



# Bayesian Signal Subspace Estimation with Compound Gaussian Sources

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

## ► To cite this version:

R. Ben Abdallah, Arnaud Breloy, M. N. El Korso, David Lautru. Bayesian Signal Subspace Estimation with Compound Gaussian Sources. Signal Processing, 2019, 167, pp.107310. 10.1016/j.sigpro.2019.107310 . hal-02305119

**HAL Id: hal-02305119**

**<https://hal.science/hal-02305119>**

Submitted on 31 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayesian signal subspace estimation with compound Gaussian sources

R. Ben Abdallah\*, A. Breloy, M.N. El Korso, D. Lautru

Paris Nanterre University, LEME EA-4416, Ville d'Avray 92410, France

## A B S T R A C T

### Keywords:

Array processing  
Subspace estimation  
Bayesian estimation  
Minimum mean square distance  
Maximum a posteriori  
Bingham distribution  
Langevin distribution  
Compound Gaussian distribution  
Majorization-minimization

In this paper, we consider the problem of low dimensional signal subspace estimation in a Bayesian context. We focus on compound Gaussian signals embedded in white Gaussian noise, which is a realistic modeling for various array processing applications. Following the Bayesian framework, we derive two algorithms to compute the maximum a posteriori (MAP) estimator and the so-called minimum mean square distance (MMSD) estimator, which minimizes the average natural distance between the true range space of interest and its estimate. Such approaches have shown their interests for signal subspace estimation in the small sample support and/or low signal to noise ratio contexts. As a byproduct, we also introduce a generalized version of the complex Bingham Langevin distribution in order to model the prior on the subspace orthonormal basis. Finally, numerical simulations illustrate the performance of the proposed algorithms.

## 1. Introduction

Subspace estimation is an ubiquitous problem in signal processing, as it is often required to infer the low-dimensional space where information lies in. It is considered as the cornerstone of a plethora of applications and algorithms such as PCA [1], DoA estimation [2], interference cancellation [3,4], reduced rank adaptive filtering [5] and signal detection [6]. Nevertheless, the subspace estimation problem becomes a challenging problem in the presence of non-standard conditions such as low sample support, low signal to noise ratio (SNR), non-Gaussian observations or presence of outliers in the training set.

Most commonly, the signal subspace is estimated through the strongest eigenvectors of the eigenvalue decomposition (EVD) of the sample covariance matrix (SCM). This corresponds to the maximum likelihood estimator (MLE) for the classical model with additive white Gaussian noise. This estimator provides an accurate estimator for high SNR and/or for large number of samples. Nevertheless, it shows its limits outside these asymptotic regimes. This estimator is also known to be sensitive to missmodeling, e.g., presence of outliers or non-Gaussian observations. A possible solution to ensure better performance in these contexts is to incorporate a prior knowledge into the estimation process. In a Bayesian context, a prior distribution of the subspace orthonormal basis can

be assumed in order to overcome the aforementioned drawbacks. This approach yields estimators such as the maximum a posterior (MAP) [7] or the minimum mean square distance (MMSD), which minimizes the expected distance between the true projection matrix and its estimate [8,9]. This is an intuitively appealing method, as it is based on a natural metric in the complex Grassmann manifold [10], i.e., the set of  $P$ -dimensional subspaces in  $\mathbb{C}^N$  (where  $P$  is the rank of subspace and  $N$  denotes the dimension of the observation space). In the context of subspace estimation, the MMSD has been introduced in [8,9]. More specifically, [9] derives a practical formulation of the MMSD estimators when the subspace of interest is parameterized by its orthonormal basis. This formulation is then used in [9] to propose MMSD estimators for two data models (namely, linear and covariance models) involving an uniform prior for the sources distribution. In [11], these results have been extended to a subspace parameterized by its CS decomposition. In [12] and [13], the authors have recently extended these concepts to Bayesian non-parametric framework in order to adaptively select the rank of the subspace to be estimated.

In this paper, we focus on the context of sources following a compound Gaussian (CG) distribution [14] embedded in white Gaussian noise [15]. This choice is motivated by the fact that the CG distribution has been considered in many modern robust signal processing applications, as it can account for local power fluctuations and presents good agreement to several real data set [16,17]. Note that this family covers a large panel of well known distributions, notably heavy-tailed ones, such as Student

\* Corresponding author.

E-mail address: 36010449@parisnanterre.fr (R.B. Abdallah).

$t$ -,  $K$ -, and Weibull distributions (cf. [16] and references therein). Hence, the considered model can accurately describe clutter (or power-fluctuating sources) plus thermal noise observations, which are common in plethora of signal processing application. As an example, this model has been used for detection in heterogeneous environment [18,19] and for robust structured covariance matrix estimation in [15,20]. Specifically, concerning the subspace estimation problem [21,22], proposed MLE algorithms for this context, and estimation bounds were derived in [23]. However, these studies were never brought to a Bayesian context, in the sense that they did not assume a prior on the subspace of interest.

In this paper, we fill this gap by deriving new Bayesian estimators in the context of CG distributed sources embedded in white Gaussian noise. First, our development requires to extend the distributions used in [9] to the case of data with complex entries. To this aim, we introduce a generalization of the real Bingham–Langevin (also referred to as Bingham–von-Mises–Fisher) distribution and we propose a practical sampling method adapted to the proposed distribution. Second, we develop an algorithm to compute the MAP estimator for the proposed model based on the Majorization–Minimization (MM) algorithm [22]. Third, we derive a Gibbs-sampler based algorithm to compute the MMSD estimator, which follows the framework of [9] (that considered a uniform prior on the sources).

Finally, numerical simulations show that the inclusion of a Bayesian prior on the subspace orthonormal basis can significantly improve the performance of the estimation process. The design of this prior depends, of course, on the considered application and comes from appropriate physical considerations/models, which is out of the scope of this paper.

The paper is organized as follows: Section 2 gives an overview on the background theory. Section 3 presents the data model and problem statement. Then, Section 4 deals with the proposed Bayesian estimators. In Section 4.1.1, numerical simulations are shown in order to assess the performance of our proposed estimators.

The following notations are adopted along this paper: italic indicates a scalar quantity, lower case boldface indicates a vector quantity and upper case boldface a matrix.  $\{w_n\}_{n=1}^N$  denotes the set of elements  $w_n$ , with  $n \in [[1, N]]$ , this writing will be contracted in  $\{w_n\}$  if there is no ambiguity.  $\text{Re}\{\cdot\}$  stands for the real part of a complex number. The superscripts  $H$  and  $T$  denote, respectively, the transpose-conjugate and the transpose operators.  $\text{Tr}\{\cdot\}$  and  $\text{etr}\{\cdot\}$  stand, respectively, for the trace and the exponential of the trace of a given matrix.  $\det(\cdot)$  is the determinant operator.  $\text{diag}(\cdot)$  denotes a diagonal matrix built from a set of elements (or a vector).  $\propto$  stands for “proportional to” and  $\stackrel{d}{=}$  stands for “has the same distribution as”.  $\text{EVD}$  allows to define the EVD of a given matrix, and a similar notation is adopted for the SVD and TSVD (thin-SVD).  $\mathcal{CN}(\mu, \mathbf{\Omega})$  is the complex normal distribution of mean  $\mu$  and covariance matrix  $\mathbf{\Omega}$ .  $\text{unif}(0, 1)$  denotes the continuous uniform distribution at the two boundaries 0 and 1.  $\mathcal{H}_N^{++}$  ( $\mathcal{H}_N^+$ ) is the set of  $N \times N$  positive (semi-)definite Hermitian matrices.  $\mathcal{U}_P^N = \{\mathbf{U} \in \mathbb{C}^{N \times P} | \mathbf{U}^H \mathbf{U} = \mathbf{I}_P\}$  is the set of  $N \times P$  semi-unitary matrices, i.e., tall matrices whose columns form an orthonormal basis.

## 2. Compound Gaussian and complex generalized Bingham Langevin distributions

This section presents the main background theory on which our derivations are based.

### 2.1. Compound Gaussian distribution

The CG distribution is a useful and well established tool in the robust signal processing literature [14]. This model is a versatile

one, as it encloses usual distributions such as Gaussian, Student  $t$ -,  $K$ -, and Weibull distributions. A  $N$ -dimensional CG observation is represented as a product of two statistically independent components. Specifically, if  $\mathbf{s} \in \mathbb{C}^N$  follows a centered CG distribution, denoted  $\mathbf{s} \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, f_\tau)$ , it has the following stochastic representation

$$\mathbf{s} \stackrel{d}{=} \sqrt{\tau} \mathbf{d}, \quad (1)$$

where

- i)  $\tau$  is a positive random scalar, called texture, of probability density function (p.d.f.)  $f_\tau$ . This parameter is statistically independent of  $\mathbf{d}$ . Depending on  $f_\tau$ , we can obtain various standard multivariate distribution for  $\mathbf{s}$  [14]. In order to design algorithm that are robust to these distributions, we consider here this parameter as unknown deterministic for each realization. This distribution will be thus denoted by  $\mathbf{s}_k \sim \mathcal{CG}(\mathbf{0}, \mathbf{\Sigma}, \tau_k)$  for each observation  $k \in [[1, K]]$ . We also denote  $\boldsymbol{\tau}$  the vector that aggregates the parameters  $\{\tau_k\}$  for a given set of observations  $\{\mathbf{s}_k\}$ .
- ii)  $\mathbf{d}$  follows a zero-mean multivariate complex Gaussian distribution of covariance matrix  $\mathbf{\Sigma}$ , denoted,  $\mathbf{d} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ . The parameter  $\mathbf{\Sigma} \in \mathcal{H}_N^+$  is referred to as the scatter matrix. Notice that if  $\mathbb{E}\{\tau\} < \infty$ , the covariance matrix of  $\mathbf{s}$  exists and is proportional to the scatter matrix, i.e.,  $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbb{E}\{\tau\} \mathbf{\Sigma}$ .

For a set of  $K$  independent and identically distributed (i.i.d.) zero-mean observations following a CG distribution, we have the representation  $\mathbf{S} \stackrel{d}{=} \mathbf{D}\mathbf{T}$ , where, the  $k$ -th column of  $\mathbf{D} \in \mathbb{C}^{N \times K}$  follows  $\mathbf{d}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$  and  $\mathbf{T} = \text{diag}(\{\sqrt{\tau_k}\})$ .

### 2.2. Complex generalized Bingham Langevin distribution (CGBL)

In order to model priors for subspaces, we focus in the following on the distribution w.r.t. the set  $\mathcal{U}_P^N$ . Among the most widely used distributions on  $\mathcal{U}_P^N$  are the Bingham and the Langevin distributions [24–26]. We present the CGBL distribution as a generalization of the aforementioned usual directional statistics to the case of matrix variables with complex entries. The CGBL is a probability distribution on the set of semi-unitary matrices which combines linear and quadratic terms that is parametrized by a set of matrices  $\{\mathbf{A}_p\} \subset \mathcal{H}_N^+$  and the matrix  $\mathbf{C}$ . We denote  $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_p\}) \in \mathbb{C}^{N \times P}$  when the p.d.f. of  $\mathbf{U}$  on  $\mathcal{U}_P^N$  reads

$$p_{\text{CGBL}}(\mathbf{U}) \propto \exp \left\{ \sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H \mathbf{A}_p \mathbf{u}_p \right\}, \quad (2)$$

where  $\mathbf{c}_p$  and  $\mathbf{u}_p$  stand for the  $p$ th column vector of, respectively,  $\mathbf{C}$  and  $\mathbf{U}$ .

**Remark 1.** From (2),  $p_{\text{CGBL}}$  promotes the concentration of each vector  $\mathbf{u}_p$  around  $\mathbf{c}_p$  and each range space  $\mathbf{u}_p \mathbf{u}_p^H$  around the subspace associated to the strongest eigenvalues of the Hermitian matrix  $\mathbf{A}_p$ . Typically, if  $\mathbf{A}_p = \mathbf{A}$  and  $\mathbf{c}_p = \mathbf{0}$ ,  $\forall p \in [[1, P]]$ , the range space  $\mathbf{U}\mathbf{U}^H$  tends to be close to the dominant space of  $\mathbf{A}$ .

Moreover, an efficient way to sample from this distribution is described in Appendix A. Finally, Table 1 lists special cases of the CGBL that correspond to standard distributions extended to the complex case.

**Remark 2.** The estimation of the Bingham and Langevin distributions parameters from a set of observations  $\{\mathbf{U}_k\}$  is well investigated theoretically in [27] and references therein for the real case. In our context we rather aim at recovering the signal subspace basis  $\mathbf{U}$  from a matrix of noisy observations  $\mathbf{Y}$ . The parameters of the CGBL distribution thus gather the available prior information and we do not address their estimation.

