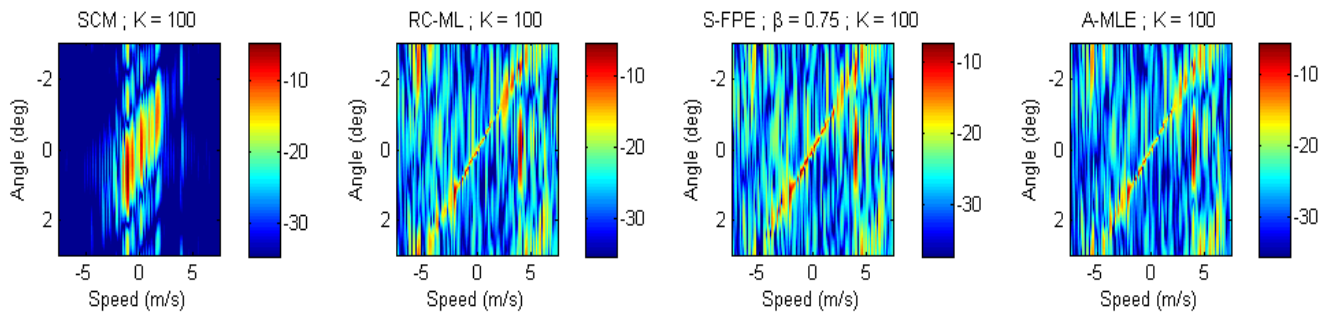


Fig. 6. Detector output on a test sample \mathbf{z}_0 containing a target at (4 m/s, 0 deg), from left to right: Λ_{SCM} (where only the clutter response is observable), Λ_{RC-ML} , Λ_{S-FPE} and Λ_{A-MLE} . $K = 100$.

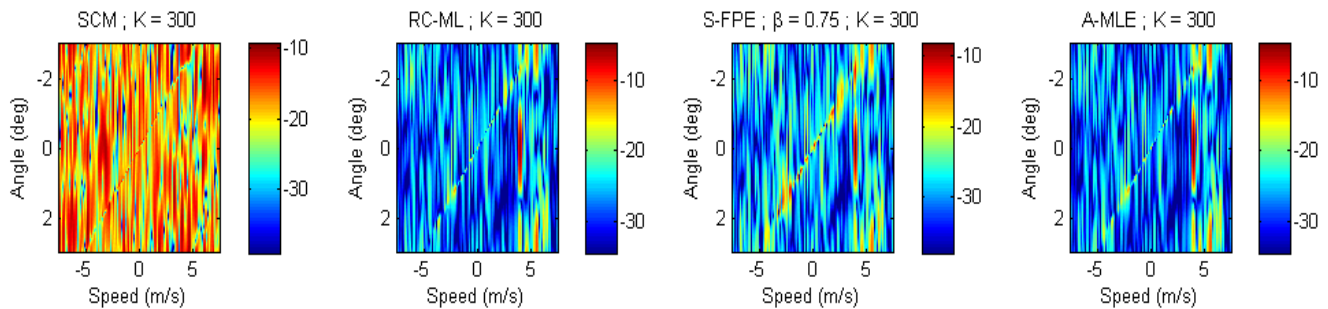


Fig. 7. Detector output on a test sample \mathbf{z}_0 containing a target at (4 m/s, 0 deg), from left to right: Λ_{SCM} , Λ_{RC-ML} , Λ_{S-FPE} and Λ_{A-MLE} . $K = 300$.

VIII. CONCLUSION

In this paper, we derived the MLE of the CM for a disturbance composed of LR-SIRV clutter plus an additive WGN. We assumed that the rank R of the LR-SIRV is known and that there is a high CNR. The problem is then separated into two parts: the estimation of the CM over the clutter subspace, and the estimation of the subspace itself. We derived the MLE expression of these parameters conditionally to each other, then resorted to a 2-Step MLE algorithm to compute the MLE of the CM. Among the derivation of the result, we discussed several relaxations and robust generalizations. Eventually, these results lead to the formulation of a very general 2-step MLE form algorithm to compute LR structured robust CM estimators. We focused then on the study of three adaptation of this algorithm namely "0-MLE", "A-MLE" and "MLE" that are not relying on the knowledge of the SIRV's texture PDF. Simulation results on NMSE and detection probability, as well as an application on a real STAP Data set, showed that this approach is useful to handle both heterogeneity of the clutter and prior on the CM rank structure. As regards the total covariance matrix estimation, the proposed estimators also lead to improvements compared to previous work on the same topic [39, 47] that were mainly focused on the clutter subspace estimation.

IX. ACKNOWLEDGMENTS

The authors would like to thank the French DGA/MI Agency for providing financial support and the STAP data set.

