Bayesian Robust Signal Subspace Estimation in Non-Gaussian Environment

R. Ben Abdallah, A. Breloy, M. N. El Korso, David Lautru

Bayesian Robust Signal Subspace Estimation in Non-Gaussian Environment

R. Ben Abdallah[1], A. Breloy[1], M. N. El Korso[1] and D. Lautru[1]
[1] LEME (EA 4416), IUT de ville d'Avray, Universit Paris Nanterre, France

***Abstract*—In this paper, we focus on the problem of low rank signal subspace estimation. Specifically, we derive new subspace estimator using the Bayesian minimum mean square distance formulation. This approach is useful to overcome the issues of low sample support and/or low signal to noise ratio. In order to be robust to various signal distributions, the proposed Bayesian estimator is derived for a model of sources plus outliers, following both a compound Gaussian distribution. In addition, the commonly assumed complex invariant Bingham distribution is used as prior for the subspace basis. Finally, the interest of the proposed approach is illustrated by numerical simulations and with a real data set for a space time adaptive processing (STAP) application.**

***Index Terms*—Subspace estimation, Bayesian estimation, minimum mean square distance, compound Gaussian, complex invariant Bingham distribution, STAP.**

I. Introduction

An ubiquitous problem in statistical signal processing is to infer the low rank signal subspace where the information lies in. It is a fundamental problem with widespread applications such as PCA [1], interference cancellation [2] and direction of arrival (DoA) estimation [3].

Classically, the subspace estimator is derived from the singular value decomposition (SVD) of the sample covariance matrix (SCM) by extracting its leading eigenvectors. The latter corresponds to the maximum likelihood (ML) estimator in a Gaussian setting. This estimation process presents good performance at high signal to noise ratio (SNR) or for large number of samples. However, outside these standard regimes, the latter usually exhibits poor performance. Additionally, it is not robust to the presence of outliers and/or heavy tailed distributed observation.

First, a possible solution to overcome the sensitivity to low SNR and/or small number of samples is to introduce prior information on the subspace of interest. More specifically, several approaches adopted a Bayesian setting in the estimation process. Notably, the minimum mean square distance (MMSD) approach showed its interest in the non standard regimes [4], [5]. The MMSD minimizes the average Euclidean distance between the true subspace and its estimate, which is a natural formulation for the problem of subspace estimation.

Second, to overcome the robustness issue, a possible alternative is to consider the compound Gaussian (CG) distribution modeling [6]. Such family of distributions provides a good fit to non Guassian real data [7] and allows to develop robust estimation procedures [8]–[10].

In order to enjoy best of both the CG modeling and the Bayesian subspace estimation approaches, we derive a new MMSD estimator for the data modeled as a CG distributed signal of interest embedded in heterogeneous noise (CG outliers plus white Gaussian noise) [11]. For this model, the complex invariant Bingham (CIB) [12] is considered as a prior of the subspace of interest.

We adopt the following notations: the operator $\text{Tr}\{.\}$ stands for the trace of a given matrix and $\mathbb{E}\{.\}$ is the expectation operator, $\text{etr}\{.\}$ is the exponential of the trace of a given matrix, $\propto$ stands for "proportional to", $\det(.)$ is the determinant operator, $\mathcal{P}_P\{.\}$ is an operator which extracts the P strongest eigenvectors of a given matrix, $\text{diag}(.)$ is a diagonal matrix built from a given vector and $\mathcal{U}_P^N = \left\{\mathbf{U} \in \mathbb{C}^{N\times P} | \mathbf{U}^H\mathbf{U} = \mathbf{I}\right\}$ is the set of $N \times P$ unitary matrices. $\{\mathbf{w}_n\}_{n\in 1\ldots N}$ denotes the set of vectors $\mathbf{w}_n, \forall n \in \{1\ldots N\}$, this writing is often be contracted in $\{\mathbf{w}_n\}$ and $\overset{d}{=}$ stands for "has the same distribution as".

II. Background

A. Compound Gaussian distribution

The CG distribution is commonly used in signal processing applications for its good fit to real data [7]. It encloses a large panel of distributions such as K-distribution, Weibull distribution and student-t distribution [6]. A CG distributed random vector $\mathbf{s}$ is denoted as $\mathbf{s} \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, \tau)$ with

$$\mathbf{s} \overset{d}{=} \sqrt{\tau}\mathbf{g} \tag{1}$$

where, τ is a positive random scalar, called texture, $\mathbf{g} \sim \mathcal{CN}(0, \mathbf{\Sigma})$ is called the speckle. In this work, we assume that there is no prior information on the texture probability density function (pdf), and consider this parameter as deterministic unknown. This relaxation provides more robustness to various signal distributions [8], and such assumption has few impact in terms of subspace estimation accuracy [13].