**Table 1**  
Special cases of the CGBL distribution.

Complex distribution	Parameters	Probability density function
Bingham–Langevin CBL( $\mathbf{C}, \Phi, \mathbf{A}$ )	$\Phi = \text{diag}(\{\phi_p\})$ , $\mathbf{A} \in \mathcal{H}_N^+$ , $\mathbf{A}_p = \phi_p \mathbf{A}$ , $\mathbf{C} \in \mathbb{C}^{N \times P}$	$p_{\text{CBL}} \propto \exp\{\sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H \mathbf{A}_p \mathbf{u}_p\}$
Bingham CB( $\Phi, \mathbf{A}$ )	$\Phi = \text{diag}(\{\phi_p\})$ , $\mathbf{A} \in \mathcal{H}_N^+$ $\mathbf{c}_p = \mathbf{0}$ , $\mathbf{A}_p = \phi_p \mathbf{A}$	$p_{\text{CB}} \propto \exp\{\sum_{p=1}^P \mathbf{u}_p^H \mathbf{A}_p \mathbf{u}_p\}$
Invariant Bingham CIB( $\kappa, \mathbf{A}$ )	$\kappa \in \mathbb{R}^+$ , $\mathbf{A} \in \mathcal{H}_N^+$ $\mathbf{A}_p = \kappa \mathbf{A}$	$p_{\text{CIB}} \propto \text{etr}\{\kappa \mathbf{U}^H \mathbf{A} \mathbf{U}\}$
Langevin CL( $\mathbf{C}$ )	$\mathbf{C} \in \mathbb{C}^{M \times R}$ , $\mathbf{A}_p = \mathbf{0}$	$p_{\text{CL}} \propto \text{etr}\{\text{Re}\{\mathbf{C}^H \mathbf{U}\}\}$

### 3. Data model

Along this paper,  $N$  denotes the size of the data,  $K$  represents the number of samples, and  $P$  is the rank of the signal subspace ( $P < N$ ). We denote by  $\mathbf{Y} \in \mathbb{C}^{N \times K}$  the data matrix,  $\mathbf{U} \in \mathcal{U}_P^N$  an unknown orthonormal basis of the signal subspace,  $\mathbf{S} \in \mathbb{C}^{P \times K}$  the matrix containing the signal of interest and  $\mathbf{N} \in \mathbb{C}^{N \times K}$  the additive noise. The conditional probability of  $\mathbf{Y}$  given  $\mathbf{U}$  is denoted by  $p(\mathbf{Y}|\mathbf{U})$  and  $\mathbb{E}_{\mathbf{U}, \mathbf{Y}}\{\cdot\}$  denotes the expectation operator applied on both  $\mathbf{U}$  and  $\mathbf{Y}$ .

The data is modeled as a sum of low-rank CG sources embedded in white Gaussian noise. This formulations is useful to model clutter (or power-fluctuating sources) plus thermal noise in several array processing applications, such as RADAR [18,19,21–23]. For this model, the samples  $\{\mathbf{y}_k\}_{k=1}^K$  (the columns of  $\mathbf{Y}$ ) are drawn as:

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{n}_k \quad (3)$$

where

- $\mathbf{s}_k \sim \mathcal{CG}(\mathbf{0}, \Sigma, \tau_k)$  are the low rank CG distributed sources. The rank  $P$  is assumed pre-established<sup>1</sup>. Moreover, the source scatter matrix is parameterized by its low-rank EVD as

$$\Sigma \stackrel{\text{EVD}}{=} \mathbf{U} \mathbf{A} \mathbf{U}^H \quad (4)$$

In addition,

- i)  $\{\tau_k\}$  are the CG textures assumed to be positive unknown deterministic.
- ii)  $\mathbf{A} = \text{diag}(\{\lambda_p\}) \in \mathbb{R}^{P \times P}$  is the diagonal matrix containing the scatter matrix eigenvalues, which are assumed to be positive unknown deterministic.
- iii)  $\mathbf{U} \in \mathcal{U}_P^N$  are the eigenvectors of the scatter matrix, whose columns spans the signal subspace basis. This basis follows the distribution  $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_p\})$ .
- $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$  is an additive white Gaussian noise of known or pre-estimated variance  $\sigma^2$ .

The data matrix can be therefore written as

$$\mathbf{Y} = \mathbf{U} \tilde{\mathbf{S}} \mathbf{T} + \mathbf{N} \quad (5)$$

with the columns of  $\tilde{\mathbf{S}} \in \mathbb{C}^{P \times K}$  distributed as  $\tilde{\mathbf{s}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{A})$  and  $\mathbf{T} = \text{diag}(\{\sqrt{\tau_k}\}) \in \mathbb{R}^{P \times P}$  is the diagonal matrix. The latter reads as a modified linear model, with unknown power fluctuations for each sample gathered in the matrix  $\mathbf{T}$ .

**Remark 3.** In this paper, we consider the hybrid Bayesian model because our main interest is incorporating a prior knowledge on the signal subspace in the estimation process. Conversely, we choose not to specify the p.d.f. of the texture parameters  $\{\tau_k\}$  (and the eigenvalues  $\{\lambda_p\}$ ) which are assumed unknown and deterministic. By doing so, we ensure more robustness to any prior mis-

match w.r.t. these parameters. Moreover, this assumption also allows for computational tractability since including a prior distribution on  $\{\tau_k\}$  in the considered model leads to integral functions that are complex to handle [32]. In the following, for sake of conciseness and with an abuse of language, the ML-MMSD hybrid Bayesian estimator (respectively ML-MAP) will be simply referred to as MMSD (respectively MAP).

By denoting

$$\Sigma_k = \tau_k \mathbf{U} \mathbf{A} \mathbf{U}^H + \sigma^2 \mathbf{I}_N \quad \forall k \in \llbracket 1, K \rrbracket \quad (6)$$

we have for each sample the conditional representation  $(\mathbf{y}_k | \mathbf{U}, \mathbf{A}, \tau_k) \sim \mathcal{CN}(\mathbf{0}, \Sigma_k)$ , leading to the conditional p.d.f. of the sample set  $\mathbf{Y}$  as

$$p(\mathbf{Y} | \mathbf{U}, \{\lambda_p\}, \{\tau_k\}) = \prod_{k=1}^K p(\mathbf{y}_k | \mathbf{U}, \{\lambda_p\}, \tau_k) \propto \prod_{k=1}^K \frac{\exp\{-\mathbf{y}_k^H \Sigma_k^{-1} \mathbf{y}_k\}}{\det(\Sigma_k)} \quad (7)$$

Thanks to the Sherman Morrison Woodbury lemma, the expression of  $\Sigma_k^{-1}$  is simplified as  $\Sigma_k^{-1} = \sigma^{-2} \mathbf{I} - \mathbf{U} \Gamma_k \mathbf{U}^H$ , where  $\Gamma_k = \sigma^{-2} \mathbf{I}_P - (\tau_k \mathbf{A} + \sigma^2 \mathbf{I}_P)^{-1}$  is a diagonal matrix of entries

$$[\Gamma_k]_{p,p} = \gamma_{k,p} = \frac{\tau_k \lambda_p}{\sigma^2 (\tau_k \lambda_p + \sigma^2)} \quad (8)$$

From (2) and (7), some manipulations allow to the posterior probability of  $\mathbf{U}$  to be rewritten as

$$\begin{aligned} p(\mathbf{U} | \mathbf{Y}, \{\tau_k\}, \{\lambda_p\}) & \quad (9) \\ & \propto p(\mathbf{Y} | \mathbf{U}, \{\tau_k\}, \{\lambda_p\}) p_{\text{CGBL}}(\mathbf{U}) \propto \prod_{k=1}^K \frac{\exp\{-\mathbf{y}_k^H (\Sigma_k)^{-1} \mathbf{y}_k\}}{\det(\Sigma_k)} p_{\text{CGBL}}(\mathbf{U}) \\ & \propto \prod_{k=1}^K \left( \prod_{p=1}^P \frac{1}{\tau_k \lambda_p + \sigma^2} \right) \exp\{-\mathbf{y}_k^H (-\mathbf{U} \Gamma_k \mathbf{U}^H + \sigma^{-2} \mathbf{I}_N) \mathbf{y}_k\} p_{\text{CGBL}}(\mathbf{U}) \\ & \propto \prod_{k=1}^K \left( \prod_{p=1}^P \frac{1}{\tau_k \lambda_p + \sigma^2} \right) \exp\left\{ \sum_{k=1}^K \mathbf{y}_k^H \mathbf{U} \Gamma_k \mathbf{U}^H \mathbf{y}_k \right\} p_{\text{CGBL}}(\mathbf{U}) \\ & \propto \left( \prod_{k=1}^K \prod_{p=1}^P \frac{1}{\tau_k \lambda_p + \sigma^2} \right) \exp\left\{ \sum_{p=1}^P \mathbf{u}_p^H \mathbf{M}_p \mathbf{u}_p \right\} p_{\text{CGBL}}(\mathbf{U}) \\ & \propto \left( \prod_{k=1}^K \prod_{p=1}^P \frac{1}{\tau_k \lambda_p + \sigma^2} \right) \exp\left\{ \sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H [\mathbf{A}_p + \mathbf{M}_p] \mathbf{u}_p \right\} \end{aligned} \quad (9)$$

with

$$\mathbf{M}_p = \sum_{k=1}^K \gamma_{k,p} \mathbf{y}_k \mathbf{y}_k^H \quad (10)$$

where  $\gamma_{k,p}$  is given in (8).

<sup>1</sup> Indeed, the proposed results can still be applied using plug-in rank estimates or by integrating physical prior knowledge on this parameter [28]. About rank estimation, the reader is referred to the overview [29] and recent methods using shrinkage [30] or random matrix theory [31].

#### 4. Bayesian subspace estimators

In this section, we aim to develop Bayesian estimators of the subspace orthonormal basis  $\mathbf{U}$  according to the data model (5). The first proposal is a MM algorithm to compute the MAP estimator. The second is an algorithm to evaluate the MMSD through MM iterations and a Gibbs sampling scheme. Additionally, we present a special case, referred to as “simplified model”, for which the MAP and the MMSD estimators coincide, and can be obtained through closed form updates. Considering these approaches, the properties of each method are listed below:

- Theoretically, the MMSD approach offers best performance in terms of expected distance between the estimated and true signal subspace projection matrices. Nevertheless, the computation of the MMSD estimator usually requires a Gibbs sampler scheme which can be computationally expensive.
- The MAP is theoretically sub-optimal (compared to the MMSD), but can generally reach good performance in practice. Moreover, the proposed algorithm to compute this estimator only involves closed form updates, which significantly reduces the computational time.
- The MMSD for the simplified is interesting because it does not require a Gibbs sampling scheme to be computed. The assumed simplification is not necessarily realistic and introduces a mismatch w.r.t. the true model, however, numerical simulations will illustrate the interest of the approach.

##### 4.1. The subspace MAP estimator

In this section, we derive a subspace MAP estimator based on the data model (3) that maximizes the posterior probability. It reads as the solution of

$$\begin{aligned} & \underset{\hat{\mathbf{U}}, \{\tau_k\}, \{\lambda_p\}}{\text{maximize}} && p(\hat{\mathbf{U}}|\mathbf{Y}, \{\tau_k\}, \{\lambda_p\}) \\ & \text{subject to} && \tau_k \geq 0 \ \forall k, \ \lambda_p \geq 0 \ \forall p \\ & && \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_p \end{aligned} \quad (11)$$

From (9) and (11), this problem can be recasted as

$$\begin{aligned} & \underset{\{\hat{\mathbf{u}}_p\}, \{\tau_k\}, \{\lambda_p\}}{\text{maximize}} && \sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \hat{\mathbf{u}}_p\} + \hat{\mathbf{u}}_p^H [\mathbf{A}_p + \mathbf{M}_p] \hat{\mathbf{u}}_p \\ & \text{subject to} && -\sum_{k=1}^K \ln(\tau_k \lambda_p + \sigma^2) \\ & && \tau_k \geq 0 \ \forall k, \ \lambda_p \geq 0 \ \forall p \\ & && \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_p \ \text{with} \ \hat{\mathbf{U}} = [\hat{\mathbf{u}}_1 | \dots | \hat{\mathbf{u}}_p] \\ & && \mathbf{M}_p = \sum_{k=1}^K \frac{\tau_k \lambda_p}{\sigma^2 (\tau_k \lambda_p + \sigma^2)} \mathbf{y}_k \mathbf{y}_k^H \end{aligned} \quad (12)$$

To solve this problem, we derive an iterative based MM algorithm that sequentially updates the variables  $\hat{\mathbf{U}}^t, \{\tau_k^t\}, \{\lambda_p^t\}$  at the  $t$ th iteration. The MM algorithm performs, at each iteration, an update of the variables by minimizing a surrogate function of the objective. This process decreases the value of the objective function at each step. We adapt here the surrogates function of [22] to our context, which leads to closed form updates of the parameters. The resulting algorithm is summed up in the box Algorithm 1. A brief explanation of the derivations is given below.

##### 4.1.1. Algorithm derivation

First, the variables  $\hat{\mathbf{U}}, \{\lambda_p\}$ , and  $\{\tau_k\}$  are initialized. This initialization can, for example, be taken from the  $P$  strongest eigenvectors and eigenvalues of the SCM for  $\hat{\mathbf{U}}$  and  $\{\lambda_p\}$ .

