Clutter Subspace Estimation in Low Rank Heterogeneous Noise Context

Arnaud Breloy, *Student Member, IEEE*, Guillaume Ginolhac, *Member, IEEE*, Frédéric Pascal, *Senior Member, IEEE* and Philippe Forster, *Member, IEEE*.

Abstract—This paper addresses the problem of the Clutter Subspace Projector (CSP) estimation in the context of a disturbance composed of a Low Rank (LR) heterogeneous clutter, modeled here by a Spherically Invariant Random Vector (SIRV), plus a white Gaussian noise (WGN). In such context, the corresponding LR adaptive filters and detectors require less training vectors than classical methods to reach equivalent performance. Unlike classical adaptive processes, which are based on an estimate of the noise Covariance Matrix (CM), the LR processes are based on a CSP estimate. This CSP estimate is usually derived from a Singular Value Decomposition (SVD) of the CM estimate. However, no Maximum Likelihood Estimator (MLE) of the CM has been derived for the considered disturbance model. In this paper, we introduce the fixed point equation that MLE of the CSP satisfies for a disturbance composed of a LR-SIRV clutter plus a zero mean WGN. A recursive algorithm is proposed to compute this solution. Numerical simulations validate the introduced estimator and illustrate its interest compared to the current state of art.

Index Terms—Covariance Matrix and Projector estimation, Maximum Likelihood Estimator, Low-Rank clutter, SIRV.

I. INTRODUCTION

IN array processing, many applications require the use of the covariance matrix (CM) of the noise: source localization techniques [1, 2], radar and sonar detection methods [3, 4], adaptive filters [5]. For instance, the optimal filter in terms of Signal to Noise Ratio (SNR) is built from the inverse of the CM of the noise and the steering vector. In practice, the CM of the noise is unknown and has to be estimated from a set of secondary data, i.e. K signal-free independent realizations of the noise. The CM estimate is then used to process sub-optimal adaptive methods [6]. Estimating the CM or other interference parameters from a given data set is a fundamental issue in signal processing. The CM estimator typically used is the Sample Covariance Matrix (SCM), which is the Maximum Likelihood Estimator (MLE) of the CM in a Gaussian environment. In this case, $2M$ (where M is the

size of the data) secondary data are required to ensure good performance of the sub-optimal filtering, i.e. a 3dB loss of the output SNR compared to optimal filtering [7].

In some applications, the disturbance is not only Gaussian, but a sum of noises with different distributions. In various contexts [5, 8], it can be modeled by the sum of a correlated noise, referred to as *clutter* (caused by the response of the environment to the emitted signal), plus a white Gaussian noise (WGN, the thermal noise, due to electronics). When the correlated disturbance has a low-rank (LR) structure, which is the case for several applications, the corresponding sub-optimal LR-filter is based onto the projector on the clutter subspace instead of the CM of the noise [9, 10]. The advantage of LR method is that estimating the clutter subspace projector requires only $2R$ secondary data (where R is the clutter rank, and generally $R \ll M$) to reach equivalent performance as for the previous scheme [11]. The projector estimate is usually derived from the Singular Value Decomposition (SVD) of an estimate of the CM, classically the SCM.

Nevertheless, the SCM is not well adapted for samples that are not Gaussian (presence of outliers, heavy tailed-distributions...). Therefore, developing filters/detectors based on the SCM in highly heterogeneous or impulsive clutter environment may lead to poor performance. To describe this kind of disturbance, one of the most general model is provided by the Complex Elliptically Symmetric distribution (CES). A detailed review of this model is provided in [12]. Among the general CES class, we will focus on the Spherically Invariant Random Vectors (SIRV) [13] also referred to as Compound-Gaussian (CG) distributions. The SIRV family covers a large panel of well-known distributions, notably heavy-tailed such as Weibull and K-distribution and has been and extensively studied in the literature [14–17]. Moreover, SIRV present good agreement to several real clutter data sets [18, 19]. Eventually, the disturbance will be modeled in this paper as a LR-SIRV clutter plus WGN. This general model has already been used in [10, 20–22].

In this context, the clutter subspace estimate may be derived from the SVD of Tyler's estimator [23], also known as the Fixed-Point estimator (FPE) in the complex case [24], which is an approached MLE of the CM for full rank SIRV noise [25, 26]. Nevertheless, this approach presents two drawbacks. Firstly, this estimator is not the MLE of the CM for the considered disturbance and its corresponding model does not take into account the LR structure of the noise. Therefore, the clutter subspace estimate based on the FPE may lead to a loss of performance [27]. Secondly, the FPE requires $K > M$

secondary data to be computed, which is a problem for high-dimensional data[1]. Moreover, this requirement does not allows to take full advantage of the LR assumption in the cases where $2R \ll M$. Finally, the FPE may not be the most appropriate for the considered model of LR-SIRV plus WGN.

In this paper we then propose to develop a direct CSP MLE, assuming that the rank R is known. This approach is inspired by [21], where an estimator of the clutter subspace had been derived only under specific hypotheses: the CM of the LR SIRV clutter is assumed to have identical eigenvalues, and the Probability Density Function (PDF) of the texture is assumed to be known. The assumption of known texture PDF has been relaxed in [33] by considering texture as an unknown deterministic parameter. The main contribution of this paper is to relax the hypothesis of identical eigenvalues of the clutter CM. Thus, a new MLE of the CSP is introduced in the context of a LR-SIRV clutter plus a WGN. This MLE appears to be defined as the fixed point of an implicit equation, so the problem of its computation is also addressed. An iterative algorithm based on a 2-Step MLE approach is therefore proposed to approach the solution. The problem of uniqueness of the solution is not addressed, therefore the proposed algorithm could be trapped in local maximums.

This paper is organized as follows. Section II presents the statistical model of the noise composed of a LR heterogeneous clutter plus a WGN. Then, Section III introduces the solution of the CSP MLE. Since this solution is shown to be defined as a fixed point of an implicit function, its computation is discussed in section IV. In section V, the proposed estimator is compared with the state of art (SCM, NSCM and Shrinkage FPE) through Monte-Carlo simulations. Section VI draws conclusions of this study.

Finally, 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} the transpose operator. $\mathcal{CN}(\mathbf{a}, \boldsymbol{\Sigma})$ is a complex Gaussian vector of mean $\mathbf{a}$ and of covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{I}_M$ is the $M \times M$ identity matrix. $|\boldsymbol{\Sigma}|$ is the determinant of the matrix $\boldsymbol{\Sigma}$ and $\mathrm{Tr}()$ is the Trace operator. $\hat{d}$ 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 whose writing will often be contracted into $\{w_n\}$. $\mathbf{e}_i$ is the ith vector of the canonical basis of appropriate dimension.