B. Complex invariant Bingham distribution

The CIB [12] is a probability distribution for a matrix $\mathbf{U} \in \mathcal{U}_P^N$ that represents a subspace orthonormal basis. We denote $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ when $\mathbf{U}$ has as pdf of the form

$$p_{\text{CIB}}(\mathbf{U}) \propto \text{etr}\left\{\kappa\mathbf{U}^H\bar{\mathbf{U}}\bar{\mathbf{U}}^H\mathbf{U}\right\} \tag{2}$$

in which κ denotes the concentration parameter and the semi unitary matrix $\bar{\mathbf{U}} \in \mathcal{U}_P^N$ represents the center of distribution. We note that the $p_{\text{CIB}}(\mathbf{U})$ is invariant to post-multiplication of $\mathbf{U}$ by any $P \times P$ unitary matrix $\mathbf{Q}$. This means that p_{CIB} also characterizes a distribution for the subspace represented by the orthogonal projector $\mathbf{U}\mathbf{U}^H$, which is our main interest.

The inclusion of an appropriate Bayesian prior can significantly improve the performance of an estimation process. However, the design of this prior depends on the considered application and comes from relevant physical considerations/models on the system. As an example, in radar processing, a physical model of the response of the background can be found in [14] and will be used to build a prior in the application section V.

C. MMSD estimator

A natural distance between a subspace basis $\mathbf{U}$ and its estimate $\widehat{\mathbf{U}}$ can be defined by the Frobenius norm of the difference between the range spaces spanned respectively by these two basis. Based on this formulation, the MMSD estimator is defined as

$$\widehat{\mathbf{U}}_{\text{MMSD}} = \arg\min_{\widehat{\mathbf{U}}} \mathbb{E}\left\{ \| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2 \right\} \tag{3}$$

where the expectation operator is applied to all parameters involved in the signal model. Based on [4], the MMSD estimator of $\mathbf{U}$ can be reformulated as the P principal eigenvectors of a given matrix $\mathbf{M}(p(\mathbf{U}|\mathbf{Y}))$ as

$$\widehat{\mathbf{U}}_{\text{MMSD}} = \mathcal{P}_P \left\{ \mathbf{M}(p(\mathbf{U}|\mathbf{Y})) \right\} \tag{4}$$

with

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Y})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y}) d\mathbf{U}. \tag{5}$$

The expression of the above matrix depends on the posterior probability distribution of $\mathbf{U}$, which is related to both its prior distribution and the signal model. In the following, we will present a model for data distributed as a CG signal of interest embedded in heterogeneous noise (CG outliers plus white Gaussian noise). We will then propose an algorithm to compute the corresponding MMSD estimator.

III. Proposed estimator

A. Data Model

Let us denote N the size of the data, K the number of samples, and P the rank of signal subspace. $\mathbf{Y} \in \mathbb{C}^{N \times N}$ the data matrix and $\mathbf{U} \in \mathcal{U}_P^N$ is the unknown orthonormal basis of the signal subspace of interest. The data is modeled as a sum of CG sources $\mathbf{s}_k$, white Gaussian noise $\mathbf{n}_k$ and CG outlier $\mathbf{c}_k$. For this model, the samples $\{\mathbf{y}_k\}$ (the columns of $\mathbf{Y}$) reads as

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{c}_k + \mathbf{n}_k, \ \forall k \in [\![1, K]\!] \tag{6}$$

For each of the parameters, we will assume the following distributions:

- The signal $\mathbf{s}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}_P, \tau_k)$ lies in low rank subspace of dimension P which is assumed pre-established[1]. The SVD of the source scatter matrix is assumed to be $\mathbf{\Sigma}_P = \mathbf{U}\mathbf{U}^H$. Notice that, as done in [9], [11], the non-null eigenvalues of $\mathbf{\Sigma}_P$ are assumed to be equal to 1. The hypothesis of eigenvalues equality is a relaxation that is made for tractability purposes but still offers interesting performance in practice [11]. Additionally, any scaling of the signal CM is absorbed in the textures τ_k of the CG distribution.
- The signal subspace orthonormal basis is distributed as $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$.
- The potential outlier is distributed as $\mathbf{c}_k \sim \mathcal{CG}(0, \mathbf{\Sigma}_c, \beta_k)$, with a covariance matrix defined by $\mathbf{\Sigma}_c = \mathbf{I}_N - \mathbf{U}\mathbf{U}^H = \mathbf{U}^{\perp}\mathbf{U}^{\perp H}$, where $\mathbf{U}^{\perp}$ is orthonormal complement of the source subspace basis. This outlier model is interesting in terms of robustness since it corresponds to the less-informative prior, as well as the worst-case outlier in terms of subspace estimation.
- The additive noise $\mathbf{n}_k$ is distributed as $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$. For the sake of clarity the variance σ^2 is assumed pre-estimated, but the generalization to unknown σ^2 is trivial.