- Update of the basis  $\hat{\mathbf{U}}$ :

By fixing  $\{\lambda_p^t\}, \{\tau_k^t\}$ , the update of the basis of interest  $\hat{\mathbf{U}}^{t+1}$  is obtained by solving

$$\begin{aligned} & \underset{\{\hat{\mathbf{u}}_p\}}{\text{maximize}} && \sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \hat{\mathbf{u}}_p\} + \hat{\mathbf{u}}_p^H [\mathbf{A}_p + \mathbf{M}_p^t] \hat{\mathbf{u}}_p \\ & \text{subject to} && \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_p \ \text{with} \ \hat{\mathbf{U}} = [\hat{\mathbf{u}}_1 | \dots | \hat{\mathbf{u}}_p] \end{aligned} \quad (13)$$

---

##### Algorithm 1: MAP for the general model (3) $\mathbf{U}_{\text{MAP}}$ .

---

**input** :  $\mathbf{Y}, \mathbf{C}, \{\mathbf{A}_p\}, \sigma^2, P, K, N$   
**output** : MAP estimators of  $\mathbf{U}, \{\tau_k\}, \{\lambda_p\}$   
**initialize**:  $\hat{\mathbf{U}}^0, \{\tau_k^0\}$  and  $\{\lambda_p^0\}$   
**1 for**  $t = 0 \dots T - 1$  **do**  
**2** | Update  $\hat{\mathbf{U}}^{t+1} = \mathcal{P}_{\text{Proc}}(\mathbf{H}^t)$ , with  $\mathbf{H}^t$  in (15)  
**3** | Update  $\tau_k^{t+1} \ \forall k$  with (18)  
**4** | Update  $\lambda_p^{t+1} \ \forall p$  with (20)  
**5 end**

---

with  $\mathbf{M}_p^t = \sum_{k=1}^K \gamma_{k,p}^t \mathbf{y}_k \mathbf{y}_k^H$  and  $\gamma_{k,p}^t = \frac{\tau_k^t \lambda_p^t}{\sigma^2 (\tau_k^t \lambda_p^t + \sigma^2)}$ . This problem has not a trivial solution due to the semi-unitary constraint. Therefore, we apply the MM procedure in order to obtain closed form updates that improve the value of the objective at each iteration. An update of the orthonormal basis can be obtained thanks to Proposition 1 in Appendix B. This update reads as

$$\hat{\mathbf{U}}^{t+1} = \mathcal{P}_{\text{Proc}}(\mathbf{H}^t) \quad (14)$$

where

$$\mathbf{H}^t = \mathbf{S}^t + 1/2\mathbf{C} \ \text{and} \ \mathbf{S}^t = [(\mathbf{A}_1 + \mathbf{M}_1^t)\mathbf{u}_1^t | \dots | (\mathbf{A}_P + \mathbf{M}_P^t)\mathbf{u}_P^t] \quad (15)$$

with  $\mathbf{u}_p^t$  is the  $p$ th column of the matrix  $\mathbf{U}^t$  and the operator  $\mathcal{P}_{\text{Proc}}$  is the projection onto the Stiefel manifold [33], defined as

$$\begin{aligned} \mathcal{P}_{\text{Proc}} : \quad \mathbb{C}^{N \times P} & \longrightarrow \mathcal{U}_P^N \\ \mathbf{Y} \stackrel{\text{TSVD}}{=} \mathbf{U}\mathbf{D}\mathbf{V}^H & \longmapsto \mathcal{P}_{\text{Proc}}\{\mathbf{Y}\} = \mathbf{U}\mathbf{V}^H \end{aligned} \quad (16)$$

- Update of  $\{\tau_k\}$ :

The optimization problem in (12) w.r.t.  $\{\tau_k\}$  for other fixed variables can be expressed as separable sub-problems in  $\tau_k$  as

$$\begin{aligned} & \underset{\tau_k}{\text{minimize}} && \sum_{p=1}^P \ln(\tau_k \lambda_p^t + \sigma^2) - \frac{\tau_k \lambda_p^t}{\tau_k \lambda_p^t + \sigma^2} z_{k,p}^{t+1} \\ & \text{subject to} && \tau_k \geq 0 \end{aligned} \quad (17)$$

with  $z_{k,p}^{t+1} = \|\mathbf{y}_k^H \mathbf{u}_p^{t+1}\|^2$ . This problem has no direct solution but a closed-form update can be obtained thanks to (66) from Proposition 2 in Appendix B for which we identify  $\tau_k = a$ ,  $\lambda_p^t = b_i$ ,  $P = I$ ,  $s_i = z_{k,p}^{t+1}$  and  $\alpha_{k,p}^t = \theta_i^t$ . Consequently, the update reads

$$\tau_k^{t+1} = \frac{1}{P} \frac{\left( \sum_{p=1}^P z_{k,p}^{t+1} \frac{\tau_k^t \lambda_p^t}{\tau_k^t \lambda_p^t + \sigma^2} \right) \left( \sum_{p=1}^P \sigma^2 \frac{\alpha_{k,p}^t}{\tau_k^t \lambda_p^t + \sigma^2} \right)}{\sum_{p=1}^P \frac{\alpha_{k,p}^t \lambda_p^t}{\tau_k^t \lambda_p^t + \sigma^2}} \quad (18)$$

- Update of  $\{\lambda_p\}$ :

By fixing the remaining variables, the optimization problem (12) w.r.t.  $\{\lambda_p\}$  is equivalent to the optimization of the following sub-problems

$$\begin{aligned} & \underset{\lambda_p}{\text{minimize}} && \sum_{k=1}^K \ln(\tau_k^{t+1} \lambda_p + \sigma^2) - \frac{\tau_k^{t+1} \lambda_p}{\tau_k^{t+1} \lambda_p + \sigma^2} z_{k,p}^{t+1} \\ & \text{subject to} && \lambda_p \geq 0 \end{aligned} \quad (19)$$

Similarly to the update of texture and by using (66), we can apply Proposition 2 in Appendix B with  $\lambda_p = a$ ,  $\tau_k^{t+1} = b_i$ ,  $K = I$  and  $\beta_{k,p}^t = \theta_i$ ,  $z_{k,p}^{t+1} = s_i$ . The updates of  $\lambda_p$  are then given by

$$\lambda_p^{t+1} = \frac{1}{K} \frac{\left( \sum_{k=1}^K z_{k,p}^{t+1} \frac{\tau_k^{t+1} \lambda_p^t}{\tau_k^{t+1} \lambda_p^t + \sigma^2} \right) \left( \sum_{k=1}^K \sigma^2 \frac{\beta_{k,p}^t}{\tau_k^{t+1} \lambda_p^t + \sigma^2} \right)}{\sum_{k=1}^K \frac{\beta_{k,p}^t \tau_k^{t+1}}{\tau_k^{t+1} \lambda_p^t + \sigma^2}}. \quad (20)$$



## 4.2. The subspace MMSD estimator

### 4.2.1. Definition

The MMSD estimator minimizes the average Euclidean distance between the true range space  $\mathcal{R}(\mathbf{U}) = \mathbf{U}\mathbf{U}^H$  and its estimate  $\mathcal{R}(\hat{\mathbf{U}}) = \hat{\mathbf{U}}\hat{\mathbf{U}}^H$ . It corresponds to a natural metric between the subspace spanned by  $\mathbf{U}$  and  $\hat{\mathbf{U}}$  in the complex Grassmann space. Straightforwardly extending the formulation of [9] to the complex case, the MMSD estimator is expressed as

$$\begin{aligned}\hat{\mathbf{U}}_{\text{MMSD}} &= \arg \min_{\hat{\mathbf{U}}} \mathbb{E}_{\mathbf{U}, \mathbf{Y}} \{ \|\hat{\mathbf{U}}\hat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H\|_F^2 \} \\ &= \arg \max_{\hat{\mathbf{U}}} \mathbb{E}_{\mathbf{U}, \mathbf{Y}} \{ \text{Tr} \{ \hat{\mathbf{U}}^H \mathbf{U}\mathbf{U}^H \hat{\mathbf{U}} \} \} \\ &= \arg \max_{\hat{\mathbf{U}}} \int \left[ \int \text{Tr} \{ \hat{\mathbf{U}}^H \mathbf{U}\mathbf{U}^H \hat{\mathbf{U}} \} p(\mathbf{U}|\mathbf{Y}) d\mathbf{U} \right] p(\mathbf{Y}) d\mathbf{Y} \quad (21)\end{aligned}$$

This integral can be maximized by directly maximizing the inner bracket w.r.t.  $\hat{\mathbf{U}}$  for all possible  $\mathbf{Y}$ , thus

$$\begin{aligned}\hat{\mathbf{U}}_{\text{MMSD}} &= \arg \max_{\hat{\mathbf{U}}} \int \text{Tr} \{ \hat{\mathbf{U}}^H \mathbf{U}\mathbf{U}^H \hat{\mathbf{U}} \} p(\mathbf{U}|\mathbf{Y}) d\mathbf{U} \\ &= \arg \max_{\hat{\mathbf{U}}} \text{Tr} \left\{ \hat{\mathbf{U}}^H \left[ \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y}) d\mathbf{U} \right] \hat{\mathbf{U}} \right\} \quad (22)\end{aligned}$$

which can be obtained as [9]

$$\hat{\mathbf{U}}_{\text{MMSD}} = \mathcal{P}_P \left\{ \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y}) d\mathbf{U} \right\} = \mathcal{P}_P \{ \mathbf{M}(p(\mathbf{U}|\mathbf{Y})) \} \quad (23)$$

in which

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Y})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y}) d\mathbf{U} \quad (24)$$

where the operator  $\mathcal{P}_P\{\cdot\}$ , that extracts the first  $P$  eigenvectors from a given matrix in  $\mathcal{H}_N^+$ , is defined by

$$\begin{aligned}\mathcal{P}_P : \mathcal{H}_N^+ &\longrightarrow \mathcal{U}_P^N \\ \mathbf{M}^{\text{EVD}} = [\mathbf{U}_P | \mathbf{U}_P^\perp] \mathbf{D} [\mathbf{U}_P | \mathbf{U}_P^\perp]^H &\longmapsto \mathcal{P}_P \{ \mathbf{M} \} = \mathbf{U}_P.\end{aligned} \quad (25)$$

The expression of the MMSD depends on  $p(\mathbf{U}|\mathbf{Y})$ , which is specified based on both the data model and the prior distribution assigned to the parameters. Usually, there is no closed-form solutions to compute  $\mathbf{M}(p(\mathbf{U}|\mathbf{Y}))$ . However, (23) can still be evaluated using the so-called induced arithmetic mean (IAM) [9] of the semi-unitary matrix, as

$$\hat{\mathbf{U}} \approx \mathcal{P}_P \left\{ \frac{1}{N_r} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{U}_{(n)} \mathbf{U}_{(n)}^H \right\} \quad (26)$$

where  $\mathbf{U}_{(n)}$  are sampled from  $p(\mathbf{U}|\mathbf{Y})$  (e.g. using the proposed method in Appendix A),  $N_{bi}$  stands for the burn-in samples (number of thrown samples from the Markov chain), and  $N_r$  is the number of samples used to evaluate the integral.

### 4.2.2. The subspace MMSD estimator for CG distributed sources

We recall that according to the data model described in Section 3,  $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_p\})$  and  $\mathbf{y}_k \sim \mathcal{CN}(\mathbf{0}, \tau_k \mathbf{U} \mathbf{U}^H + \sigma^2 \mathbf{I}_N)$ . Based on (22) and (6), the MMSD estimator of  $\mathbf{U}$  is expressed as the solution of the following optimization problem

$$\begin{aligned}\underset{\hat{\mathbf{U}}, \{\tau_k\}, \{\lambda_p\}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U}, \mathbf{Y}} \{ \|\hat{\mathbf{U}}\hat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H\|_F^2 \} \\ \text{subject to} \quad & \tau_k \geq 0 \quad \forall k, \quad \lambda_p \geq 0 \quad \forall p \\ & \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_P\end{aligned} \quad (27)$$

In order to solve this optimization problem, we derive in the following Section an iterative algorithm that sequentially updates the variables  $\hat{\mathbf{U}}$ ,  $\{\tau_k\}$  and  $\{\lambda_p\}$ . The update of  $\hat{\mathbf{U}}$  requires a Gibbs sampling scheme, while for updating both the texture  $\{\tau_k\}$  and the eigenvalues  $\{\lambda_p\}$ , we use the MM procedure from Section 4.1. The overall algorithm is summed up in the box Algorithm 1.

### 4.2.3. Algorithm derivation

The initialization of the variables  $\hat{\mathbf{U}}^0$ ,  $\{\lambda_p^0\}$  and  $\{\tau_k^0\}$  is done as for the MAP estimator. The updates of the blocks  $\hat{\mathbf{U}}$ ,  $\{\lambda_p\}$  and  $\{\tau_k\}$  are detailed below

- Update of the basis  $\hat{\mathbf{U}}$ :

For fixed blocks  $\{\tau_k^t\}$  and  $\{\lambda_p^t\}$ , the update  $\hat{\mathbf{U}}^{t+1}$  is obtained by solving the following problem

$$\begin{aligned}\underset{\hat{\mathbf{U}}}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{U}, \mathbf{Y}} \{ \|\hat{\mathbf{U}}\hat{\mathbf{U}}^H - \mathbf{U}\mathbf{U}^H\|_F^2 \} \\ \text{subject to} \quad & \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_P\end{aligned} \quad (28)$$

Thanks to the expression given in Section 4.2.1, the update is obtained by

$$\hat{\mathbf{U}}^{t+1} = \mathcal{P}_P \{ \mathbf{M}(p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}, \{\lambda_p^t\})) \} \quad (29)$$

with

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}, \{\lambda_p^t\})) = \int \mathbf{U}\mathbf{U}^H p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}, \{\lambda_p^t\}) d\mathbf{U} \quad (30)$$

The posterior probability in (9) is recognized as  $(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}, \{\lambda_p^t\}) \sim \text{CGBL}(\mathbf{C}, \{\mathbf{G}_p^t\})$  with  $\mathbf{G}_p^t = \mathbf{A}_p + \mathbf{M}_p^t$ ,  $\forall p \in [[1, P]]$ . With this general distribution, there is no closed form for computing the integral in (30). Nevertheless, the update can be evaluated by the IAM as given in (26) where  $\mathbf{U}_{(n)}^t$  are sampled as  $\mathbf{U}_{(n)}^t \sim \text{CGBL}(\mathbf{C}, \{\mathbf{G}_p^t\})$ . In order to do so, an efficient Gibbs sampling procedure to draw the CGBL distribution is given in Algorithm 6 of Appendix A. Special

---

#### Algorithm 2: MMSD for the general model (3) $\mathbf{U}_{\text{MMSD}}$ .

---

**input** :  $\mathbf{Y}, \mathbf{C}, \{\mathbf{A}_p\}, \sigma^2, P, K, N, N_{bi}, N_r$   
**output** :  $\hat{\mathbf{U}}_{\text{MMSD}}, \{\tau_k\}, \{\lambda_p\}$   
**initialize**:  $\hat{\mathbf{U}}^0, \{\tau_k^0\}$  and  $\{\lambda_p^0\}$   
**1 for**  $t = 0 \dots T - 1$  **do**  
**2**    **for**  $n = 1 \dots N_{bi} + N_r$  **do**  
**3**       Sample  $\mathbf{U}_{(n)} = \text{CGBL}(\mathbf{C}, \{\mathbf{G}_p\})$  ccf Appendix A  
**4**    **end**  
**5**    Update  $\hat{\mathbf{U}}^{t+1} \approx \mathcal{P}_P \left\{ \frac{1}{N_r} \sum_{n=N_{bi}+1}^{N_{bi}+N_r} \mathbf{U}_{(n)} \mathbf{U}_{(n)}^H \right\}$   
**6**    Update  $\tau_k^{t+1} \forall k$  with (18)  
**7**    Update  $\lambda_p^{t+1} \forall p$  with (20)  
**8 end**

---

cases for the sampling scheme required on this update are given in Table 2.