REFERENCES

- [1] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., MIT, Lexington, Mass., USA, Tech. Rep., 1994.
- [2] L. Scharf, *Statistical signal processing: detection, estimation and time series analysis*. Addison-Wesley Publishing Co., 1990.
- [3] J. R. Guerci, *Space-time adaptive processing for radar*. Artech House, 2003.
- [4] S. Haykin, *Adaptive radar signal processing*. John Wiley & Sons, 2007.
- [5] R. Maronna, R. Martin, and V. Yohai, *Robust Statistics*. John Wiley, 2006.
- [6] P. J. Huber and E. M. Ronchetti, *Robust statistics*. John Wiley & Sons Inc, 2009.
- [7] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5597–5625, 2012.
- [8] K. Yao, "A representation theorem and its applications to spherically invariant random processes," *Information Theory, IEEE Transactions on*, vol. 19, no. 5, pp. 600 – 608, 1973.
- [9] J. Billingsley, "Ground clutter measurements for surface-sited radar," MIT, Tech. Rep. 780, 1993.
- [10] E. Conte, A. De Maio, and C. Galdi, "Statistical analysis of real clutter at different range resolutions," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 40, no. 3, pp. 903–918, 2004.
- [11] 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, 2006.
- [12] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Compound-gaussian clutter modeling with an inverse gaussian texture distribution," *Signal Processing Letters, IEEE*, vol. 19, no. 12, pp. 876–879, 2012.
- [13] R. A. Maronna, "Robust M -estimators of multivariate location and scatter," *The Annals of Statistics*, vol. 4, no. 1, pp. 51–67, 1976.
- [14] D. Tyler, "A distribution-free M -estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
- [15] 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," *Signal Processing*,

- IEEE Transactions on*, vol. 56, no. 1, pp. 34–48, 2008.
- [16] 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," *Signal Processing, IEEE Transactions on*, vol. 50, no. 8, pp. 1908–1915, 2002.
- [17] F. Gini and A. Farina, "Vector subspace detection in compound-Gaussian clutter. Part I: survey and new results," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 4, pp. 1295–1311, 2002.
- [18] F. Gini, A. Farina, and M. Montanari, "Vector subspace detection in compound-Gaussian clutter. Part II: performance analysis," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 4, pp. 1312–1323, 2002.
- [19] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, 2002.
- [20] E. Ollila and D. Tyler, "Distribution-free detection under complex elliptically symmetric clutter distribution," in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 413–416, 2012.
- [21] O. Besson and Y. Abramovich, "Adaptive detection in elliptically distributed noise and under-sampled scenario," *Signal Processing Letters, IEEE*, vol. 21, no. 12, pp. 1531–1535, 2014.
- [22] M. Greco, S. Fortunati, and F. Gini, "Maximum likelihood covariance matrix estimation for complex elliptically symmetric distributions under mismatched condition," *Signal Processing Elsevier*, vol. 104, pp. 381–386, 2014.
- [23] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, 2004.
- [24] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4097–4107, 2011.
- [25] Y. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach - part 1: The over-sampled case," *Signal Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.
- [26] O. Besson and Y. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach - part 2: The under-sampled case," *Signal Processing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2013.
- [27] F. Pascal, Y. Chitour, and Y. Quek, "Generalized robust shrinkage estimator and its application to stap detection problem," *Signal Processing, IEEE Transactions on*, vol. 62, no. 21, pp. 5640–5651, 2014.
- [28] Y. Sun, P. Babu, and D. Palomar, "Regularized tyler's scatter estimator: Existence, uniqueness, and algorithms," *Signal Processing, IEEE Transactions on*, vol. 62, no. 19, pp. 5143–5156, Oct 2014.
- [29] E. Ollila and D. Tyler, "Regularized m -estimators of scatter matrix," *Signal Processing, IEEE Transactions on*, vol. 62, no. 22, pp. 6059–6070, 2014.
- [30] A. Wiesel, "Geodesic convexity and covariance estimation," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6182–6189, 2012.
- [31] I. Soloveychik and A. Wiesel, "Group symmetry and non-gaussian covariance estimation," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1105–1108, 2013.
- [32] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 29–38, 2012.
- [33] I. Soloveychik and A. Wiesel, "Tyler's covariance matrix estimator in elliptical models with convex structure," *Signal Processing, IEEE Transactions on*, vol. 62, no. 20, pp. 5251–5259, 2014.
- [34] M. Rangaswamy, I. Kirsteins, B. Freburger, and D. Tufts, "Signal detection in strong low rank Compound-Gaussian interference," in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 144–148, 2000.
- [35] M. Rangaswamy, F. Lin, and K. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing Elsevier*, vol. 84, pp. 1653–1665, 2004.
- [36] R. Raghavan, "Statistical interpretation of a data adaptive clutter subspace estimation algorithm," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 2, pp. 1370–1384, 2012.
- [37] G. Ginolhac, P. Forster, F. Pascal, and J. Ovarlez, "Derivation of the bias of the normalized sample covariance matrix in a heterogeneous noise with application to low rank STAP filter," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 514–518, 2012.
- [38] G. Ginolhac, P. Forster, F. Pascal, and J.-P. Ovarlez, "Performance of two low-rank STAP filters in a heterogeneous noise," *Signal Processing, IEEE Transactions on*, vol. 61, no. 1, pp. 57–61, 2013.
- [39] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Clutter subspace estimation in low rank heterogeneous noise context," *Signal Processing, IEEE Transactions on*, vol. 63, no. 9, pp. 2173–2182, 2015.
- [40] L. E. Brennan and F. Staudaher, "Subclutter visibility demonstration," RL-TR-92-21, Adaptive Sensors Incorporated, Tech. Rep., 1992.
- [41] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276–280, 1986.
- [42] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *Signal Processing Magazine, IEEE*, vol. 21, no. 4, pp. 36–47, 2004.
- [43] R. R. Nadakuditi and J. W. Silverstein, "Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 468–480, 2010.
- [44] L. Huang, C. Qian, H. C. So, and J. Fang, "Source enumeration for large array using shrinkage-based detectors with small samples," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 51, no. 1, pp. 344–357, 2015.
- [45] L. Huang and H. C. So, "Source enumeration via mdl criterion based on linear shrinkage estimation of noise subspace covariance matrix," *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4806–4821, 2013.
- [46] 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 the European Signal Processing Conference (EUSIPCO)*, 2013.
- [47] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *Signal Processing, IEEE Transactions on*, vol. 64, no. 8, pp. 1986–1998, 2016.
- [48] M. Tipping and C. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, p. 611–622, 1999.
- [49] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Robust estimation of the clutter subspace for a low rank heterogeneous noise under high clutter to noise ratio assumption," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 66–70, 2014.
- [50] J. Manton, "Optimization algorithms exploiting unitary constraints," *Signal Processing, IEEE Transactions on*, vol. 50, no. 3, pp. 635–650, 2002.
- [51] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.
- [52] T. Abrudan, J. Eriksson, and V. Koivunen, "Steepest descent algorithms for optimization under unitary matrix constraint," *Signal Processing, IEEE Transactions on*, vol. 56, no. 3, pp. 1134–1147, 2008.
- [53] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Technical note on clutter subspace projector MLE computation: Algorithm analysis," *Technical Report, available at sup-elec.fr/offres/file_inline_src/426/426_P_13413_547e056b16508_18.pdf*.
- [54] D. Rabideau and A. Steinhardt, "Improved adaptive clutter cancellation through data-adaptive training," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, no. 3, pp. 879–891, 1999.
- [55] 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.
- [56] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Cfar property and robustness of the lowrank adaptive normalized matched filters detectors in low rank compound gaussian context," in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2014, pp. 301–304.
- [57] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, no. 1, pp. 501–515, 2014.
- [58] A. Breloy, Y. Sun, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "A robust signal subspace estimator," *Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP)*, 2016.
- [59] S. Kraut and L. Scharf, "The cfar adaptive subspace detector is a scale-invariant glrt," *Signal Processing, IEEE Transactions on*, vol. 47, no. 9, pp. 2538–2541, 1999.
- [60] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *Signal Processing, IEEE Transactions on*, vol. 49, no. 1, pp. 1–16, 2001.