These hypotheses lead to $(\mathbf{y}_k | \tau_k, \beta_k, \mathbf{U}) \sim \mathcal{CN}(0, \mathbf{\Sigma}_k)$ with

$$\mathbf{\Sigma}_k = \tau_k \mathbf{U}\mathbf{U}^H + \beta_k \mathbf{U}^{\perp}\mathbf{U}^{\perp H} + \sigma^2 \mathbf{I}_N \tag{7}$$

Then, the pdf of $\mathbf{Y}$ conditional to $\{\tau_k\}, \{\beta_k\}, \mathbf{U}$ reads

$$\begin{aligned} p(\mathbf{Y}|\{\tau_k\}, \{\beta_k\}, \mathbf{U}) &= \prod_{k=1}^{K} p(\mathbf{y}_k | \tau_k, \beta_k, \mathbf{U}) \\ &\propto \prod_{k=1}^{K} \frac{\exp\{-\mathbf{y}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{y}_k\}}{\det(\mathbf{\Sigma}_k)} \end{aligned} \tag{8}$$

Thanks to the Sherman Morisson Woodbury lemma, the expression of $\mathbf{\Sigma}_k^{-1}$ is given by

$$\mathbf{\Sigma}_k^{-1} = \frac{1}{\sigma^2 + \beta_k} \mathbf{I} - \frac{\tau_k - \beta_k}{(\sigma^2 + \tau_k)(\sigma^2 + \beta_k)} \mathbf{U}\mathbf{U}^H \tag{9}$$

From (2) and (8), the posterior probability of $(\mathbf{U}|\mathbf{Y}, \{\tau_k\}, \{\beta_k\})$ is expressed as

$$\begin{aligned} p\left(\mathbf{U}|\mathbf{Y}, \{\tau_k\}, \{\beta_k\}\right) &\propto p\left(\mathbf{Y}|\mathbf{U}, \{\tau_k\}, \{\beta_k\}\right) p_{\text{CIB}}(\mathbf{U}) \\ &\propto \prod_{k=1}^{K} \frac{\exp\{-\mathbf{y}_k^H \mathbf{\Sigma}_k^{-1} \mathbf{y}_k\}}{\det(\mathbf{\Sigma}_k)} p_{\text{CIB}}(\mathbf{U}) \\ &\propto \prod_{k=1}^{K} h_k \exp\{\frac{\tau_k}{\sigma^2(\sigma^2 + \tau_k)} \mathbf{y}_k^H \mathbf{U}\mathbf{U}^H \mathbf{y}_k\} \\ &\propto \left(\prod_{k=1}^{K} h_k\right) \text{etr}\{\mathbf{U}^H \mathbf{M} \mathbf{U}\} \text{etr}\left\{\kappa \mathbf{U}^H \bar{\mathbf{U}}\bar{\mathbf{U}}^H \mathbf{U}\right\} \\ &\propto \left(\prod_{k=1}^{K} h_k\right) \text{etr}\{\mathbf{U}^H (\mathbf{M} + \kappa \bar{\mathbf{U}}\bar{\mathbf{U}}^H) \mathbf{U}\} \end{aligned} \tag{10}$$

with

$$\begin{cases} h_k = \dfrac{1}{(\tau_k + \sigma^2)^P (\beta_k + \sigma^2)^{N-P}} \\ \mathbf{M} = \mathbf{Y} \text{diag}\left(\dfrac{\tau_1 - \beta_1}{(\sigma^2 + \tau_1)(\sigma^2 + \beta_1)}, \ldots, \dfrac{\tau_K - \beta_K}{(\sigma^2 + \tau_K)(\sigma^2 + \beta_K)}\right) \mathbf{Y}^H \end{cases}$$

In the following, we propose an iterative algorithm to compute the MMSD estimator for such model.

[1] The proposed results can be applied using plug-in rank estimates or by integrating physical prior knowledge on this parameter [15] (an example is given in Section V).