- Update of the eigenvalues  $\{\lambda_p\}$  and the textures  $\{\tau_k\}$ :

For fixed  $\hat{\mathbf{U}}^{t+1}$ , the update of the eigenvalues and the texture is equivalent to solve respectively the ML problem (19) and (17) since these parameters are unknown deterministic (cf. remark 2). Consequently, the updates of  $\{\tau_k^{t+1}\}$  and  $\{\lambda_p^{t+1}\}$  are obtained respectively from (18) and (20).

### 4.3. Simplified model: white CG model with CIB prior

In this Section, we focus on a special case that we refer to as simplified model, where  $\Sigma = \lambda \mathbf{U}\mathbf{U}^H$ . This relaxation of the true model, e.g. used in [32], allows for interesting simplifications that significantly reduce the computational time of the estimation procedure. Note that any scaling on the scatter can be absorbed in the textures parameters as  $\tilde{\tau} = \lambda \tau$ , so we can assume  $\Sigma = \mathbf{U}\mathbf{U}^H$ . For this model,  $\mathbf{y}_k | \mathbf{U}, \tau_k \sim \mathcal{CN}(\mathbf{0}, \Sigma_k)$ , thus, the covariance reads

$$\Sigma_k = \tau_k \mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I}_N \quad (31)$$

Using the Sherman Morrison Woodbury lemma,  $\Sigma_k^{-1}$  reads as

$$\Sigma_k^{-1} = (\tau_k \mathbf{U}\mathbf{U}^H + \sigma^2 \mathbf{I})^{-1} = \sigma^{-2} \mathbf{I} - \frac{\tau_k}{\sigma^2(\tau_k + \sigma^2)} \mathbf{U}\mathbf{U}^H, \quad \forall k \quad (32)$$

**Table 2**Posterior distributions for standard priors on  $\mathbf{U}$  under simplified (SM) and general (GM) models.

The prior complex distribution of $\mathbf{U}$	Posterior distribution $p(\mathbf{U} \mathbf{Y})$ for a given model (SM/GM)	To be sampled for MMSD
Bingham CB( $\mathbf{A}, \Phi$ )	SM: $p(\mathbf{U} \mathbf{Y})_{\text{SM}} \propto \exp\left\{\sum_{p=1}^P \mathbf{u}_p^H (\phi(p)\mathbf{A} + \mathbf{W}) \mathbf{u}_p\right\}$ GM: $p(\mathbf{U} \mathbf{Y})_{\text{GM}} \propto \exp\left\{\sum_{p=1}^P \mathbf{u}_p^H (\phi(p)\mathbf{A} + \mathbf{G}_p) \mathbf{u}_p\right\}$	CB( $\{\phi(p)\mathbf{A} + \mathbf{W}\}$ ) CB( $\{\phi(p)\mathbf{A} + \mathbf{G}_p\}$ )
Langevin CL( $\mathbf{C}$ )	SM: $p(\mathbf{U} \mathbf{Y})_{\text{SM}} \propto \text{etr}\{\text{Re}\{\mathbf{C}^H \mathbf{U}\} + \mathbf{U}^H \mathbf{W} \mathbf{U}\}$ GM: $p(\mathbf{U} \mathbf{Y})_{\text{GM}} \propto \exp\left\{\sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H \mathbf{G}_p \mathbf{u}_p\right\}$	CBL( $\mathbf{C}, \mathbf{I}_P, \mathbf{W}$ ) CGBL( $\mathbf{C}, \{\mathbf{G}_p\}$ )
Invariant Bingham CIB( $\kappa, \mathbf{A}$ )	SM: $p(\mathbf{U} \mathbf{Y})_{\text{SM}} \propto \text{etr}\{\mathbf{U}^H (\kappa \mathbf{A} + \mathbf{W}) \mathbf{U}\}$ GM: $p(\mathbf{U} \mathbf{Y})_{\text{GM}} \propto \exp\left\{\sum_{p=1}^P \mathbf{u}_p^H (\kappa \mathbf{A} + \mathbf{G}_p) \mathbf{u}_p\right\}$	Closed form $\mathcal{P}_P\{\kappa \mathbf{A} + \mathbf{W}\}$ CB( $\{\kappa \mathbf{A} + \mathbf{G}_p\}$ )

Then, the p.d.f.  $p(\mathbf{Y}|\mathbf{U}, \{\tau_k\})$  reduces to

$$p(\mathbf{Y}|\mathbf{U}, \{\tau_k\}) \propto \prod_{k=1}^K p(\mathbf{y}_k|\mathbf{U}, \tau_k) \propto \prod_{k=1}^K \frac{\exp\{-\mathbf{y}_k^H \Sigma_k^{-1} \mathbf{y}_k\}}{\det(\Sigma_k)} \quad (33)$$

$$\propto \exp\left\{\sum_{k=1}^K \frac{\tau_k}{\sigma^2(\tau_k + \sigma^2)} \mathbf{U}^H \mathbf{y}_k \mathbf{y}_k^H \mathbf{U}\right\} \left(\prod_{k=1}^K (\tau_k + \sigma^2)^{-P}\right) \quad (34)$$

where  $\mathbf{W} = \mathbf{Y} \mathbf{B} \mathbf{Y}^H$  and  $\mathbf{B} = \text{diag}\left(\left\{\frac{\tau_1}{\sigma^2(\tau_1 + \sigma^2)} \cdots \frac{\tau_K}{\sigma^2(\tau_K + \sigma^2)}\right\}\right)$ . In order to obtain closed form expression, we assign to  $\mathbf{U}$  a CIB distribution, i.e.,  $\mathbf{U} \sim \text{CIB}(\kappa, \mathbf{A})$ , thus its p.d.f. reads as

$$p_{\text{CIB}}(\mathbf{U}) \propto \text{etr}\{\kappa \mathbf{U}^H \mathbf{A} \mathbf{U}\} \quad (35)$$

#### 4.3.1. The MMSD estimator for the simplified model

In this case, the MMSD estimator of the basis  $\hat{\mathbf{U}}$  is defined as the minimizer of the following problem

$$\begin{aligned} & \underset{\hat{\mathbf{U}}, \{\tau_k\}}{\text{minimize}} && \mathbb{E}_{\mathbf{U}, \mathbf{Y}}\{\|\hat{\mathbf{U}} \hat{\mathbf{U}}^H - \mathbf{U} \mathbf{U}^H\|_F^2\} \\ & \text{subject to} && \tau_k \geq 0, \forall k \\ & && \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_P \end{aligned} \quad (36)$$

Following the previous lines, we develop an iterative estimation method by solving (36) w.r.t.  $\hat{\mathbf{U}}$  and  $\{\tau_k\}$  sequentially. The box [Algorithm 3](#) sums-up the main steps of the estimation process

**Algorithm 3:** MMSD estimation of subspace for the simplified model  $\hat{\mathbf{U}}_{\text{sMMSD}}$ .

---

**input** :  $\mathbf{Y}, \mathbf{A}, \sigma^2, P, K, N, \kappa$   
**output** :  $\hat{\mathbf{U}}_{\text{sMMSD}}, \{\tau_k\}$   
**initialize**:  $\hat{\mathbf{U}}^0, \{\tau_k^0\}$   
**1 while** stop criterion **unreached do**  
**2** | Update  $\hat{\mathbf{U}}^t = \mathcal{P}_P\{\kappa \mathbf{A} + \mathbf{W}^t\}$   
**3** | Update  $\tau_k^t = \max\left(\frac{\mathbf{y}_k^H \hat{\mathbf{U}}^t \hat{\mathbf{U}}^t \mathbf{y}_k}{P} - \sigma^2, 0\right)$   
**4 end**

---

which are detailed in the following.

- Update of the basis  $\hat{\mathbf{U}}$ :

Now, we assume that the block  $\{\tau_k^t\}$  is fixed, consequently, the updated  $\hat{\mathbf{U}}^{t+1}$  is the solution of the following problem

$$\begin{aligned} & \underset{\hat{\mathbf{U}}}{\text{minimize}} && \mathbb{E}_{\mathbf{U}, \mathbf{Y}}\{\|\hat{\mathbf{U}} \hat{\mathbf{U}}^H - \mathbf{U} \mathbf{U}^H\|_F^2\} \\ & \text{subject to} && \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_P \end{aligned} \quad (37)$$

In this case, the updated basis  $\hat{\mathbf{U}}^{t+1}$  is derived as the MMSD estimator in (23) leading to

$$\hat{\mathbf{U}}^{t+1} = \mathcal{P}_P\{\mathbf{M}(p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}))\} \quad (38)$$

where

$$\mathbf{M}(p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\})) = \int \mathbf{U} \mathbf{U}^H p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}) d\mathbf{U} \quad (39)$$

From (35) and (33), the posterior probability  $p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\})$  reads as

$$\begin{aligned} p(\mathbf{U}|\mathbf{Y}, \{\tau_k^t\}) & \propto p(\mathbf{Y}|\mathbf{U}, \{\tau_k^t\}) p_{\text{CIB}}(\mathbf{U}) \propto \text{etr}\{\kappa \mathbf{U}^H \mathbf{A} \mathbf{U}\} \text{etr}\{\mathbf{U}^H \mathbf{W}^t \mathbf{U}\} \\ & \propto \text{etr}\{\mathbf{U}^H (\kappa \mathbf{A} + \mathbf{W}^t) \mathbf{U}\} \end{aligned} \quad (40)$$

with  $\mathbf{W}^t = \mathbf{Y} \mathbf{B}^t \mathbf{Y}^H$  and  $\mathbf{B}^t = \text{diag}\left(\left\{\frac{\tau_1^t}{\sigma^2(\tau_1^t + \sigma^2)} \cdots \frac{\tau_K^t}{\sigma^2(\tau_K^t + \sigma^2)}\right\}\right)$ . Using Proposition 1 from [9], we notice that the updated basis admits the following closed form expression

$$\hat{\mathbf{U}}^{t+1} = \mathcal{P}_P\left\{\int \mathbf{U} \mathbf{U}^H \text{etr}\{\mathbf{U}^H (\kappa \mathbf{A} + \mathbf{W}^t) \mathbf{U}\} d\mathbf{U}\right\} = \mathcal{P}_P\{\kappa \mathbf{A} + \mathbf{W}^t\} \quad (41)$$

This specific model provides a closed-form solution with an interesting interpretations. Indeed, this MMSD appears naturally as the principal subspace of the sum of the SCM using scaled samples and scaled prior subspace projector.