II. STATISTICAL MODEL AND EXPRESSION OF THE LIKELIHOOD FUNCTION

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

$$\mathbf{z}_k = \mathbf{n}_k + \mathbf{c}_k . \tag{1}$$

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

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

where the power of the WGN σ^2 is assumed to be known[2]. It may be considered unitary since it appears to be only a scale factor fully determined by the Clutter to Noise Ratio (CNR). Therefore, for simplification we will assume $\sigma^2 = 1$ without loss of generality (w.l.o.g.). The LR-SIRV [13] $\mathbf{c}$ is a M-dimensional zero mean complex Gaussian vector (the *speckle*) of CM $\boldsymbol{\Sigma}_c$, multiplied by the square root of a positive random power factor (the *texture*) τ. We will assume that there is no prior information on the texture PDF. In that case, one can consider that each realization of the texture τ_k is an unknown deterministic positive parameter. Each realization $\mathbf{c}_k$ follows then, conditionally to τ_k:

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

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

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

with $\mathbf{C}_c$ the $R \times R$ diagonal matrix of the eigenvalues and $\mathbf{V}_c$ the $M \times R$ concatenation of the eigenvectors. We also define the Clutter Subspace orthogonal projector:

$$\boldsymbol{\Pi}_c = \sum_{r=1}^{R} \mathbf{v}_r \mathbf{v}_r^H = \mathbf{V}_c \mathbf{V}_c^H , \tag{5}$$

and its complementary $\boldsymbol{\Pi}_c^\perp = \mathbf{I}_M - \boldsymbol{\Pi}_c$.

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

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

with

$$\boldsymbol{\Sigma}_k = \mathbf{I}_M + \tau_k \boldsymbol{\Sigma}_c . \tag{7}$$

The likelihood of the dataset is then:

$$f(\{\mathbf{z}_k\} | \{\boldsymbol{\Sigma}_c\}, \{\tau_k\}) = \prod_{k=1}^{K} \frac{e^{-\mathbf{z}_k^H \boldsymbol{\Sigma}_k^{-1} \mathbf{z}_k}}{\pi^M |\boldsymbol{\Sigma}_k|} . \tag{8}$$

Up to a constant, the log-likelihood is:

$$\log f(\{\mathbf{z}_k\} | \{\boldsymbol{\Sigma}_c\}, \{\tau_k\}) = -\sum_{k=1}^{K} \mathbf{z}_k^H \boldsymbol{\Sigma}_k^{-1} \mathbf{z}_k - \sum_{k=1}^{K} \ln |\boldsymbol{\Sigma}_k| . \tag{9}$$

Using the definition of $\boldsymbol{\Sigma}_k$, one has:

$$\sum_{k=1}^{K} \ln |\boldsymbol{\Sigma}_k| = \sum_{k=1}^{K} \sum_{r=1}^{R} \log(1 + \tau_k c_r) . \tag{10}$$

---

[1] Different methods are currently proposed to overcome this issue. Recent and promising approaches are based on the regularization of the FPE algorithm [28–32].

[2] This 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.

The log-likelihood, up to a constant, can then be rewritten:

$$\log(f) = -\sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{V}_c (\tau_k \mathbf{C}_c + \mathbf{I}_R)^{-1} \mathbf{V}_c^H \mathbf{z}_k$$

$$-\sum_{k=1}^{K} \mathbf{z}_k^H (\mathbf{I}_M - \mathbf{\Pi}_c) \mathbf{z}_k - \sum_{k=1}^{K} \sum_{r=1}^{R} \log(1 + \tau_k c_r) \ . \quad (11)$$

The inversion of the matrix of the first term gives:

$$\mathbf{V}_c (\tau_k \mathbf{C}_c + \mathbf{I}_R)^{-1} \mathbf{V}_c^H = \sum_{r=1}^{R} \frac{1}{\tau_k c_r + 1} \mathbf{v}_r \mathbf{v}_r^H \ . \quad (12)$$

Leading to the following log-likelihood expression (ignoring constant terms):

$$\log f = \sum_{k=1}^{K} \sum_{r=1}^{R} \left[\frac{\tau_k c_r}{1 + \tau_k c_r} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \log(1 + \tau_k c_r) \right] \quad (13)$$

For identifiability and tractability purpose, we will also define the factors $d_r^k = \tau_k c_r$, representing the power of the clutter for each realization k and direction r. Which leads to the following relaxed log-likelihood:

$$\log f = \sum_{k=1}^{K} \sum_{r=1}^{R} \left[\frac{d_r^k}{1 + d_r^k} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - \log(1 + d_r^k) \right] \quad (14)$$

Where the parameters $\{d_r^k\}$ will be considered as unknown, deterministic and positive. This relaxed model appears as "more general" since it involves more parameters (RK instead of $R + K$). However, it does not take into account the inherent link between the d_r^k's. This problem will be discussed in section IV-A.

Let us now address the problem of the MLE of the CSP basis $\mathbf{V}_c$. Note that the CSP is defined $\mathbf{\Pi}_c = \mathbf{V}_c \mathbf{V}_c^H$, consequently, the term "MLE of the clutter subspace" will here indifferently refers to the "MLE of the CSP" $\hat{\mathbf{\Pi}}_c$ or the "MLE a clutter subspace basis" $\{\hat{\mathbf{v}}_r\}_{r \in [\![1,R]\!]}$. Nevertheless, we note that a basis defining the MLE of the CSP $\mathbf{\Pi}_c$ is not unique since it can be defined up to a rotation.

III. MLE OF THE CSP

The first part of this section is devoted to the expression of the MLE of the side parameters d_r^k and of the clutter subspace basis $\{\mathbf{v}_r\}$. The second part discusses the interpretation of the proposed result.

A. Clutter Subspace MLE

Considering the relaxed model (14), one has the following Proposition and Theorem:

Proposition 1 *The MLE under positivity constraint of d_r^k, denoted $\hat{d}_r^k$ for specific $r \in [\![1, R]\!]$ and $k \in [\![1, K]\!]$ is:*

$$\hat{d}_r^k = \begin{cases} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - 1 & if \ \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k > 1 \\ 0 & else \end{cases} , \quad (15)$$

Proof: The derivative of the expression (14) w.r.t. d_r^k, for specific $r \in [\![1, R]\!]$ and $k \in [\![1, K]\!]$ is:

$$\frac{\partial \ln f(\{\mathbf{z}_k\} | \{\mathbf{v}_r\}, \{d_r^k\})}{\partial d_r^k} = -\frac{1}{1 + d_r^k} + \frac{\mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k}{(1 + d_r^k)^2} \ . \quad (16)$$

This expression is cancelled to identify $\hat{d}_r^k$ the MLE of d_r^k. Nevertheless this power factor is assumed to be a positive value. Since the likelihood is strictly decreasing after its maximum $\hat{d}_r^k$, the MLE under the positivity constraint is then given by this Proposition 1. ∎

Theorem 1 *The MLE of the clutter subspace basis $\{\hat{\mathbf{v}}_r\}_{r \in [\![1,R]\!]}$ is the solution of the following constrained optimization problem:*

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

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

$$\hat{\mathbf{M}}_r = \sum_{k=1}^{K} \hat{\alpha}_r^k(\hat{\mathbf{v}}_r) \mathbf{z}_k \mathbf{z}_k^H \ , \quad (17)$$

with

$$\hat{\alpha}_r^k(\hat{\mathbf{v}}_r) = \frac{\hat{d}_r^k}{1 + \hat{d}_r^k} \ , \quad (18)$$

and with $\hat{d}_r^k$ the MLE of the parameters d_r^k

Proof: c.f. Appendix A ∎

The computation of such solution is not trivial and will be discussed in section IV.