B. MMSD Algorithm derivation

According to the data model (6) and the generic expression of the MMSD (4), the subspace estimator of $\mathbf{U}$ is expressed as the solution of the following optimisation problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}},\{\tau_k\},\{\beta_k\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{Y},\mathbf{U}}\left\{\| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2\right\} \\ \text{subject to} \quad & \tau_k \geq 0,\ \beta_k \geq 0 \ \ \forall k \\ & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_P \end{aligned} \tag{11}$$

This problem is hard to solve due to the non-convex constraints and the expectation function. Therefore, we propose a block coordinate descent algorithm that updates sequentially the basis of interest $\mathbf{U}$ and the texture parameters $\{\tau_k\}$ and $\{\beta_k\}$. The algorithm has a guaranteed convergence in terms of objective value and is summed up in the box Algorithm 1.

1) Update of the basis $\mathbf{U}$*:* For fixed texture $\{\tau_k^t\}$ and $\{\beta_k^t\}$, the update of $\mathbf{U}$ is obtained by solving the following problem

$$\begin{aligned} \underset{\widehat{\mathbf{U}}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{Y},\mathbf{U}}\left\{\| \widehat{\mathbf{U}}\widehat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H \|_F^2\right\} \\ \text{subject to} \quad & \widehat{\mathbf{U}}^H\widehat{\mathbf{U}} = \mathbf{I}_P \end{aligned} \tag{12}$$

From (10) and (4), the expression of the updated orthonormal basis $\widehat{\mathbf{U}}^{t+1}$ reads

$$\begin{aligned} \widehat{\mathbf{U}}^{t+1} &= \mathcal{P}_P\left\{\int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y},\{\tau_k^t\},\{\beta_k^t\})d\mathbf{U}\right\} \\ &= \mathcal{P}_P\left\{\int \mathbf{U}\mathbf{U}^H \text{etr}\{\mathbf{U}^H(\mathbf{M}^t + \kappa\bar{\mathbf{U}}\bar{\mathbf{U}}^H)\mathbf{U}\}d\mathbf{U}\right\} \end{aligned} \tag{13}$$

with

$$\mathbf{M}^t = \mathbf{Y}\text{diag}\left(\left\{\frac{\tau_k^t - \beta_k^t}{(\sigma^2+\tau_k^t)(\sigma^2+\beta_k^t)}\right\}\right)\mathbf{Y}^H \tag{14}$$

Using Proposition 1 of [4] extended to the complex case, we can obtain a closed form solution of the update as

$$\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_P\left\{\mathbf{M}^t + \kappa\bar{\mathbf{U}}\bar{\mathbf{U}}^H\right\} \tag{15}$$

This update has the same complexity as the deterministic case [11]. It does not require a Gibbs sampling step [4].

2) Update of textures $\{\tau_k\}$ *and* $\{\beta_k\}$*:* For fixed basis $\widehat{\mathbf{U}}^{t+1}$, the update of texture $\{\tau_k\}$ and $\{\beta_k\}$ are the solutions of the problem

$$\begin{aligned} \underset{\{\tau_k\},\{\beta_k\}}{\text{maximize}} \quad & p(\mathbf{Y}|\widehat{\mathbf{U}}^{t+1},\{\tau_k\},\{\beta_k\}) \\ \text{subject to} \quad & \tau_k \geq 0,\ \beta_k \geq 0 \ \forall k \end{aligned} \tag{16}$$

The above problem is equivalent to minimizing the negative log-likelihood as

$$\begin{aligned} \underset{\{\tau_k\},\{\beta_k\}}{\text{minimize}} \quad & \sum_{k=1}^K \ln\left(\det(\boldsymbol{\Sigma}_k)\right) + \mathbf{y}_k^H\boldsymbol{\Sigma}_k^{-1}\mathbf{y}_k \\ \text{subject to} \quad & \tau_k \geq 0,\ \beta_k \geq 0 \ \forall k \end{aligned} \tag{17}$$

The objective function is separable in the τ_k's and the β_k's. Following the same methodology as in [11], the blocks $\{\widehat{\tau}_k^{t+1}\}$ and $\{\widehat{\beta}_k^{t+1}\}$ are updated as

$$\widehat{\tau}_k^{t+1} = \max\left(\frac{\mathbf{y}_k^H\widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H}\mathbf{y}_k}{P} - \sigma^2, 0\right),\ \forall k \tag{18}$$