- Update of the texture parameter  $\{\tau_k\}$

For fixed  $\hat{\mathbf{U}}^{t+1}$ , the update of  $\{\tau_k\}$  is obtained by maximizing the p.d.f.  $p(\mathbf{Y}|\hat{\mathbf{U}}^{t+1}, \{\tau_k\})$  as

$$\begin{aligned} & \underset{\{\tau_k\}}{\text{maximize}} && p(\mathbf{Y}|\hat{\mathbf{U}}^{t+1}, \{\tau_k\}) \\ & \text{subject to} && \tau_k \geq 0, \forall k \end{aligned} \quad (42)$$

with  $\Sigma_k = \tau_k \hat{\mathbf{U}}^{t+1} \hat{\mathbf{U}}^{t+1H} + \sigma^2 \mathbf{I}_N$ ,  $\forall k$ . Minimizing the negative log-likelihood is equivalent to solve

$$\begin{aligned} & \underset{\{\tau_k\}}{\text{minimize}} && \sum_{k=1}^K \ln(\det(\Sigma_k^{t+1})) + \mathbf{y}_k^H (\Sigma_k^{t+1})^{-1} \mathbf{y}_k \\ & \text{subject to} && \tau_k \geq 0, \forall k \end{aligned} \quad (43)$$

which leads to

$$\tau_k^{t+1} = \max\left(\frac{\mathbf{y}_k^H \hat{\mathbf{U}}^{t+1} \hat{\mathbf{U}}^{t+1H} \mathbf{y}_k}{P} - \sigma^2, 0\right), \forall k \quad (44)$$

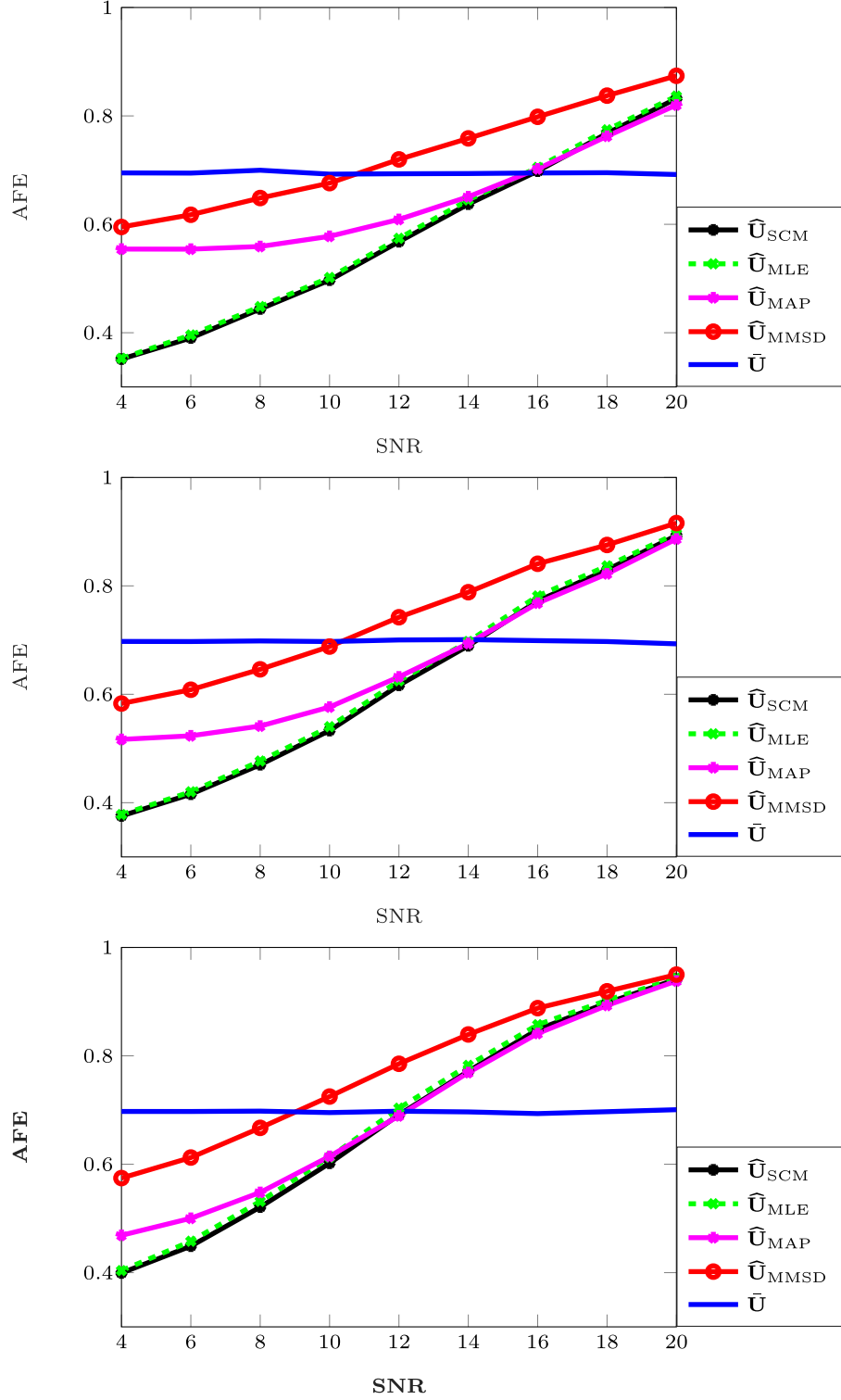
#### 4.3.2. Link with the MAP estimator for the simplified model

From (40), the update of the basis of interest is the solution of the following problem:

$$\begin{aligned} & \underset{\hat{\mathbf{U}}}{\text{maximize}} && \hat{\mathbf{U}}^H (\kappa \mathbf{A} + \mathbf{W}^t) \hat{\mathbf{U}} \\ & \text{subject to} && \hat{\mathbf{U}}^H \hat{\mathbf{U}} = \mathbf{I}_P \end{aligned} \quad (45)$$

Given that  $\kappa \mathbf{A} + \mathbf{W}^t \in \mathcal{H}_N^+$ , the updated basis for the MAP estimator is

$$\hat{\mathbf{U}}^{t+1} = \mathcal{P}_P\{\kappa \mathbf{A} + \mathbf{W}^t\} \quad (46)$$



**Fig. 1.** AFE w.r.t. SNR for  $P=5$ ,  $N=20$ ,  $\mathbf{U} \sim \text{CL}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ ,  $\kappa=80$ ,  $\nu=0.5$ , from top to bottom:  $K=3P$ ,  $K=4P$  and  $K=6P$ .

which corresponds to (41). Furthermore, we can notice that the MAP update of  $\{\tau_k^{t+1}\}$  is identical to (44). Therefore, in this case, the MAP estimator coincides with the MMSE estimator.

#### 4.4. Notes on complexity and convergence analysis

##### 4.4.1. Computational complexities

In this subsection, we detail the complexity of the proposed algorithms. Notice that we focus on the cost of each variable up-

dates. The total complexity of the algorithms is to be scaled by the number of iterations. For both the MAP and the MMSE, the update of the textures  $\{\tau_k\}$  and the eigenvalues  $\{\lambda_p\}$  are obtained in closed form that only involve scalar multiplications/additions ( $\mathcal{O}(NKP)$ ). The bottleneck of each algorithm lies in the update of the eigenvectors  $\mathbf{U}$ :

- MAP: The derivation of  $\mathbf{U}_{\text{MAP}}$  requires the computation of  $\mathbf{H}^t$  in (15) ( $\mathcal{O}(NKP)$ ) and its TSVD of  $\mathbf{H}^t$  in (14) ( $\mathcal{O}(N^2P + P^2N)$ ).



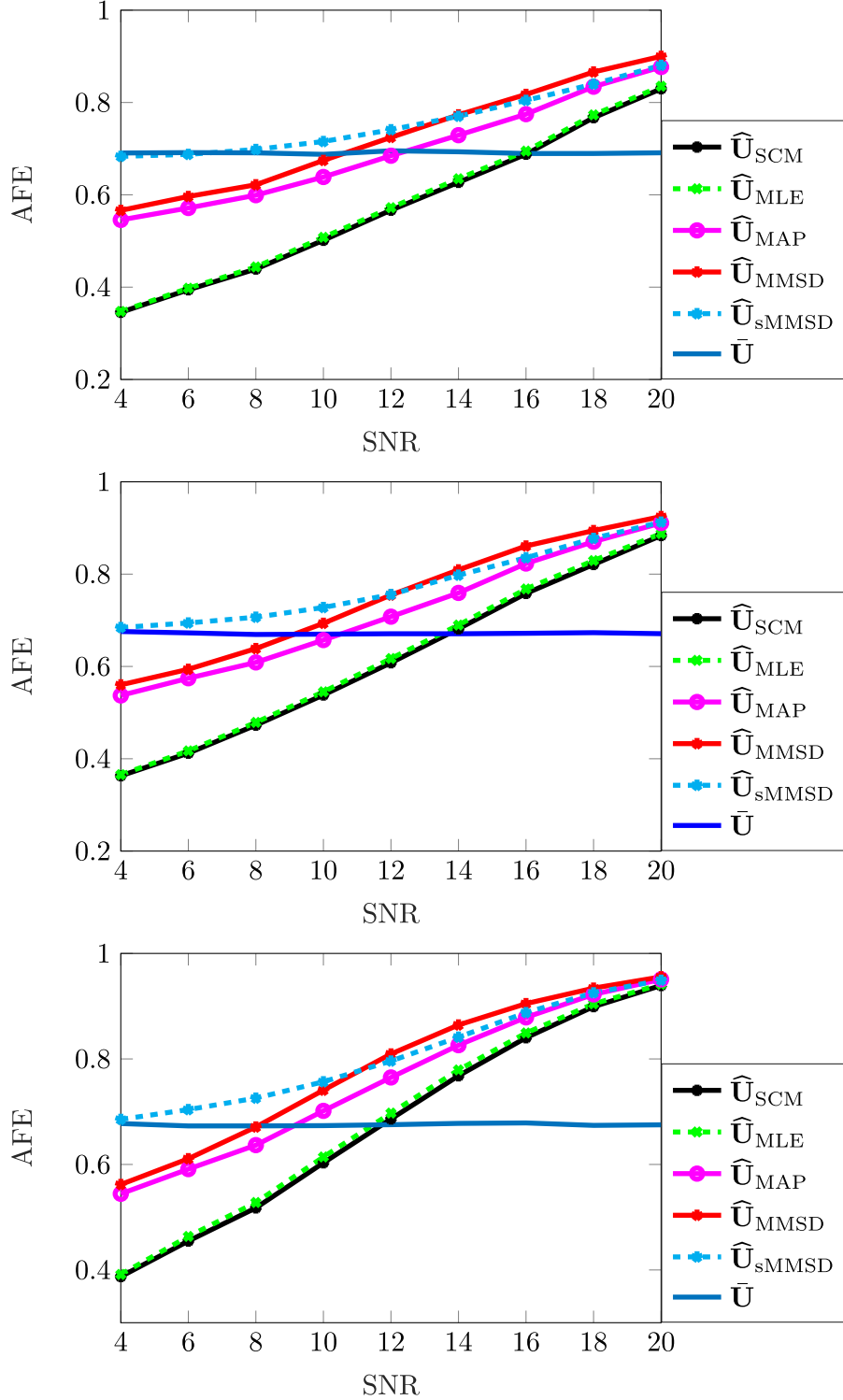


Fig. 2. AFE w.r.t. SNR for  $P=5$ ,  $N=20$ ,  $\nu=0.5$ ,  $\mathbf{U} \sim \text{CB}(\kappa_0 \Phi, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ ,  $\kappa_0=300$  from top to bottom:  $K=3P$ ,  $K=4P$  and  $K=6P$ .

- MMMSD: The update of  $\mathbf{U}_{\text{MMMSD}}$  requires the evaluation of the matrix  $\mathbf{M}$  in (30) with a gibbs sampler (which cost cannot be evaluated analytically) and the computation of its  $P$  strongest eigenvectors ( $\mathcal{O}(N^2P)$ ). Nevertheless, (30) can be obtained in closed-form for the simplified model ( $\mathcal{O}(N^2P)$ ).

#### 4.4.2. Convergence analysis

The MM algorithm ensures a monotonic decrement of the objective function at each iteration [34]. A convergence anal-

ysis for the MAP algorithm can be directly conducted as in [22,35]. However, this analysis cannot be applied directly to the MMMSD since this estimator requires to approximate the exact update by the Gibbs sampler in (26). Therefore, the convergence of Algorithm 2 remains an opened question. Nevertheless, Section 4.1.1 shows that the numerical performance obtained with Algorithm 2 is satisfactory.

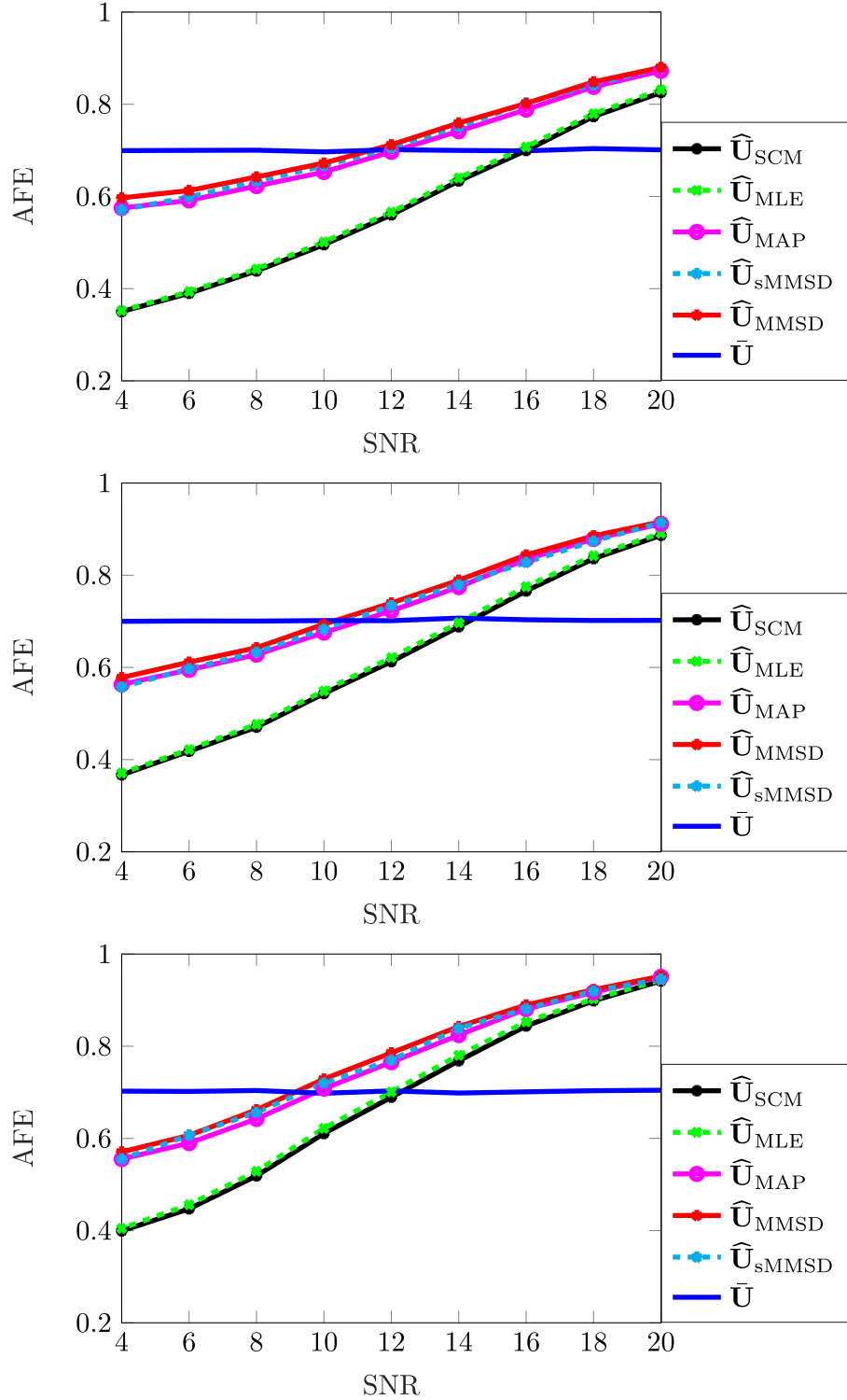


Fig. 3. AFE w.r.t. SNR for  $P=5$ ,  $N=20$ ,  $\nu=0.5$ ,  $\mathbf{U} \sim \text{CIB}(\kappa, \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)$ ,  $\kappa=50$ , from top to bottom:  $K=3P$ ,  $K=4P$  and  $K=6P$ .