B. Discussion

One may notice that the matrices $\{\hat{\mathbf{M}}_r\}$ defining different $\hat{\mathbf{v}}_r$'s in (17) are SCMs of the data scaled by estimated factors $\hat{\alpha}_r^k$. Since the power of the WGN is $\sigma^2 = 1$. One can also identify the terms:

$$\hat{\alpha}_r^k = \frac{\hat{d}_r^k}{1 + \hat{d}_r^k} = \frac{\widehat{CNR}_r^k}{1 + \widehat{CNR}_r^k} \ , \quad (19)$$

where $\widehat{CNR}_r^k = \hat{d}_r^k / \sigma^2$ represents the estimated CNR of the realization $\mathbf{z}_k$ over the subspace $\mathbf{\Pi}_r$. The $\hat{\alpha}_r^k$ factors are then corresponding to the MLE of the "power proportion" of the clutter over the subspace defined by $\mathbf{\Pi}_r$. So $\hat{\alpha}_r^k$ give more weight to $\mathbf{z}_k$'s with a strong CNR in the considered subspace. Moreover, the threshold imposed by the positivity constraint on $\hat{d}_r^k$ in Proposition 1 means that if the CNR of $\mathbf{z}_k$ is too low in the subspace $\mathbf{\Pi}_r$, this training vector is not taken into account. The same type of factor appears in [21] and [33] but its MLE expression differs since it involves the texture PDF and does not take into account the possible difference of the true CM eigenvalues. Yet, the ML estimation strategy is similar.

Usually, robust estimates of the full-rank CM tends to scale the data by a factor inversely proportional to the power of the noise (or the texture), such as for the Normalized SCM (NSCM) [34, 35], FPE [24–26], LR-FPE [36]. In this paper, the problem is the estimation of the CSP. The matrices $\hat{\mathbf{M}}_r$ do not define a MLE of the CM but only an intermediary to

obtain the LR clutter subspace MLE. Here, the scale factor $\hat{\alpha}_r^k$ involved 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. This approach is therefore robust from the point of view of estimating the CSP (not the CM). Such estimation method will, in principle, not ensure good robustness to outliers, as detailed in [36]. However it may be useful in some specific cases, where the clutter is not well represented over the secondary data set [37]. This scenario also corresponds to a low/average CNR and/or very impulsive clutter.

Moreover, for increasing CNR, one can intuitively expect that the scaling factors $\hat{\alpha}_r^k \simeq 1$. This leads to the following approximation $\mathbf{M}_r \simeq \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H$, meaning that the simple SVD of the SCM can provide a "good" approximation of the MLE of the CSP (at least for high CNR and low heterogeneity). This result is quite unexpected since the SCM is not a robust estimate of the CM in presence of SIRV distributed noise.

IV. CLUTTER SUBSPACE COMPUTATION

One may notice that the matrices $\hat{\mathbf{M}}_r$ are involving the vectors $\{\boldsymbol{v}_r\}$ so the clutter CM MLE is defined as a fixed point of an implicit function and is not directly reachable. Similar problem occurs in [21], and has been solved by using an Expectation-Maximization (EM) [38] algorithm which leads to a recursive method to compute the solution. In our case, the PDF of the d_r^k's is not assumed to be known and the EM approach can not be applied. Another solution, as proposed in [33], is the 2-Step MLE algorithm[3]. We will propose as well a 2-Step form algorithm to approach the solution, nevertheless resolution of the second step is here more complex than in [33], as shown in section IV-B.

In this section, we will propose and study an algorithm to approach this solution. This algorithm consists in alternatively maximizing the likelihood in terms of $\{d_r^k\}$ and $\{\mathbf{v}_r\}$. Since it corresponds to an alternate maximization of a bounded function, the convergence is ensured. Nevertheless, uniqueness of the solution is not proven. Moreover, we propose an optional regularization of the $\{d_r^k\}$ estimates for the first step. Its validity will be empirically demonstrated through the simulations section.

Since we will resort to iterative algorithms to reach a possible MLE solution, a good initialization is crucial to avoid possible local maximums of f. As discussed in [21] the SVD of the SCM provides a "good first guess" of the CSP. This intuition is confirmed by the last remark made in section III-B. The initial $\boldsymbol{\Pi}_c^{(0)}$ will be thus taken as the R strongest eigenvectors of the SCM.

A. Step 1: estimating $\{\hat{d}_r^k\}^{(n)}$ for fixed $\{\mathbf{v}_r\}^{(n-1)}$

At this step, one may simply use Proposition 1 to obtain the

---

[3] A parallel may here be explicited with previous works: when the disturbance is only composed by a full rank SIRV, [26] showed that the FPE algorithm corresponds to an approached EM proposed by [39] in the case of unknown texture PDF.

---

updated estimates $\{\hat{d}_r^k\}$. Nevertheless, Proposition 1 provides "bad" estimates since there is only one observation for each parameter. We propose then an optional regularization of these estimates. Although it is purely algorithmic, this regularization will be shown to provide a noticeable gain in performance.

As described in Section II, a SIRV corresponds to a specific case of the considered relaxed model with $d_r^k = \tau_k c_r$, where τ_k is the texture of the kth sample and c_r is the rth eigenvalue of the clutter CM. This section aims at proposing a specific method for recovering the parameters τ_k and c_r out of estimated d_r^k from Proposition 1. Indeed, in that case, the d_r^k's are correlated and one can use their inherent structure to improve the estimation process as well as to obtain separated estimates of both parameters. Firstly, identifiability conditions have to be set. Indeed, for any scale factor $\gamma \neq 0$ the couples $\{\tau_k, c_r\}$ and $\{\gamma\tau_k, c_r/\gamma\}$ may both describe the same SIRV $d_r^k = \tau_k c_r$. According to previous works in CM estimates, we will use the classical condition $\mathrm{Tr}(\boldsymbol{\Sigma}_c) = R$, which will therefore impose the constraint $\sum_{r=1}^{R} c_r = R$. Consequently, the problem is to separate variables τ_k and c_r from a given matrix $\mathbf{D}$ with $[\mathbf{D}]_{k,r} = d_r^k$:

$$\mathbf{D} = \begin{pmatrix} d_1^1 & \cdots & d_1^K \\ \vdots & \ddots & \vdots \\ d_R^1 & \cdots & d_R^K \end{pmatrix} , \qquad (20)$$

Since $d_r^k = \tau_k c_r$ the problem consists in finding two vectors $\boldsymbol{\tau}^{(n)}$ (the concatenated textures estimates) and $\mathbf{c}^{(n)}$ (the concatenated eigenvalues estimates) that satisfies

$$\mathbf{D} = \mathbf{c}\boldsymbol{\tau}^T = \begin{pmatrix} c_1 \\ \vdots \\ c_R \end{pmatrix} \begin{pmatrix} \tau_1 & \cdots & \tau_K \end{pmatrix} . \qquad (21)$$