Arnaud Breloy 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 his Ph.D. degree from University of Paris Saclay in Signal Processing in 2015. His thesis was done in collaboration between SATIE (ENS Cachan) and SONDRRA (CentraleSupélec) laboratories. He is currently assistant professor at University Paris X and the LEME laboratory. His research interests focuses on statistical signal processing, array and radar signal processing

and robust estimation methods.



Guillaume Ginolhac received the degree of Electrical Engineer in 1997, the Ph.D. degree in Signal Processing from the Institut National Polytechnique de Grenoble (INPG) in 2001 and the HDR degree from ENS Cachan in 2011. He was Associate Professor of Université Paris X in the SATIE Lab from 2002 to 2010. HE is currently professor of Université de Savoie Mont-Blanc in the LISTIC Lab. His research interests are in estimation and detection theory with applications to array processing and radar/sonar.



Frédéric Pascal 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, advised by Pr. Philippe Forster: "Detection and Estimation in Compound Gaussian Noise" in 2006. This Ph.D thesis was in collaboration with the French Aerospace Lab (ONERA), Palaiseau, France. 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 (resp. Jan. 2012 - Dec. 2013), he was an Assistant Professor (resp. Associate Professor) in SONDRRA, the French-Singaporean laboratory of SUPELEC. In 2012, he obtained a Research Directorship Habilitation (HDR) thesis in Signal Processing from the University of Paris-Sud. From January 2014, Frederic Pascal is a full Professor in the L2S laboratory at SUPELEC. From January 2015, Frederic Pascal is an elected senior member of the IEEE Signal Processing Society SAM technical committee. His research interests contain estimation, detection and classification in statistical signal processing and radar processing.



Philippe Forster was born in Brest, France, in 1960. He received the Agrégation de physique appliquée degree from the Ecole Normale Supérieure de Cachan, France, in 1983 and the Ph.D. degree in electrical engineering from the Université de Rennes, France, in 1988. Currently, he is Professor of Electrical Engineering at Paris Ouest Nanterre University, Ville d'Avray, France, and member of the SATIE Laboratory at the Ecole Normale Supérieure de Cachan. His research interests are in estimation and detection theory with applications to array processing, radar, and digital communications.

processing, radar, and digital communications.