## 5. Numerical simulations

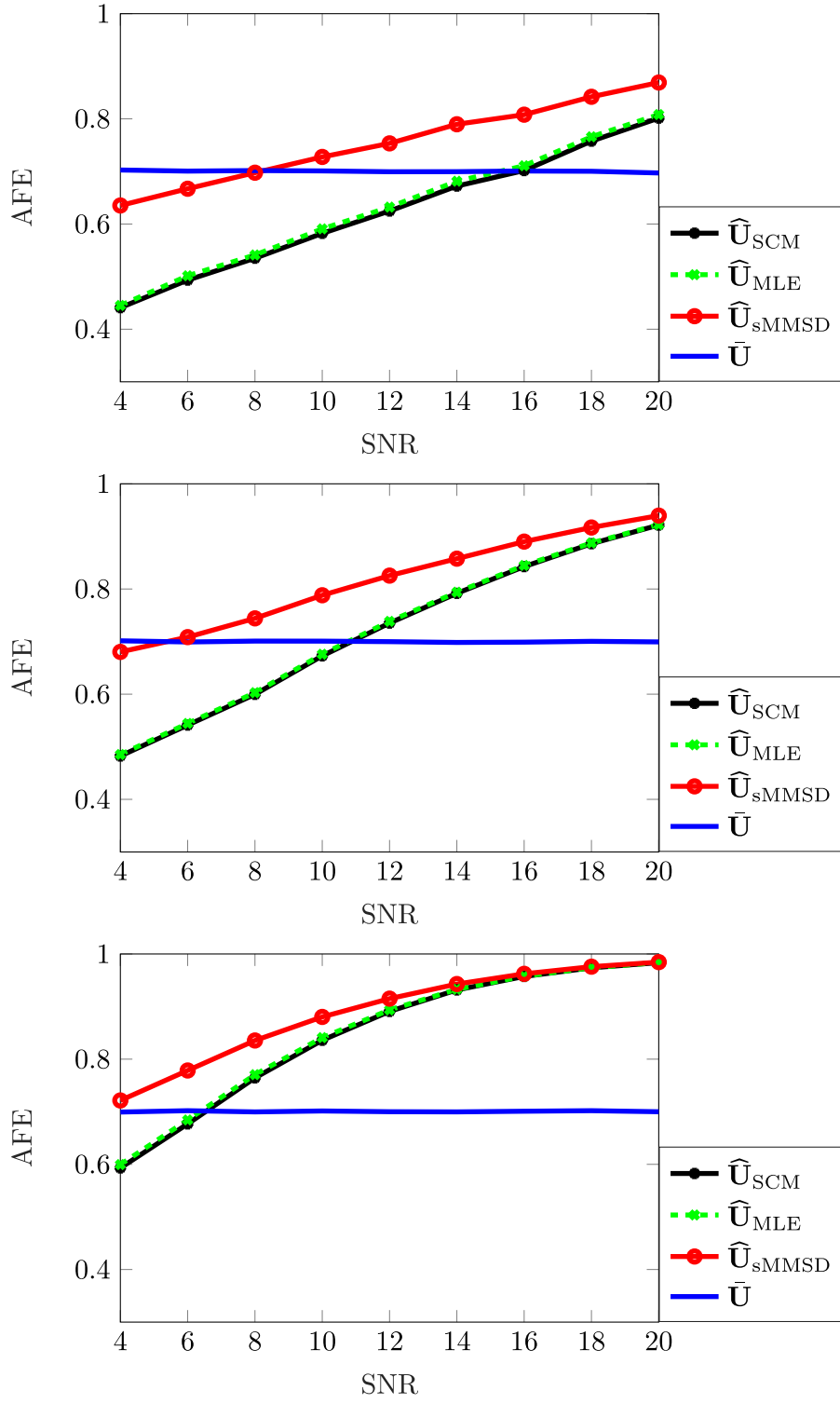
### 5.1. Setup

To illustrate the performance of the proposed estimators, we evaluate their average fraction of energy (AFE) through Monte Carlo simulations. The AFE is considered as an adequate criteria of performance for subspace estimation, since it evaluates the closeness of the true range space  $\mathbf{U}\mathbf{U}^H$  towards its estimate  $\hat{\mathbf{U}}\hat{\mathbf{U}}^H$ . The

AFE of a given estimator  $\hat{\mathbf{U}}$  is expressed as:

$$\text{AFE}(\hat{\mathbf{U}}) = \mathbb{E}\{\text{Tr}\{\mathbf{U}^H \hat{\mathbf{U}}\hat{\mathbf{U}}^H \mathbf{U}\}\}/P \quad (47)$$

The samples are generated from the model in Section 3, i.e.  $\mathbf{y}_k \sim \mathcal{CN}(0, \tau_k \mathbf{U}\mathbf{A}\mathbf{U}^H + \sigma^2 \mathbf{I})$ . The texture parameters  $\{\tau_k\}$  follow a Gamma distribution parameterized by its shape  $\nu$  which reflects the heterogeneity of the sources, i.e.,  $\tau_k \sim \Gamma(\nu, \frac{1}{\nu})$ ,  $\forall k$  (thus  $\mathbb{E}\{\tau\} = 1$ ). We set  $[\mathbf{A}]_{p,p} = (P+1-p)/(\sum_{i=1}^P i)$  and  $\sigma^2$  to fix the



**Fig. 4.** AFE w.r.t. SNR for  $P=5$ ,  $N=20$ ,  $\kappa=50$ ,  $\nu=0.5$ , from top to bottom:  $K=2P$ ,  $K=3P$  and  $K=4P$ .

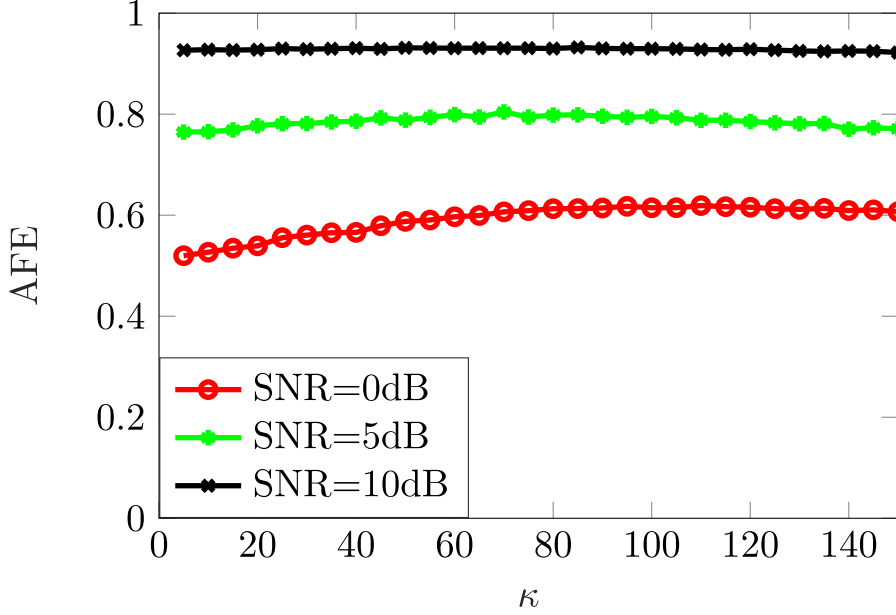


Fig. 5. AFE of the MMSD w.r.t. assumed  $\kappa$  for various SNR,  $P=5$ ,  $N=20$ ,  $\nu=1$ ,  $K=30$ ,  $\mathbf{U} \sim \text{CIB}(\kappa_0, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ , with true parameter  $\kappa_0=60$ .

signal to noise ratio as  $\text{SNR} = \text{Tr}\{\mathbf{\Lambda}\}/\sigma^2$ . We consider different scenarios for the distribution of  $\mathbf{U}$ :

- S1:  $\mathbf{U}$  follows the complex Langevin distribution  $\mathbf{U} \sim \text{CL}(\kappa\bar{\mathbf{U}})$ ,
- S2:  $\mathbf{U}$  follows the complex Bingham distribution  $\mathbf{U} \sim \text{CB}(\kappa\Phi, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$  where  $[\Phi]_{p,p} = (P+1-p)/(\sum_{i=1}^P i)$ ,
- S3:  $\mathbf{U}$  follows the complex invariant Bingham distribution  $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ ,

where  $\bar{\mathbf{U}} \in \mathcal{U}_P^N$  are the first vectors of the canonical basis and the concentration parameters ( $\kappa$  and  $\Phi$ ) are set so that  $\text{AFE}(\bar{\mathbf{U}})$  has the same value for all the scenarios. We compare the following estimators: i)  $\hat{\mathbf{U}}_{\text{SCM}} = \mathcal{P}_P\{\mathbf{Y}\mathbf{Y}^H\}$ , the estimator built from the EVD of the SCM; ii)  $\hat{\mathbf{U}}_{\text{MLE}}$  is the subspace MLE computed with EBMM Algorithm from [22]; iii)  $\hat{\mathbf{U}}_{\text{MAP}}$  is the proposed MAP estimator, computed with Algorithm 1; iv)  $\hat{\mathbf{U}}_{\text{MMSD}}$  is the proposed MMSD estimator computed with Algorithm 2; v)  $\hat{\mathbf{U}}_{\text{sMMSD}}$  the simplified MMSD estimator, that assumes  $\mathbf{\Lambda} = \mathbf{I}$  and  $\mathbf{U} \sim \text{CIB}(\kappa, \bar{\mathbf{U}}\bar{\mathbf{U}}^H)$ , computed with Algorithm 3. This estimator is evaluated for S2 and S3 but the relaxation is not suited for S1 (where the true prior is a complex Langevin); vi)  $\bar{\mathbf{U}}$  is the center of the prior distribution on  $\mathbf{U}$ .

## 5.2. Results

Fig. 1 displays the AFE in function of SNR for various sample size  $K$  in scenario S1. In this case, the SCM exhibits good performance in the standard regimes (high SNR and/or large  $K$ ). The textures parameter is  $\nu=0.5$  so the sources are mildly impulsive. Therefore the MLE exhibits performances close to the SCM as it can be expected (differences will be observed in the following). However, both show their limits at low SNR. In this challenging context, Bayesian estimators can leverage the prior information and exhibit better performance in terms of AFE. Interestingly, for the complex Langevin prior, the MMSD outperforms the MAP, which reaches performance close to SCM/MLE as the SNR increases.

Fig. 2 displays the AFE in function of SNR for various sample size  $K$  in scenario S2. The same general observations as in the previous Figure can be drawn. For the complex Bingham prior case, the MMSD still outperforms the MAP, but not as significantly as in the scenario S1. We also observe that sMMSD, that assumes

equals eigenvalues and a mismatched (averaged) prior, offers an interesting performance versus computational time trade-off when it comes to estimate only the signal subspace. By construction of the true prior, the first column-vectors of  $\mathbf{U}$  exhibits less variance than the last ones. By uniformly averaging the prior for each vectors, sMMSD introduces a bias towards the center of distribution, which explains its performance close to  $\bar{\mathbf{U}}$  at low SNR.

Fig. 3 displays the AFE in function of SNR for various sample size  $K$  in scenario S3. Here, the sMMSD assumes the true prior and is only mismatched by assuming equals eigenvalues. In this scenario the sMMSD exhibits performance almost identical to the MMSD, which suggest that it is acceptable to relax the eigenvalue estimation when it comes to estimate only the signal subspace.

Fig. 4 displays the AFE in function of SNR for various sample size  $K$  in the actual simplified model, i.e., the scenario S3 where  $\mathbf{\Lambda} = \mathbf{I}$ . In this context, the MMSD and the MAP coincide with sMMSD and we still observe the interest of the Bayesian approach in challenging contexts (low SNR and/or  $K$ ).

## 5.3. Robustness to the concentration parameter and the signal distribution

First, we study the effect of a miss-selected concentration parameter  $\kappa$  on the AFE of the proposed Bayesian estimator. The setup of Fig. 5 is the same as for Fig. 4 (simplified model where the MMSD and MAP coincide) and displays the AFE of the MMSD estimator w.r.t. the assumed  $\kappa$ , while the true concentration parameter  $\kappa_0$  is fixed. This figure illustrates that, for a reasonable range of  $\kappa$ , the AFE of the MMSD estimator remains almost unchanged. Thus, the proposed method appears robust to a reasonable miss-selection of the concentration parameters of the assumed prior distribution.