Algorithm 1: Robust MMSD subspace estimator

input : $\mathbf{Y}$, $\bar{\mathbf{U}}$, σ^2, P, K, N, κ
output : $\widehat{\mathbf{U}}_{\text{MMSD}}$, $\{\tau_k\}$, $\{\beta_k\}$
initialize: $\widehat{\mathbf{U}}^0$, $\{\tau_k^0\}$, $\{\beta_k^0\}$
while *stop criterion unreached* **do**
  Update $\widehat{\mathbf{U}}^{t+1} = \mathcal{P}_P\left\{\kappa\bar{\mathbf{U}}\bar{\mathbf{U}}^H + \mathbf{M}^t\right\}$
  Update $\widehat{\tau}_k^{t+1} = \max\left(\frac{\mathbf{y}_k^H\widehat{\mathbf{U}}^{t+1}\widehat{\mathbf{U}}^{t+1^H}\mathbf{y}_k}{P} - \sigma^2, 0\right)$
  Update $\widehat{\beta}_k^{t+1} = \max\left(\frac{\mathbf{y}_k^H\widehat{\mathbf{U}}^{\perp^{t+1}}\widehat{\mathbf{U}}^{\perp^{t+1^H}}\mathbf{y}_k}{N-P} - \sigma^2, 0\right)$
end

---

and

$$\widehat{\beta}_k^{t+1} = \max\left(\frac{\mathbf{y}_k^H\widehat{\mathbf{U}}^{\perp^{t+1}}\widehat{\mathbf{U}}^{\perp^{t+1^H}}\mathbf{y}_k}{N-P} - \sigma^2, 0\right),\ \forall k \tag{19}$$

In the following, we illustrate the interest of the proposed approach through numerical simulations and STAP detection application.

IV. Numerical simulations

A. Setup

The performance of the proposed estimator is illustrated through Monte-Carlo simulations. We evaluate the average fraction of energy (AFE) of a given estimator $\widehat{\mathbf{U}}$ as

$$\text{AFE}(\widehat{\mathbf{U}}) = \mathbb{E}\{\text{Tr}\{\mathbf{U}^H\widehat{\mathbf{U}}\widehat{\mathbf{U}}^H\mathbf{U}\}\}/P \tag{20}$$

The AFE is considered since it evaluates the closeness of the true range space $\mathbf{U}\mathbf{U}^H$ towards its estimate $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^H$.

The data matrix $\mathbf{Y}$ is generated according to the model in (6). The white Gaussian noise is generated as $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$. The signal follows a CG distribution as in (1) with $\mathbf{s}_k \overset{d}{=} \sqrt{\tau_k}\mathbf{g}_k\ \forall k$, $\tau_k \sim \Gamma(\nu, \frac{1}{\nu})$, $\mathbf{g}_k \sim \mathcal{CN}(0, \alpha\mathbf{U}\mathbf{U}^H)$, and signal to noise ratio SNR=$\log(\frac{\alpha}{\sigma^2})$. When outliers are present in a sample, they are generated as $\mathbf{c}_k \overset{d}{=} \sqrt{\beta_k}\mathbf{d}_k\ \forall k$, $\beta_k \sim \Gamma(\nu', \frac{1}{\nu'})$, $\mathbf{d}_k \sim \mathcal{CN}(0, \gamma(\mathbf{I} - \mathbf{U}\mathbf{U}^H))$, and outlier to noise ratio ONR=$\log(\frac{\gamma}{\sigma^2})$. The signal subspace basis follows $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, where κ denotes the concentration parameter and the center of distribution $\bar{\mathbf{U}}$ is randomly chosen.

B. List of estimators

We compare the following estimators:

- $\widehat{\mathbf{U}}_{\text{SCM}}$: the estimator built from the SVD of the SCM.
- $\widehat{\mathbf{U}}_{\text{MLE}}$: the estimator from [11], corresponding to the MLE estimator for the considered context, while assuming no prior distribution on the orthonormal basis $\mathbf{U}$.
- $\widehat{\mathbf{U}}_{\text{MMSD-U}}$: the MMSD estimator for uniformly distributed sources [4].
- $\widehat{\mathbf{U}}_{\text{MMSD-G}}$: the MMSD estimator in presence of Gaussian sources [5].
- $\widehat{\mathbf{U}}_{\text{MMSD-CG}}$: the proposed MMSD estimator detailed in the box Algorithm 1.
- $\bar{\mathbf{U}}$ is the center of distribution on $\mathbf{U}$ (non adaptive estimator).