The identifiability constraint imposes $\sum_{r=1}^{R} c_r = R$, so each column of $\mathbf{D}$ satisfies:

$$\sum_{r=1}^{R} d_r^k = \sum_{r=1}^{R} c_r \tau_k = \tau_k R , \qquad (22)$$

which leads to:

$$\tau_k = \sum_{r=1}^{R} d_r^k / R . \qquad (23)$$

One can then obtain the c_r's by the following ratios:

$$c_r = \frac{\sum_{k=1}^{K} d_r^k}{\sum_{k=1}^{K} \tau_k} \qquad (24)$$

One can derive regularized $\hat{d}_r^{k(n)}$ denoted $\tilde{d}_r^{k(n)}$ by applying the same process to the matrix $[\hat{\mathbf{D}}]_{k,r} = \hat{d}_r^k$, where the expression of $\hat{d}_r^k$ is given by Proposition 1. The regularization is then obtained with:

$$\tilde{d}_r^{k(n)} = \hat{\tau}_k^{(n)} \hat{c}_r^{(n)} = \frac{\sum_{r=1}^{R} \hat{d}_r^{k(n)} \sum_{k=1}^{K} \hat{d}_r^{k(n)}}{\sum_{k=1}^{K} \sum_{r=1}^{R} \hat{d}_r^{k(n)}} . \qquad (25)$$

To conclude, the exact MLE of the parameters $\{\tau_k, c_r\}$ are not directly obtainable. However the MLE of the relaxed parameters d_r^k are. Moreover, one can derive estimates of $\{\tau_k, c_r\}$ (that are not exact MLE) through the set $\{\hat{d}_r^k\}$

using their inherent structure. The obtained $\{\tau_k, c_r\}$ can be used to derive regularized estimates of the d_r^k's, in the sense that they satisfy the structure imposed by the non relaxed model. Validity and interest of this regularization approach will be justified in practice in the simulation sections.

B. Step 2: estimating $\{\mathbf{v}_r\}^{(n)}$ for fixed $\{d_r^k\}^{(n)}$

The second step involves the maximization of the first term of f in (14) w.r.t. $\mathbf{v}_r$'s for fixed d_r^k's.

$$\underset{\{\mathbf{v}_r\}}{\text{maximize}} \quad f_0(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \mathbf{v}_k^H \mathbf{M}_r^{(n)} \mathbf{v}_k$$
$$\text{subject to} \quad \mathbf{v}_r^H \mathbf{v}_r = 1, \ r \in [\![1,R]\!]$$
$$\mathbf{v}_r^H \mathbf{v}_j = 0, \ r,j \in [\![1,R]\!], \ r \neq j$$

with

$$\hat{\mathbf{M}}_r^{(n)} = \sum_{k=1}^{K} \frac{d_r^{k(n)}}{1 + d_r^{k(n)}} \mathbf{z}_k \mathbf{z}_k^H, \quad (26)$$

to obtain the updated clutter subspace basis $\{\hat{\mathbf{v}}_r\}_{r\in[\![1,R]\!]}^{(n+1)}$.

The optimization problem defining the MLE of the CSP basis belongs to the Non-Convex Quadratic Constrained Quadratic Program (QCQP) with Non-Convex constraints. Indeed, the objective is maximizing a quadratic function (unbounded) over a bounded set (due to the normality constraint). Moreover, the orthogonality constraints $\mathbf{v}_r^H \mathbf{v}_j = 0$ for $r \neq j$ are bilinear and non convex. The problem is close to a SVD problem: it would be one if the matrices $\mathbf{M}_r$ were all equal. It is also not a joint-diagonalization problem because the objective function is not function of $\sum_{r=1}^{R} ||\text{diag}(\mathbf{V}^H \hat{\mathbf{M}}_r \mathbf{V})||^2$. For this specific problem, there is to our knowledge no closed-form solution. Therefore, one has to resort to iterative algorithms to reach local maximums of f_0, as discussed below.

Such optimization problem is not trivial. To reach a local maximum of the function, we will use algorithm (15) from [40]: "Modified Steepest descent on Stiefel manifold". We refer to other possible algorithms [41, 42] and we also propose an other algorithm based on an heuristic with promising properties in [43]. 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 extremas. However, one could expect that previous step provides a point near to the global optimum. The algorithm could therefore be "wisely" initialized.

To sum up, the proposed algorithm, uniting steps 1 and 2 is described in the box Algorithm 1.

V. SIMULATIONS RESULTS

This section deals with numerical simulations to illustrate interest of the introduced estimator and its associated algorithm.

Algorithm 1 : 2-Step MLE algorithm for CSP estimation

1: Initialize $\{\mathbf{v}_r^{(0)}\}$ as the R strongest eigenvectors of the SCM
2: **for** $n = 1 \to N$ **do**
3:     Estimate $\{d_r^k\}^{(n)}$ conditionally to $\{\mathbf{v}_r\}^{(n-1)}$ with Proposition 1
4:     Optional regularization $\{d_r^k\}^{(n)} \leftarrow \{\tilde{d}_r^k\}^{(n)}$ with (25)
5:     Compute the set of matrix $\{\mathbf{M}_r\}^{(n)}$ with (17)
6:     Apply algorithm 15 of [40] to locally maximize

$$\sum_{r=1}^{R} \mathbf{v}_r^H \mathbf{M}_r^{(n)} \mathbf{v}_r$$

    under unitary constraints, the solution updates $\{\mathbf{v}_r\}^{(n)}$
7: **end for**
8: $\hat{\mathbf{\Pi}}_c = \sum_{r=1}^{R} \mathbf{v}_r^{(N)} \mathbf{v}_r^{(N)H}$

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 $(\mathbf{c}_k | \tau_k) \sim \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{\Sigma}_c)$, with a random texture τ_k generated for each sample. The texture PDF is a Gamma distribution (leading to a K-distributed clutter) of shape parameter ν et 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 first R eigenvectors and eigenvalues of a Toeplitz matrix of correlation parameter $\rho \in [0,1]$ (obtained through SVD). This matrix is then scaled so $\text{Tr}(\mathbf{\Sigma}_c) = \gamma M$ for a given constant γ. This scaling[4] allows to set the CNR thanks to the definition:

$$\text{CNR} = \frac{\mathbb{E}(\tau)\text{Tr}(\mathbf{\Sigma}_\mathbf{c})}{\text{Tr}(\sigma^2 \mathbf{I}_M)} = \gamma \quad (27)$$

Tested estimators: We will study the following estimators of the CSP:

- SCM: The CSP estimate derived from the SVD of the classical Sample Covariance Matrix

$$\hat{\mathbf{\Sigma}}_{\text{SCM}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \quad (28)$$

- NSCM: The CSP estimate derived from the SVD of the Normalized SCM, which is a consistent estimate of the CSP in the considered context [22].

$$\hat{\mathbf{\Sigma}}_{\text{NSCM}} = \frac{1}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \mathbf{z}_k} \quad (29)$$

- S-FPE: The CSP estimate derived from the SVD of the Shrinkage-FPE, also known as Diagonally-Loaded FPE [28, 30–32], which is a regularization of Tyler's estimator [23] that can be computed for $K < M$, and is defined and unique for $\beta \in (\max(0, 1 - K/M), 1]$ by:

$$\hat{\mathbf{\Sigma}}_{\text{S-FPE}}(\beta) = (1-\beta)\frac{M}{K} \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H}{\mathbf{z}_k^H \hat{\mathbf{\Sigma}}_{\text{S-FPE}}^{-1}(\beta)\mathbf{z}_k} + \beta \mathbf{I}_M \quad (30)$$

---

[4]Note that this scaling is not directly consistent with the constraint $\sum_{r=1}^{R} c_r = R$ imposed in the estimation section IV-A, however it will not impact results since the scale ambiguity is removed when using variables $\tilde{d}_r^k = \hat{c}_r \hat{\tau}_k$.