Second, we study the performance of the proposed method w.r.t. the signal distribution, parameterized by the shape  $\nu$ . The setup of Fig. 6 is the same as for Fig. 4 and displays the AFE of the sMMSD estimator w.r.t  $\nu$  for various SNRs. When  $\nu \rightarrow 1$ , the signal tends to be more impulsive (i.e., heavy tailed distributed). In this context, we can notice a slight difference between the SCM and the MLE, which illustrates the interest of taking the non-Gaussianity into account. Interestingly, the performance drop hap-

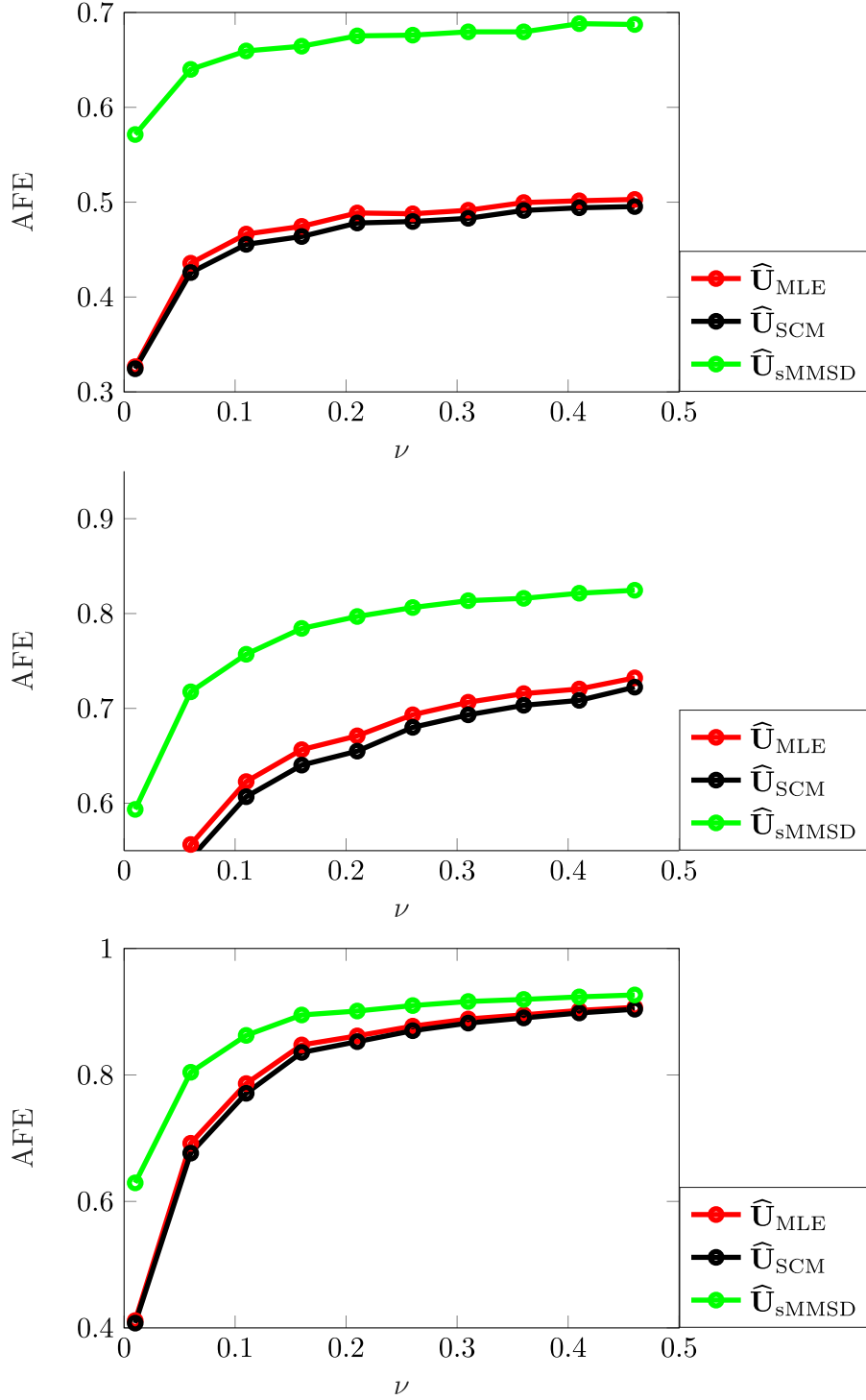


Fig. 6. AFE w.r.t.  $\nu$  for  $P=5$ ,  $N=20$ ,  $K=30$ ,  $\mathbf{U} \sim \text{CIB}(\kappa_0, \tilde{\mathbf{U}}\tilde{\mathbf{U}}^H)$  from top to bottom SNR=0dB, SNR=5dB and SNR=10dB.

pens at lower  $\nu$  for the Bayesian estimator, which shows the interest of exploiting both the non-Gaussian assumption and the prior knowledge.

## 6. Conclusion

In this paper, we considered a Bayesian approach for subspace estimation. First, we introduced a generalized version of the complex Bingham Langevin distribution (CBL) in order to model the prior distribution of the subspace orthonormal basis. Second, we

formulated the MAP and the MMSD estimators of the signal subspace in the context of CG distributed sources and CGBL distributed subspace. Finally, simulations illustrated the interest of the proposed approach in critical regimes (low SNR and/or low sample size).

## Declaration of Competing Interest

We have no conflicts of interest to disclose.

## Acknowledgements

This work was supported in part by ANR-ASTRID MARGARITA (ANR-17-ASTR-0015).

## Appendix A. Generation of the CGBL distribution

In this appendix, we present the procedure in order to sample a semi-unitary random matrix  $\mathbf{U} \in \mathcal{U}_p^N$  from the complex generalized Bingham Langevin distribution (CGBL), i.e.,  $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_p\})$ , where  $\mathbf{C} \in \mathbb{C}_p^N$  and  $\{\mathbf{A}_p\} \subset \mathcal{H}_N^+$ . The p.d.f. of  $\mathbf{U}$  is given by

$$\begin{aligned} p_{\text{CGBL}}(\mathbf{U}|\mathbf{C}, \{\mathbf{A}_p\}) &= \prod_{p=1}^P p_{\text{CVGBL}}(\mathbf{u}_p|\mathbf{c}_p, \mathbf{A}_p) \\ &\propto \exp \left\{ \sum_{p=1}^P \text{Re}\{\mathbf{c}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H \mathbf{A}_p \mathbf{u}_p \right\} \\ &\propto \prod_{p=1}^P \exp \left\{ \text{Re}\{\mathbf{c}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H \mathbf{A}_p \mathbf{u}_p \right\} \end{aligned} \quad (48)$$

where  $\mathbf{u}_p$  and  $\mathbf{c}_p$  stand for the  $p$ th column of respectively  $\mathbf{U}$  and  $\mathbf{C}$ . Most of the upcoming results shown below are based on the methodology of [25,26]. In the following, we first recall the sampling procedure of a real vector Bingham Langevin (vBL) distributed vector. Then, we define and sample a complex vGBL distributed vector. Finally, we deduce the generation of complex a GBL distributed matrix.

### A1. Sampling a real Bingham Langevin distributed vector

The vBL distribution [25] is a probability distribution on the set of unitary real vectors which combines linear and quadratic terms denoted as  $\mathbf{u} \sim \text{vBL}(\mathbf{c}, \mathbf{A})$  where the matrix  $\mathbf{A}$  is a symmetric matrix and  $\mathbf{c}$  is a real vector. The p.d.f. of  $\mathbf{u} \sim \text{vBL}(\mathbf{c}, \mathbf{A})$  is proportional to

$$f_{\text{vBL}}^*(\mathbf{u}) = \exp\{\mathbf{c}^T \mathbf{u} + \mathbf{u}^T \mathbf{A} \mathbf{u}\} \quad (49)$$

In [26], an acceptance-rejection scheme is proposed to sample the vBL distribution using an angular central Gaussian distribution denoted as  $\text{ACG}(\mathbf{\Omega})$  with  $\mathbf{\Omega}$  is a symmetric positive definite matrix. Its p.d.f. reads as

$$f_{\text{ACG}}^*(\mathbf{u}) = |\mathbf{\Omega}|^{1/2} (\mathbf{u}^T \mathbf{\Omega} \mathbf{u})^{-N/2} \quad (50)$$

with

$$\mathbf{\Omega} = \mathbf{I} + (2/b)(\gamma \mathbf{I} - \mathbf{A} - 1/2 \mathbf{c} \mathbf{c}^T) \quad (51)$$

where  $\gamma = \max(\text{eig}(\mathbf{A} + 1/2 \mathbf{c} \mathbf{c}^T))$  and  $b$  satisfies the following equality

$$\sum_{i=1}^N \frac{1}{b + 2\beta_i} = 1 \quad (52)$$

with  $\{\beta_i\}$  denotes the eigenvalues of the matrix  $\gamma \mathbf{I} - \mathbf{A} - 1/2 \mathbf{c} \mathbf{c}^T$ . This sampling technique is summed up in the box Algorithm 4.

### A2. Definition and sampling of the complex vector generalized Bingham Langevin distributed vector

Let us start first with defining the relation between the CVGBL distribution and the vBL distribution. Based on [26], a given complex unitary random vector  $\mathbf{u} \in \mathbb{C}^N$  with  $\mathbf{u} \sim \text{CVGBL}(\mathbf{c}, \mathbf{A})$  has a.d.f. of the form

$$p_{\text{CVGBL}}(\mathbf{u}) \propto \exp\{\text{Re}\{\mathbf{c}^H \mathbf{u}\} + \mathbf{u}^H \mathbf{A} \mathbf{u}\}$$

where  $\mathbf{A} \in \mathcal{H}_N^+$  and  $\mathbf{c}$  is a complex vector. Let us denote

$$\mathbf{u} = \mathbf{u}_1 + i\mathbf{u}_2, \quad \mathbf{A} = \mathbf{A}_1 + i\mathbf{A}_2 \quad \text{and} \quad \mathbf{c} = \mathbf{c}_1 + i\mathbf{c}_2 \quad (53)$$

---

**Algorithm 4:** Acceptance-rejection sampling scheme for the vBL distribution.

---

```

input :  $\mathbf{A}, \mathbf{c}$ 
output :  $\mathbf{u}$ 
1 Compute  $\gamma = \max(\text{eig}(\mathbf{A} + 1/2 \mathbf{c} \mathbf{c}^T))$ 
2 Compute  $b^*$  satisfying (52)
3 Compute
    $M^*(b^*) = \exp\{1/2 + \gamma\} \exp\{-(N - b^*)/2\} (N/b^*)^{N/2} |\mathbf{\Omega}|^{-1/2}$ 
4 Compute  $\mathbf{\Omega} = \mathbf{I} + (2/b^*)(\gamma \mathbf{I} - \mathbf{A} - 1/2 \mathbf{c} \mathbf{c}^T)$ 
5 repeat
6   Sample  $\mathbf{y} \sim \mathcal{N}(0, \mathbf{\Omega}^{-1})$ 
7   Compute  $\mathbf{u} = \mathbf{y}/\|\mathbf{y}\|$ 
8   Compute  $f_{\text{ACG}}^*(\mathbf{u}) = |\mathbf{\Omega}|^{1/2} (\mathbf{u}^T \mathbf{\Omega} \mathbf{u})^{-N/2}$ 
9   Compute  $f_{\text{vBL}}^*(\mathbf{u}) = \exp\{\mathbf{c}^T \mathbf{u} + \mathbf{u}^T \mathbf{A} \mathbf{u}\}$ 
10  Sample  $u \sim \text{Unif}(0, 1)$ 
11 until  $u < f_{\text{vBL}}^*(\mathbf{u}) / (M^*(b^*) f_{\text{ACG}}^*(\mathbf{u}))$ ;
12 Accept  $\mathbf{u}$ 

```

---

where  $\mathbf{u}_1, \mathbf{c}_1$  are respectively the real parts of  $\mathbf{u}, \mathbf{c}$  and  $\mathbf{u}_2, \mathbf{c}_2$  are respectively the imaginary parts of  $\mathbf{u}, \mathbf{c}$ . The matrix  $\mathbf{A}_1$  is symmetric and  $\mathbf{A}_2$  is a skew-symmetric matrix. In the following, we aim to introduce a relation between the vBL distribution and the CVGBL distribution.

$$\begin{aligned} p_{\text{CVGBL}}(\mathbf{u}) &\propto \exp\{\text{Re}\{\mathbf{c}^H \mathbf{u}\} + \mathbf{u}^H \mathbf{A} \mathbf{u}\} \\ &= \exp\{\text{Re}\{(\mathbf{c}_1 + i\mathbf{c}_2)^H (\mathbf{u}_1 + i\mathbf{u}_2)\}\} \\ &\quad \exp\{(\mathbf{u}_1 + i\mathbf{u}_2)^H (\mathbf{A}_1 + i\mathbf{A}_2) (\mathbf{u}_1 + i\mathbf{u}_2)\} \\ &\propto \exp\{\mathbf{c}_1^T \mathbf{u}_1 + \mathbf{c}_2^T \mathbf{u}_2 + \mathbf{u}_1^T \mathbf{A}_1 \mathbf{u}_1 + i\mathbf{u}_1^T \mathbf{A}_1 \mathbf{u}_2 + i\mathbf{u}_1^T \mathbf{A}_2 \mathbf{u}_1 \\ &\quad - \mathbf{u}_1^T \mathbf{A}_2 \mathbf{u}_2 - i\mathbf{u}_2^T \mathbf{A}_1 \mathbf{u}_1 + \mathbf{u}_2^T \mathbf{A}_1 \mathbf{u}_2 + \mathbf{u}_2^T \mathbf{A}_2 \mathbf{u}_1 + i\mathbf{u}_2^T \mathbf{A}_2 \mathbf{u}_2\} \\ &\propto \exp\{\mathbf{c}_1^T \mathbf{u}_1 + \mathbf{c}_2^T \mathbf{u}_2 + \mathbf{u}_1^T \mathbf{A}_1 \mathbf{u}_1 - \mathbf{u}_1^T \mathbf{A}_2 \mathbf{u}_2 + \mathbf{u}_2^T \mathbf{A}_1 \mathbf{u}_2 \\ &\quad + \mathbf{u}_2^T \mathbf{A}_2 \mathbf{u}_1\} \end{aligned} \quad (54)$$