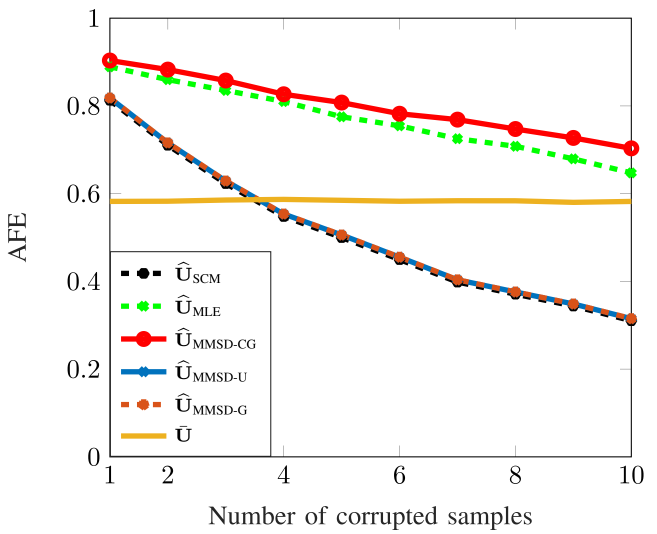

Fig. 1. AFE w.r.t. number of corrupted samples for N=30, K=20, P=5, $\nu = \nu'$=1, ONR=SNR=15dB, $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, $\kappa = 60$.

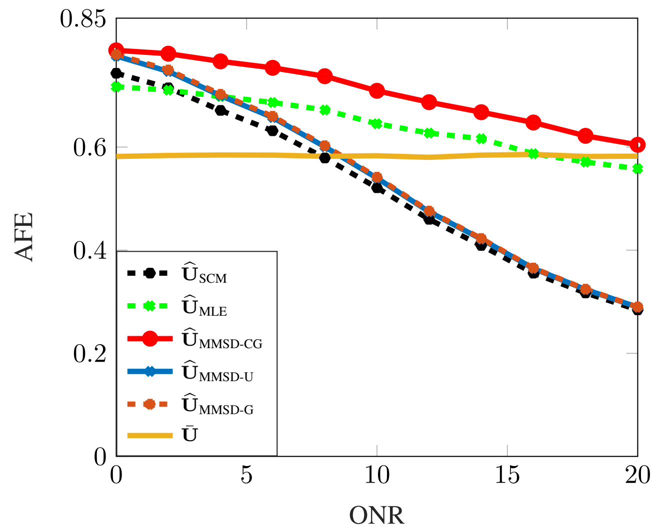

Fig. 2. AFE w.r.t. ONR (in dB) for N=30, K=20, P=5, ν=ν'=1, SNR=10dB, $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$, $\kappa = 60$.

C. Results

Figure 1 displays the AFE of the different estimators w.r.t. the number of corrupted samples where we distinguish two types of samples: the corrupted ones, generated as $\mathbf{y}_k = \mathbf{s}_k + \mathbf{c}_k + \mathbf{n}_k$, and the non corrupted ones, generated as $\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k$. In this context, the non-robust estimators $\widehat{\mathbf{U}}_{\text{SCM}}$, $\widehat{\mathbf{U}}_{\text{MMSD-U}}$, and $\widehat{\mathbf{U}}_{\text{MMSD-G}}$ exhibit poor performances due to the presence of outliers. Conversely, the estimators $\widehat{\mathbf{U}}_{\text{MLE}}$ and $\widehat{\mathbf{U}}_{\text{MMSD-CG}}$, show a better resistance to sample corruptions since they account for outliers in their derivation. The proposed estimator $\widehat{\mathbf{U}}_{\text{MMSD-CG}}$ reaches a better AFE than $\widehat{\mathbf{U}}_{\text{MLE}}$ thanks to the inclusion of the prior information.

Figure 2 displays the AFE in function of ONR where the data are generated as: $\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k$, $\forall k \in [\![1, K-1]\!]$ and only the last sample contains an outlier as $\mathbf{y}_K = \mathbf{s}_K + \mathbf{c}_K + \mathbf{n}_K$. In this scenario, the same conclusion as in Figure 1 can be drawn. Interestingly, the proposed estimator $\widehat{\mathbf{U}}_{\text{MMSD-CG}}$ can exploit both the Bayesian prior information and the outlier modeling to resist high ONR contexts.

V. Application

A. STAP detection

STAP is a technique used in airborne radar in order to detect moving targets embedded in interference background [14]. Under $\mathcal{H}_1$ (a target is present) the $K+1$ available samples are assumed to follow the model

$$\mathcal{H}_1 : \begin{cases} \mathbf{z}_0 = \mathbf{d} + \mathbf{c}_0 + \mathbf{n}_0, \\ \mathbf{z}_k = \mathbf{c}_k + \mathbf{n}_k, \ \forall k \in [\![1, K]\!] \end{cases} \tag{21}$$