In the presented simulations, the S-FPE is computed with the algorithm described in [32] with a diagonal loading coefficient arbitrarily fixed to $\beta = 0.1$ if $K > M$ and $\beta = 1 - K/M + \epsilon$ if $K < M$ (with $\epsilon = 0.02$).

- A-MLE: the approached MLE of the CSP, derived in the context of LR-SIRV plus WGN assuming that the eigenvalues of the clutter CM are all equals [33].
- MLE: The MLE introduced in this paper. We present the results for 3 different algorithms:
  - Algorithm 1 without regularization of the d_r^k's estimates.
  - Algorithm 1 with regularization of the d_r^k's estimates.
  - As benchmark, we will also presents the results for the algorithm using actual d_r^k, which only involves the second step.

For all the simulations, estimators that require iterative algorithm have been computed with 30 iterations. The second step of Algorithm 1 has also been computed with 30 iterations.

A. Validation Simulations : NMSE

In these first simulations, the studied criterion is the Normalized Mean Square Error (NMSE), namely for a given estimator $\hat{\mathbf{\Pi}}_c$:

$$\frac{||\hat{\mathbf{\Pi}}_c - \mathbf{\Pi}_c||^2}{||\mathbf{\Pi}_c||^2} \quad , \tag{31}$$

with $||\mathbf{\Pi}_c||^2 = R$.

Figures 1 and 2 present respectively the NMSE of the estimators versus K for low CNR (3dB) and average CNR (10dB). First, we notice that in both cases, the NSCM reaches the worst performance. The S-FPE also reaches bad performances compared to others methods. In both cases, the MLE involving known d_r^k reaches the best performances, which was expected for a benchmark. We also notice the MLE approach improves the results compared to the SCM, especially for average CNR (figure 1). More precisely, the algorithm involving regularized $\tilde{d}_r^k$ is achieving the best results, while the one using $\hat{d}_r^k$ is close to the SCM. This illustrates the interest of the regularization of the side parameters. However, we also notice that A-MLE reach performance close to the MLE (with regularization) so it seems that when it comes to estimating clutter subspace only, precise information over the eigenvalues is not necessary. In the context of higher CNR (figure 2) the SCM tends to have the same performance as the MLE based algorithms, as it was inferred in the discussion section.

The NMSE criterion does not fully reflects the clutter nulling performance of an estimator $\hat{\mathbf{\Pi}}_c$, which will be considered in the next section. However we believe that the presented results could be useful for others applications. For example in the context of heterogeneous correlated sources embedded in WGN. The NMSE criterion could be relevant is the projector is function of parameters of interest $\boldsymbol{\theta}$: $\mathbf{\Pi}_c(\boldsymbol{\theta})$, like in MUSIC [1] application.
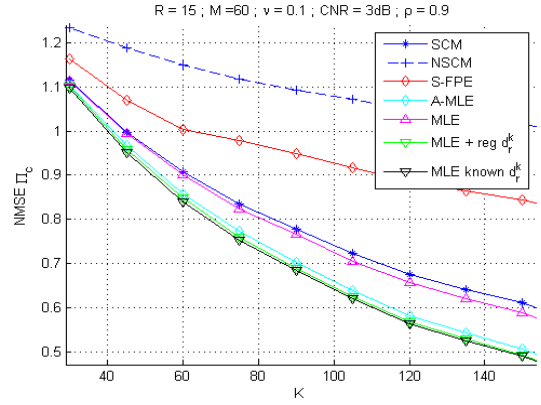

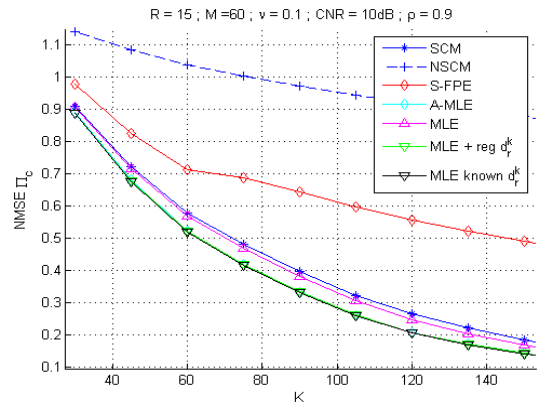Fig. 1. NMSE versus K. $M = 60$, $R = 15$, $\nu = 0.1$, CNR=3dB, $\rho = 0.9$.

Fig. 2. CSR versus K. $M = 60$, $R = 15$, $\nu = 0.1$, CNR=10dB, $\rho = 0.9$.

B. Clutter Suppression Ratio (CSR) simulations

In this section, we will study the following criterion, namely Clutter Suppression Ratio (CSR):

$$10 \log_{10} \left[\sum_{l=1}^{L} \mathbf{x}_l^H \left[\mathbf{I}_M - \hat{\mathbf{\Pi}}_c \right] \mathbf{x}_l \right] - 10 \log_{10} \left[\sum_{l=1}^{L} \mathbf{x}_l^H \mathbf{x}_l \right]$$

for a set of L clutter-plus-noise vectors $\{x_l\}$, independent of the vectors used for estimating the clutter subspace. This criterion assesses the clutter nulling performance of a given CSP estimator $\hat{\mathbf{\Pi}}_c$. Therefore, it is more reflective of an applicative use, as this not directly linked to the NMSE. For a given CNR γ, the analytic performance bound of this criterion is:

$$10 \log_{10} \left(\frac{M - R}{(\gamma + 1)M} \right) \tag{32}$$

For these simulations, according to previous results, we only kept the MLE estimator computed with algorithm 1 and regularization of the d_r^k parameters. Figures 3 to 5 present respectively the CSR of the estimators versus K for various CNR (resp. 3dB, 10dB, 20dB). The same conclusions as in the previous section can be drawn: when the CNR is high, the SCM has close performance to the MLE. When the CNR is low, the MLE approach provides a gain in terms of clutter
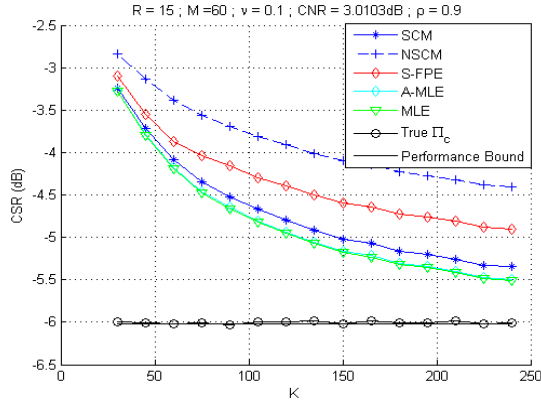
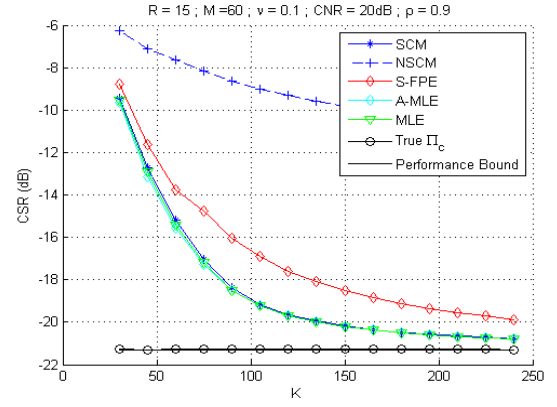Fig. 3. CSR versus K. $M = 60$, $R = 15$, $\nu = 0.1$, CNR=3dB, $\rho = 0.9$.

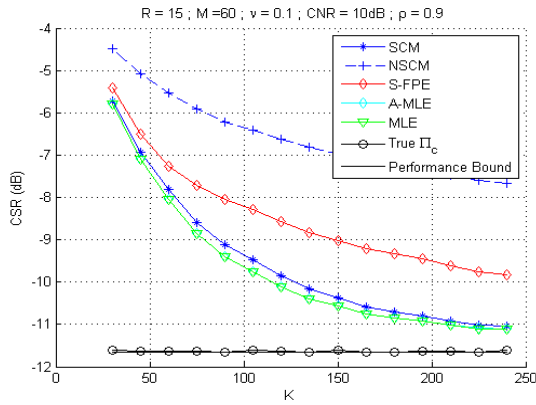Fig. 5. CSR versus K. $M = 60$, $R = 15$, $\nu = 0.1$, CNR=20dB, $\rho = 0.9$.

Fig. 4. CSR versus K. $M = 60$, $R = 15$, $\nu = 0.1$, CNR=10dB, $\rho = 0.9$.

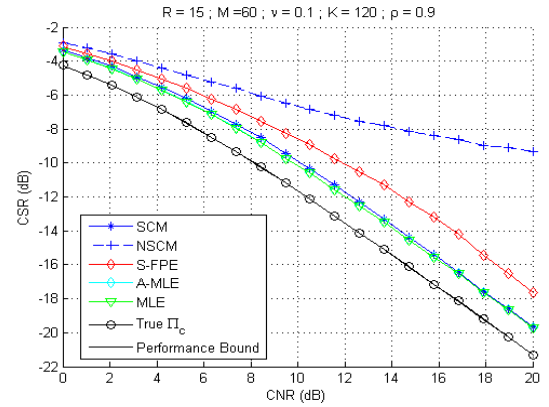Fig. 6. CSR versus CNR. $M = 60$, $R = 15$, $\nu = 0.1$, $K = 120$, $\rho = 0.9$.

Fig. 7. CSR versus $10\log_{10}(\nu)$. $M = 60$, $R = 15$, CNR=10dB, $K = 120$, $\rho = 0.9$.

nulling performance. However, there is no strong difference between A-MLE and MLE regarding to the CSR criterion.

To confirm the previous conclusions, figure 6 presents the CSR of the estimators versus the CNR. One can observe that SCM and MLE's tends to reach the same performance when the CNR increases and that the difference between SCM and MLE is visible only for low and average CNR. Figures 7 presents the CSR of the estimators versus ν. The parameter ν reflects the heterogeneity of the clutter: if it is close to 0, the clutter will be highly heterogeneous. On the contrary, if $\nu \gg 1$, the clutter will tend to have a more Gaussian behavior. First of all, we notice that the performance of every estimator decreases when the noise becomes more heterogeneous. However, this figure illustrates the robustness of the MLE approach: indeed, one can observe that every estimator will tend to have the same performance in the Gaussian case i.e. for increasing ν. However the proposed MLE (and also the A-MLE) is improving the CSR (compared to other methods) for heterogeneous clutter.

VI. CONCLUSION

In this paper have been proposed an estimator of the rank R CSP in the context of a LR-SIRV clutter (of known rank R) plus WGN, as modeled in Eq. (7) and (8) respectively. Th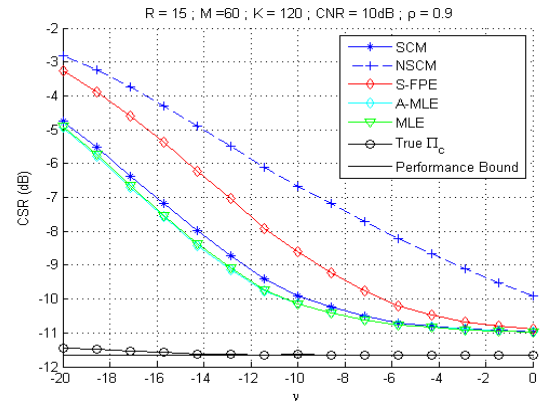is estimator is derived in Theorem 1 as the MLE of the CSP without prior knowledge of the SIRV texture PDF. The MLE of the CSP for the considered context is not obtainable in a closed form solution but as the fixed point of an implicit function. Therefore this paper proposed a 2-Step based MLE algorithm to compute this solution. The proposed algorithm ensures the convergence as it increases the likelihood at each step. However the problem of uniqueness of the solution has not been considered and it is possible that other rank R subspaces exist that satisfies Theorem 1, meaning that the proposed

algorithm could be trapped in local maximums. Simulation results showed that the introduced estimator outperforms the current state of art in terms of NMSE (accuracy of estimation) as well as in terms of CSR (clutter nulling perfromance). More precisely, the CSP MLE provides a robust estimator from the point of view of estimating the CSP: it improves the results when the sample set does not meet standard conditions (high CNR and low heterogeneity) i.e. when there is "less" clutter. Interestingly, a statistical interpretation (confirmed by simulations) has also shown that the SCM could provide a good approximation of the MLE in high CNR context. This result is unexpected since it does not involve robust estimate of the CM [12], even in the presence of an heterogeneous noise. The considered model fits several applications where the noise may have a LR heterogeneous structure such as STAP radar [5], hyperspectral detection or array processing. However, the problem of unknown rank R should be addressed in further works.