For each sample $k \in [\![0, K]\!]$, we observe the contribution of an additive noise, which is modeled as a sum of low rank clutter $\mathbf{c}_k$ embedded in white Gaussian noise $\mathbf{n}_k$. The sample $\mathbf{z}_0$ is the tested cell which may contain a potential moving target $\mathbf{d} = \alpha \mathbf{p}$ with α is the amplitude and $\mathbf{p}$ is the steering vector. In STAP, the steering vector $\mathbf{p}$ is parameterized by the target speed and its DoA. Usually, a detection test is performed over a grid (speed$\times$DoA) in order to test for every possibility. In general, the secondary data $\mathbf{z}_k$, $\forall k \in [\![1, K]\!]$ are assumed to be i.i.d., signal outlier free, and are used in the derivation of an adaptive detection process. However, we will test the robustness of this process to corruption by targets in the following.

B. Experimental setup

The data is provided by the French agency DGA/MI and described in [16] (in french). We note that the ground clutter response $\mathbf{c}_k$ lies in low dimensional subspace spanned by the orthonormal basis $\mathbf{U}_c$ of rank P, that can be evaluated thanks to the Bernnan rule [17]. A common detection procedure in this context is the use of the low rank adaptive normalized matched filter (LR-ANMF):

$$\widehat{\Lambda}(\widehat{\mathbf{\Pi}}_c) = \frac{|\mathbf{p}^H \widehat{\mathbf{\Pi}}_c^\perp \mathbf{z}_0|^2}{|\mathbf{p}^H \widehat{\mathbf{\Pi}}_c^\perp \mathbf{p}||\mathbf{z}_0^H \widehat{\mathbf{\Pi}}_c^\perp \mathbf{z}_0|} \tag{22}$$

where $\widehat{\mathbf{\Pi}}_c = \widehat{\mathbf{U}}_c \widehat{\mathbf{U}}_c^H$ denotes the clutter subspace projector (CSP) estimated from the secondary data, and $\widehat{\mathbf{\Pi}}_c^\perp$ is its orthogonal complement. From (22), we note that the performance detection of the LR-ANMF depends on accuracy of the clutter subspace estimation step. We will compare the following detectors:

- $\widehat{\Lambda}_{\text{SFPE}}$ where the CSP is built from the SVD regularized Tyler's estimator [18]. The regularization parameter is chosen so that the best visible results are obtained.
- $\widehat{\Lambda}_{\text{MLE}}$ where the CSP is built from the robust estimator proposed in [11].
- $\widehat{\Lambda}_{\text{MMSD}}$ where the CSP is built from the proposed MMSD subspace estimator. The prior $\bar{\mathbf{U}}$ of p_{CIB} is computed from