APPENDIX A
PROOF OF THEOREM 1

Before beginning the proof of Theorem 1, unitary constraint on $\{\mathbf{v}_r\}$ has to be expressed. Indeed maximizing the likelihood w.r.t. $\{\mathbf{v}_r\}$ without constraints does not inherently ensure fundamental properties on $\{\mathbf{v}_r\}$. The $\{\mathbf{v}_r\}$ has to form an ortho-normal basis of the estimated clutter subspace, therefore the vectors $\mathbf{v}_r$ have to satisfy unitary constraints, i.e.:

- $\mathbf{v}_r^H \mathbf{v}_r = 1$, for $r \in [\![1, R]\!]$
- $\mathbf{v}_r^H \mathbf{v}_j = 0$, for $r, j \in [\![1, R]\!]$ and $r \neq j$

The unitary constraint will be imposed using the Lagrange multipliers method. Let $\mathcal{L}(\{\mathbf{v}_r\})$ be the Lagrange function associated to the unitary constraints on $\{\mathbf{v}_r\}$:

$$\mathcal{L}(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \lambda_r (\mathbf{v}_r^H \mathbf{v}_r - 1)$$
$$+ \sum_{r=1}^{R} \sum_{l=r+1}^{R} \mu_{r,l} \, \mathbf{v}_r^H \mathbf{v}_l + \sum_{r=1}^{R} \sum_{l=r+1}^{R} \mu_{r,l}^H \, \mathbf{v}_l^H \mathbf{v}_r \, , \quad (33)$$

where λ_r are the Lagrange multipliers associated to the normalization constraint and $\mu_{r,l}$ the ones associated to the orthogonality constraint[5]. One has the following Lemma :

Lemma 1 *The derivative of this $\mathcal{L}(\{\mathbf{v}_r\})$ w.r.t. $\mathbf{v}_j^H$ for $j \in [\![1, R]\!]$ is:*

$$\frac{\partial \mathcal{L}(\{\mathbf{v}_r\})}{\partial \mathbf{v}_j^H} = \mathbf{V}\mathbf{\Lambda}\mathbf{e}_j \, , \quad (34)$$

where $\mathbf{V}$ is the concatenation of the vectors $\mathbf{v}_r$, $\mathbf{e}_j$ is the jth vector of the canonical basis, and $\mathbf{\Lambda}$ is the Hermitian matrix defined in defined in (36).

Proof: One has

$$\frac{\partial \mathcal{L}(\{\mathbf{v}_r\})}{\partial \mathbf{v}_j^H} = \lambda_j \mathbf{v}_j + \sum_{l=j+1}^{R} \mu_{j,l} \mathbf{v}_l + \sum_{r=1}^{j-1} \mu_{r,j}^H \mathbf{v}_r \quad (35)$$

---
[5]The presence of the complex conjugate sum associated to the $\mu_{r,l}$'s is required since we are dealing with complex vectors.

This expression may be rewritten using the $R \times R$ Lagrange constraints matrix $\mathbf{\Lambda}$, defined by:

$$[\mathbf{\Lambda}]_{i,j} = \begin{cases} \mu_{i,j}^H & \text{if } i < j \\ \lambda_j & \text{if } i = j \\ \mu_{j,i} & \text{if } i > j \end{cases} \, , \quad (36)$$

which leads to Lemma 1. At this point, it is important to notice that the matrix $\mathbf{\Lambda}$ is Hermitian since $[\mathbf{\Lambda}]_{i,j} = [\mathbf{\Lambda}]_{j,i}^H$. ∎

To derive the MLE of $\{\mathbf{v}_r\}$, one replace the expression of the MLE of the d_r^k (function of the $\mathbf{v}_r$'s) to obtain the reduced log-likelihood. The obtained expression is then differentiated w.r.t. the $\mathbf{v}_r$'s to identify the geometric relation that should satisfy the MLE. Then, Lemma 2 allows to express the MLE as a solution of a constrained optimization problem, which concludes the proof of Theorem 1. We first derive Lemma 2:

Lemma 2 *Let $\{\mathbf{M}_r\}$ be a set of R $M \times M$ matrices. The following constrained optimization problem*

$$\begin{aligned} \underset{\{\mathbf{v}_r\}}{maximize} \quad & f_0(\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \mathbf{v}_k^H \mathbf{M}_r \mathbf{v}_k \\ subject \ to \quad & \mathbf{v}_r^H \mathbf{v}_r = 1 \, , \ r \in [\![1, R]\!] \\ & \mathbf{v}_r^H \mathbf{v}_j = 0 \, , \ r, j \in [\![1, R]\!] \, , \ r \neq j \end{aligned}$$

has local solutions that are satisfying:

$$\mathbf{V}^H [\mathbf{M}_1\mathbf{v}_1|...|\mathbf{M}_R\mathbf{v}_R] = \mathbf{\Lambda} \quad (37)$$

where $\mathbf{V}$ is the concatenation of $\{\mathbf{v}_r\}$, and $\mathbf{\Lambda}$ is the Hermitian matrix of the Lagrange multipliers associated to unitary constraints.

Proof: Using the Lagrange constraints method, one has to maximize the functional $f_0(\{\mathbf{v}_r\}) - \mathcal{L}(\{\mathbf{v}_r\})$. One can cancel the derivative of this functional w.r.t $\mathbf{v}_j$:

$$\frac{\partial (f_0 - \mathcal{L})(\{\hat{\mathbf{v}}_r\})}{\partial \mathbf{v}_j^H} = 0 \Leftrightarrow \mathbf{M}_j \hat{\mathbf{v}}_j = \hat{\mathbf{V}}\mathbf{\Lambda}\mathbf{e}_j, \quad (38)$$

Then, all the R solutions w.r.t. each $\hat{\mathbf{v}}_r$ may be concatenated in a single equation:

$$[\mathbf{M}_1\mathbf{v}_1|...|\mathbf{M}_R\mathbf{v}_R] = \mathbf{V}\mathbf{\Lambda} \quad (39)$$

Multiplying 39 left by $\mathbf{V}^H$ and using the relation $\mathbf{V}^H\mathbf{V} = \mathbf{I}_R$ (verified under the unitary constraint) concludes the proof of this lemma. ∎

Let us now turn to the formal proof of Theorem 1 :

Proof: The d_r^k's are replaced by their MLE expression from Proposition 1 in (14) in order to obtain the reduced log-likelihood $\hat{f}$:

$$\hat{f}(\{\mathbf{z}_k\}|\{\mathbf{v}_r\}) = \sum_{r=1}^{R} \sum_{k=1}^{K} \mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k - RK$$
$$- \sum_{r=1}^{R} \sum_{k=1}^{K} \ln(\mathbf{z}_k^H \mathbf{v}_r \mathbf{v}_r^H \mathbf{z}_k) \, , \quad (40)$$

the derivative of $\hat{f}$ w.r.t. $\mathbf{v}_j^H$, $j \in [\![1, R]\!]$ is:

$$
\begin{aligned}
\frac{\partial \hat{f}}{\partial \mathbf{v}_j^H} &= \sum_{k=1}^{K} \mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j - \sum_{k=1}^{K} \frac{\mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j}{\mathbf{z}_k^H \mathbf{v}_j \mathbf{v}_j^H \mathbf{z}_k} \\
&= \sum_{k=1}^{K} \frac{\mathbf{z}_k^H \mathbf{v}_j \mathbf{v}_j^H \mathbf{z}_k - 1}{\mathbf{z}_k^H \mathbf{v}_j \mathbf{v}_j^H \mathbf{z}_k} \mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j \\
&= \sum_{k=1}^{K} \frac{\hat{d}_j^k}{1 + \hat{d}_j^k} \mathbf{z}_k \mathbf{z}_k^H \mathbf{v}_j
\end{aligned}
\tag{41}
$$