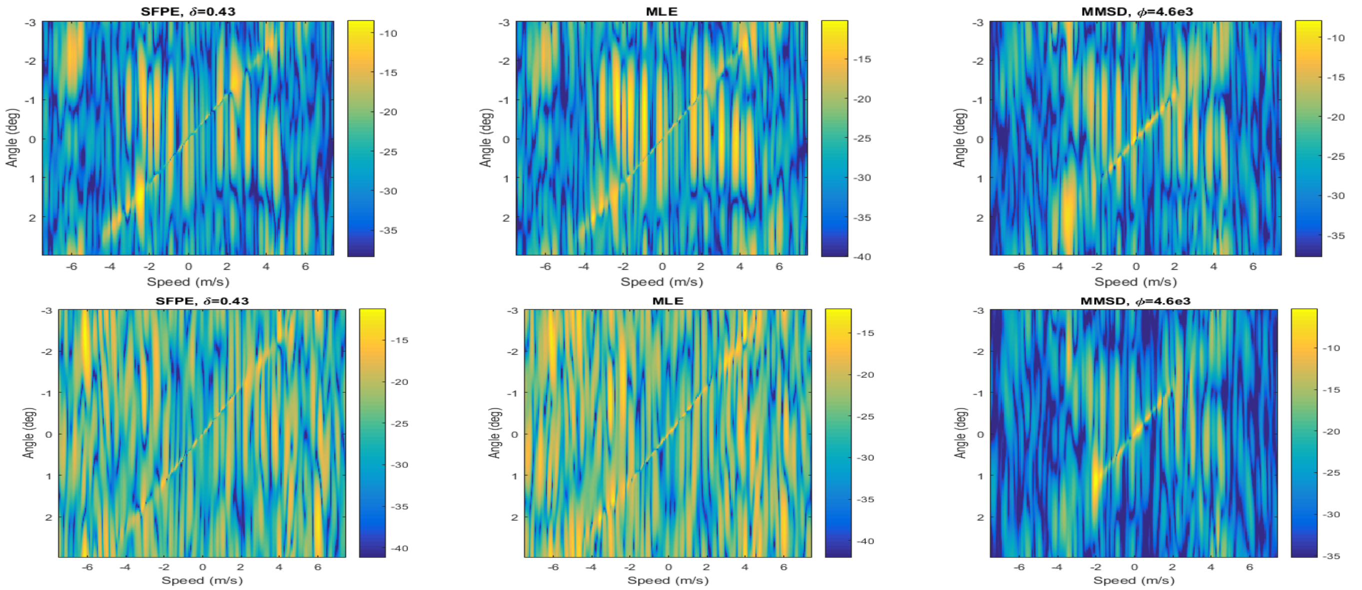

Fig. 3. Output of the detectors for 1 (top) and 2 (bottom) corrupted sample(s) in the secondary data set. $K \approx 3P$ so $K < N$, $P = 46$, $N = 256$.

the SVD of the STAP covariance matrix model from [14]. The concentration parameter κ is adjusted manually since the latter is unknown (adaptive estimation procedures will be considered in future works). It is also worth mentioning that this detector offers a trade-off between the MLE detector $\mathbf{\Lambda}_{\text{MLE}}$ (for $\kappa = 0$) and the "prior only" detector $\mathbf{\Lambda}_{\text{prior}}(\bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ (for $\kappa \to \infty$).

In our setup, the tested cell $\mathbf{z}_0$ contains 10 synthetic targets that can be detected by applying the LR-ANMF over a grid (speed$\times$DoA). We will focus on a challenging scenario, where few secondary data is available ($K \approx 3P < N$), and where some of them are corrupted by the presence of the same targets as in $\mathbf{z}_0$. We test the robustness to this corruption i.e., if targets are still visible in the detector's output.

C. Results

Figure 3 displays the output of the given detectors in presence of 1 corrupted samples and 2 corrupted samples. All of the tested detectors appear robust to a single corruption, even with a challengingly low sample support. The use of the physical prior in $\widehat{\mathbf{\Lambda}}_{\text{MMSD}}$ appear to reduce (at least visually) the false alarms in some areas. Most interestingly, the proposed $\widehat{\mathbf{\Lambda}}_{\text{MMSD}}$ allows for interference rejection and target detection when more than one sample is corrupted, thanks to a robust formulation that also includes prior knowledge.

References

[1] I. Jolliffe, *Principal component analysis*, Springer, 2011.

[2] R. Grover, D. A. Pados, and M. J. Medley, "Subspace direction finding with an auxiliary-vector basis," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 758–763, 2007.

[3] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Academic Press Library in Signal Processing*, vol. 3, pp. 651–717. Elsevier, 2014.

[4] O. Besson, N. Dobigeon, and J. Y. Tourneret, "Minimum mean square distance estimation of a subspace," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5709–5720, 2011.

[5] R. Ben Abdallah, A. Breloy, M. N. El Korso, D. Lautru, and H. H. Ouslimani, "Minimum mean square distance estimation of subspaces in presence of gaussian sources with application to STAP detection," *Journal of Physics: Conference Series, IOP Publishing*, vol. 904, no. 1, pp. 012010, 2017.

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

[7] O. Esa, D. E. Tyler, V. Koivunen, and H. P. Vincent, "Compound gaussian clutter modeling with an inverse gaussian texture distribution," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 876–879, 2012.

[8] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2008.

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

[10] 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," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1986–1998, 2016.

[11] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, D. P. Palomar, and F. Pascal, "A robust signal subspace estimator," *Statistical Signal Processing Workshop (SSP),IEEE*, pp. 1–4, 2016.

[12] K. V. Mardia and P. E. Jupp, "Distributions on spheres," *Directional statistics*, vol. 898, pp. 182, 2000.

[13] A. Breloy, L. Le Magoarou, G. Ginolhac, F. Pascal, and P. Forster, "Maximum likelihood estimation of clutter subspace in non homogeneous noise context," *Signal Processing Conference (EUSIPCO), Proceedings of the 21st European*, pp. 1–5, 2013.

[14] J. Ward, "Space time adaptive processing for airborne RADAR," Tech. Rep., MIT, Lexington, Mass., USA, December 1994.

[15] N. A. Goodman and J. M. Stiles, "On clutter rank observed by arbitrary arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 178–186, Jan 2007.

[16] J. P. Ovarlez, F. Le Chevalier, and S. Bidon, "Les données de club STAP, introduction au STAP," *Traitement de signal*, vol. 28, 2011.

[17] L. E. Brennan and F.M. Staudaher, "Subclutter visibility demonstration," Tech. Rep., 1992.

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