Eventually, if we denote $\hat{\mathbf{M}}_r$ the matrices:

$$
\hat{\mathbf{M}}_r = \sum_{k=1}^{K} \hat{\alpha}_r^k(\mathbf{v}_r) \mathbf{z}_k \mathbf{z}_k^H \ ,
\tag{42}
$$

with

$$
\hat{\alpha}_r^k(\mathbf{v}_r) = \frac{\hat{d}_r^k}{1 + \hat{d}_r^k} \ ,
\tag{43}
$$

the derivative of $\hat{f}$ w.r.t. $\mathbf{v}_j^H$ is then expressed as:

$$
\frac{\partial \hat{f}(\{\mathbf{z}_k\} | \{\mathbf{v}_r\})}{\partial \mathbf{v}_j^H} = \hat{\mathbf{M}}_j \mathbf{v}_j
\tag{44}
$$

Note that the case $\hat{d}_r^k = 0$ (due to the threshold in Proposition 1) has not been not considered w.l.o.g. since it only means that $\mathbf{z}_k$ is not taken into account in the definition of $\hat{\mathbf{M}}_r$. Indeed, in that case $\hat{\alpha}_r^k = 0$. One can also redefine each sum in (40) by only involving the $\hat{d}_r^k \neq 0$ and obtain the same result.

Combining the Lagrange constraint method and $\hat{f}$ leads to maximize the following functional g w.r.t. the $\mathbf{v}_r$'s:

$$
g(\{\mathbf{v}_r\}) = \hat{f}(\{\mathbf{z_k}\} | \{\mathbf{v}_r\}) - \mathcal{L}(\{\mathbf{v}_r\}) \ ,
\tag{45}
$$

which derivative w.r.t $\mathbf{v}_j$, given by (34) and (44) and is cancelled for:

$$
\frac{\partial g(\{\hat{\mathbf{v}}_r\})}{\partial \mathbf{v}_j^H} = 0 \Leftrightarrow \hat{\mathbf{M}}_j \hat{\mathbf{v}}_j = \hat{\mathbf{V}} \mathbf{\Lambda} \mathbf{e}_j,
\tag{46}
$$

All the R solutions w.r.t. each $\hat{\mathbf{v}}_r$ may be concatenated in a single equation:

$$
\left[\hat{\mathbf{M}}_1 \hat{\mathbf{v}}_1 | ... | \hat{\mathbf{M}}_R \hat{\mathbf{v}}_R \right] = \hat{\mathbf{V}} \mathbf{\Lambda}
\tag{47}
$$

Using the relation $\hat{\mathbf{V}}^H \hat{\mathbf{V}} = \mathbf{I}_R$ (verified under the unitary constraint) gives:

$$
\hat{\mathbf{V}}^H \left[\hat{\mathbf{M}}_1 \hat{\mathbf{v}}_1 | ... | \hat{\mathbf{M}}_R \hat{\mathbf{v}}_R \right] = \mathbf{\Lambda}
\tag{48}
$$

One can then apply Lemma 2 to expresses the MLE of the CSP basis as solution of a constrained optimization problem, which concludes the proof of Theorem 1. ∎

REFERENCES

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

[2] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariant techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on* vol. 37, no. 7, pp. 984–995, July 1989.

[3] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 208 – 216, 1992.

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

[5] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Lab., MIT, Lexington, Mass., USA, Tech. Rep., December 1994.

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

[7] I. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-10, no. 6, pp. 853 – 863, November 1974.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing - Detection Theory*. Prentice-Hall PTR, 1998, vol. 2.

[9] I. Kirsteins and D. Tufts, "Adaptive detection using a low rank approximation to a data matrix," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 30, pp. 55 – 67, 1994.

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

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

[12] 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.

[13] 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, September 1973.

[14] F. Gini and M. Greco, "Suboptimum approach to adaptive coherent radar detection in compound-Gaussian clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, no. 3, pp. 1095–1104, 1999.

[15] E. Conte, M. Lops, and G. Ricci, "Asymptotically optimum radar detection in compound-Gaussian clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 31, no. 2, pp. 617–625, 1995.

[16] 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.

[17] 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.

[18] J. Billingsley, "Ground clutter measurements for surface-sited radar," MIT, Tech. Rep. 780, February 1993.

[19] M. Rangaswamy, D. Weiner, and A. Ozturk, "Non-Gaussian vector identification using spherically invariant random processes," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 29, no. 1, pp. 111–124, January 1993.

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

[21] 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.

[22] 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.

[23] D. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[24] 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, January 2008.

[25] 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, August 2002.

[26] 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, December 2002.

[27] J. H. Michels, M. Rangaswamy, and B. Himed, "Performance of parametric and covariance based STAP tests in compound-Gaussian clutter," *Digital Signal Processing*, vol. 12, pp. 307–328, April-July 2002.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] E. Conte and G. Ricci, "Performance prediction in compound-gaussian clutter," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 30, no. 2, pp. 611–616, 1994.

[35] S. Bausson, F. Pascal, P. Forster, J. Ovarlez, and P. Larzabal, "First- and second-order moments of the normalized sample covariance matrix of spherically invariant random vectors," *IEEE Signal Processing Letters*, vol. 14, no. 6, pp. 425 – 428, June 2007.

[36] 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 ICASSP*, Florence, Italy, May 2014.

[37] D. Rabideau and A. Steinhardt, "Improved adaptive clutter cancellation trough data-adaptive training," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, no. 3, pp. 879 – 891, July 1999.

[38] N. M. L. A. P. Dempster and D. B. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm," in *Royal Statistical society*, 1976.

[39] N. Pulsone and R. S. Raghavan, "Analysis of an adaptive CFAR detector in non-Gaussian interference," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 35, no. 3, pp. 903–916, 1999.

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

[41] 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.

[42] 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, March 2008.

[43] A. Breloy, G. Ginolhac, F. Pascal, and P. Forster, "Technical note on clutter subspace projector mle computation: Algorithm analysis," *Technical Report, available at supelec.fr/offres/file_inline_src/426/426_P_13413_547e056b16508_18.pdf*

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.
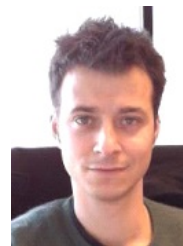
Frédéric Pascal received the Master's degree ("Probabilities, Statistics and Applications: Signal, Image et Networks") with merit, in Applied Statistics from University Paris VII - Jussieu, Paris, France, in 2003. Then, he received the Ph.D. degree of Signal Processing, from University Paris X - Nanterre, 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 SONDRA, 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.

Arnaud Breloy graduated from Ecole Centrale Marseille and recived a Master's degree of Signal and Image Processing from university of Aix-Marseille in 2012-2013. He is currently Ph.D Student at the ENS Cachan, SATIE laboratoy. His research interests focuses on statistical signal processing, array and radar signal processing, robust and low rank statistical methods.