Given that  $\mathbf{A}_1$  is a symmetric matrix and  $\mathbf{A}_2$  is a skew-symmetric matrix, we have

$$\mathbf{u}_2^T \mathbf{A}_2 \mathbf{u}_2 = 0, \quad \mathbf{u}_1^T \mathbf{A}_2 \mathbf{u}_1 = 0 \quad \text{and} \quad \mathbf{u}_1^T \mathbf{A}_1 \mathbf{u}_2 = \mathbf{u}_2^T \mathbf{A}_1 \mathbf{u}_1$$

Then,

$$p_{\text{CVGBL}}(\mathbf{u}) \propto \exp\{\tilde{\mathbf{c}}^T \tilde{\mathbf{u}} + \tilde{\mathbf{u}}^T \tilde{\mathbf{A}} \tilde{\mathbf{u}}\} \quad (55)$$

with

$$\tilde{\mathbf{u}}^T = [\mathbf{u}_1^T, \mathbf{u}_2^T], \quad \tilde{\mathbf{c}}^T = [\mathbf{c}_1^T, \mathbf{c}_2^T] \quad \text{and} \quad \tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & -\mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 \end{bmatrix}$$

Finally,

$$\mathbf{u} \sim \text{CVGBL}(\mathbf{c}, \mathbf{A}) \Leftrightarrow \tilde{\mathbf{u}} \sim \text{vBL}(\tilde{\mathbf{c}}, \tilde{\mathbf{A}}) \quad (56)$$

with  $\tilde{\mathbf{u}} \in \mathbb{R}^{2N}$ ,  $\tilde{\mathbf{c}} \in \mathbb{R}^{2N}$  and  $\tilde{\mathbf{A}} \in \mathbb{R}^{2N \times 2N}$  a symmetric matrix. Algorithm 5 details the generation of the unit complex random

---

**Algorithm 5:** The generation of the unit complex random vector  $\mathbf{u} \sim \text{CVGBL}(\mathbf{c}, \mathbf{A})$ .

---

```

input :  $\mathbf{c}, \mathbf{A}$ 
output :  $\mathbf{u}$ 
1 Compute the  $2N$  real unit vector  $\tilde{\mathbf{c}}$  from the vector  $\mathbf{c}$ 
2 Compute the  $2N \times 2N$  real symmetric matrix  $\tilde{\mathbf{A}}$  from the matrix  $\mathbf{A}$ 
3 Sample the real unit random vector  $\tilde{\mathbf{u}} = \text{vBL}(\tilde{\mathbf{c}}, \tilde{\mathbf{A}})$ 
4 Sample the complex unit random vector  $\mathbf{u}$  from  $\tilde{\mathbf{u}}$ 

```

---

vector  $\mathbf{u} \sim \text{CVGBL}(\mathbf{c}, \mathbf{A})$ .



### A3. Sampling a CGBL distributed matrix

The random matrix CGBL is sampled from the markov chain monte carlo method [25]. Hence, we generate samples  $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_p\})$  which converge in distribution to  $p_{\text{CGBL}}$ . The procedure is detailed in Algorithm 6 and is similar to [25].

**Algorithm 6:** The generation of the semi-unitary matrix  $\mathbf{U} \sim \text{CGBL}(\mathbf{C}, \{\mathbf{A}_p\})$ .

---

```

input   :  $\mathbf{C}, \{\mathbf{A}_p\}$ 
output  :  $\mathbf{U}$ 
initialize:  $\mathbf{U}^{(0)} \leftarrow \mathbf{U}_{\text{init}}$  (a semi-unitary matrix)
1 while stop criterion unreached do
2   for  $p \in \{1, \dots, P\}$  in random order do
3     Compute the null space  $\mathbf{N}$  of the matrix  $\mathbf{U}_{[:,p]}$ 
4     Compute the unit vector  $\mathbf{u} = \mathbf{N}^H \mathbf{U}_{[:,p]}$ 
5     Compute  $\tilde{\mathbf{c}} = \kappa \mathbf{N}^H \mathbf{c}_p$  and  $\tilde{\mathbf{A}} = \mathbf{N}^H \mathbf{A}_p \mathbf{N}$ 
6     Update the complex unit vector  $\mathbf{u} \sim \text{CVGBL}(\tilde{\mathbf{c}}, \tilde{\mathbf{A}})$ 
7     Update  $\mathbf{u}_p = \mathbf{N} \mathbf{u}$ 
8   end
9 end

```

---

### Appendix B. Details on the derivation of the MM algorithm

We derive two propositions needed for the proposed algorithms design. These propositions are generic. The first proposition is used for the update of the orthonormal basis. The second proposition is useful to derive updates w.r.t. the texture parameter  $\{\tau_k\}$  and the eigenvalues  $\{\lambda_p\}$ .

**Proposition 1.** Let  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \in \mathcal{U}_p^N$ ,  $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_p] \in \mathbb{C}^{N \times P}$  and  $\{\mathbf{Z}_p\} \subset \mathcal{H}_N^+$ . The function

$$f(\mathbf{U}) = \sum_{p=1}^P \text{Re}\{\mathbf{q}_p^H \mathbf{u}_p\} + \mathbf{u}_p^H [\mathbf{Z}_p] \mathbf{u}_p \quad (57)$$

is lower bounded at  $\mathbf{U}^t$  as

$$\begin{aligned} f(\mathbf{U}) &\geq \sum_{p=1}^P \mathbf{u}_p^H (\mathbf{Z}_p \mathbf{u}_p^t + 1/2 \mathbf{q}_p) + (\mathbf{u}_p^t \mathbf{Z}_p + 1/2 \mathbf{q}_p^H) \mathbf{u}_p + \text{const} \\ &= \text{Tr}\{\mathbf{U}^H \mathbf{H}^t\} + \text{Tr}\{\mathbf{H}^t \mathbf{U}\} + \text{const} = -\|\mathbf{U} - \mathbf{H}^t\|_F^2 + \text{const} \end{aligned} \quad (58)$$

with equality when  $\mathbf{U} = \mathbf{U}^t = [\mathbf{u}_1^t, \dots, \mathbf{u}_p^t]$  and  $\mathbf{H}^t = 1/2 \mathbf{Q} + [\mathbf{Z}_1 \mathbf{u}_1^t, \dots, \mathbf{Z}_p \mathbf{u}_p^t]$ . The surrogate function reads as

$$f(\mathbf{U}|\mathbf{U}^t) = -\|\mathbf{U} - \mathbf{H}^t\|_F^2 \quad (59)$$

Maximizing the above function under unitary constraints is equivalent to solve

$$\begin{aligned} &\underset{\mathbf{U}}{\text{minimize}} && \|\mathbf{U} - \mathbf{H}^t\|_F^2 \\ &\text{subject to} && \mathbf{U}^H \mathbf{U} = \mathbf{I}_p \end{aligned} \quad (60)$$

which is an orthogonal Procrustes problem [22] that has a unique solution given as

$$\hat{\mathbf{U}}^{t+1} = \mathcal{P}_{\text{Proc}}(\mathbf{H}^t) \quad (61)$$

where the projection onto the set  $\mathcal{U}_p^N$  is denoted by the operator

$$\begin{aligned} \mathcal{P}_{\text{Proc}} : \mathbb{C}^{N \times P} &\longrightarrow \mathcal{U}_p^N \\ \mathbf{Y} \stackrel{\text{TSVD}}{=} \mathbf{U} \mathbf{D} \mathbf{V}^H &\longmapsto \mathcal{P}_{\text{Proc}}\{\mathbf{Y}\} = \mathbf{U} \mathbf{V}^H \end{aligned} \quad (62)$$

with  $\stackrel{\text{TSVD}}{=}$  defines the thin-singular value decomposition of a given matrix.

**Proposition 2.** Let us consider  $a, \{b_i\}$  and  $\{s_i\}$  where  $a > 0, b_i > 0$  and  $s_i > 0, \forall i \in [[1, I]]$ . The objective function

$$g(a) = \sum_{i=1}^I \left( \ln(ab_i + \sigma^2) - \frac{ab_i s_i}{ab_i + \sigma^2} \right) \quad (63)$$

is upper bounded by

$$g(a) \leq A \ln(Ba + C) - D \ln(a) \quad (64)$$

with

$$\begin{cases} \theta_i^t = 1 + s_i \frac{a^t b_i}{a^t b_i + \sigma^2} & A = \sum_{i=1}^I \theta_i^t \\ B = \frac{\sum_{i=1}^I \frac{\theta_i^t b_i}{b_i a^t + \sigma^2}}{\sum_{i=1}^I \theta_i^t} & C = \sigma^2 \frac{\sum_{i=1}^I \frac{\theta_i^t}{b_i a^t + \sigma^2}}{\sum_{i=1}^I \theta_i^t} \\ D = \sum_{i=1}^I s_i \frac{a^t b_i}{a^t b_i + \sigma^2} \end{cases}$$

then, the surrogate function reduces to

$$g(a|a^t) = A \ln(Ba + C) - D \ln(a) \quad (65)$$

with equality at  $a = a^t$ . The minimizer of the above function under positivity constraint is given as

$$a^{t+1} = \frac{DC}{B(A-D)} = \frac{1}{I} \frac{\left( \sum_{i=1}^I s_i \frac{a^t b_i}{a^t b_i + \sigma^2} \right) \left( \sum_{i=1}^I \sigma^2 \frac{\theta_i^t}{a^t b_i + \sigma^2} \right)}{\sum_{i=1}^I \frac{\theta_i^t b_i}{a^t b_i + \sigma^2}} \quad (66)$$

**Proof.** The proof of Propositions 1 and 2 are similar to [22].  $\square$

### References

- [1] I. Jolliffe, *Principal Component Analysis*, Springer, 2011.
- [2] R.O. Schmidt, Multiple emitter location and signal parameter estimation, *IEEE Trans. Antennas Propag.* 34 (3) (1986) 276–280.
- [3] J. Liu, M.S. Shbat, V. Tuzlukov, Interference cancellation and DOA estimation by generalized receiver applying LMS and MUSIC algorithms, *Prog. Electromagn. Res. Symp. Proc.* (2013) 187–190.
- [4] R. Grover, D.A. Pados, M.J. Medley, Subspace direction finding with an auxiliary-vector basis, *IEEE Trans. Signal Process.* 55 (2) (2007) 758–763.
- [5] R.C.D. Lamare, R. Sampaio-Neto, Reduced-rank adaptive filtering based on joint iterative optimization of adaptive filters, *IEEE Signal Process. Lett.* 14 (12) (2007) 980–983.
- [6] M. Haardt, M. Pesavento, F. Röemer, M.N. El Korso, Subspace methods and exploitation of special array structures, in: M. Viberg (Ed.), *Electronic Reference in Signal Processing: Array and Statistical Signal Processing*, vol. 3, Academic Press Library in Signal Processing, Elsevier Ltd., 2014, pp. 651–717. ISBN 978-0-12-411597-2. Chapter 2.15.
- [7] O. Besson, N. Dobigeon, J. Tournet, Joint Bayesian estimation of close subspaces from noisy measurements, *IEEE Signal Process. Lett.* 21 (2014) 168–171.
- [8] A. Srivastava, A Bayesian approach to geometric subspace estimation, *IEEE Trans. Signal Process.* 48 (5) (2000) 1390–1400.
- [9] O. Besson, N. Dobigeon, J.Y. Tournet, Minimum mean square distance estimation of a subspace, *IEEE Trans. Signal Process.* 59 (12) (2011) 5709–5720.
- [10] Y. Chikuse, *Statistics on Special Manifold*, Springer-Verlag, New York, 2003.
- [11] O. Besson, N. Dobigeon, J.Y. Tournet, CS decomposition based Bayesian subspace estimation, *IEEE Trans. Signal Process.* 60 (8) (2012) 4210–4218.
- [12] C. Elvira, P. Chainais, N. Dobigeon, Bayesian nonparametric principal component analysis, *arXiv:1709.05667* (2017).
- [13] E. Clement, C. Pierre, D. Nicolas, Bayesian nonparametric subspace estimation, in: *Acoustics, Speech and Signal Processing (ICASSP)*, 2017 IEEE International Conference on, 2017, pp. 2247–2251.
- [14] E. Ollila, D.E. Tyler, V. Koivunen, H.V. Poor, Complex elliptically symmetric distributions: survey, new results and applications, *IEEE Trans. Signal Process.* 60 (11) (2012) 5597–5625.
- [15] A. Breloy, Y. Sun, P. Babu, G. Ginolhac, D.P. Palomar, F. Pascal, A robust signal subspace estimator, in: *Statistical Signal Processing Workshop (SSP)*, IEEE, 2016, pp. 1–4.
- [16] O. Esa, D.E. Tyler, V. Koivunen, H.P. Vincent, Compound gaussian clutter modeling with an inverse gaussian texture distribution, *IEEE Signal Process. Lett.* 19 (12) (2012) 876–879.
- [17] M. Greco, F. Gini, M. Rangaswamy, Statistical analysis of measured polarimetric clutter data at different range resolutions, *IEE Proc.-RADAR Sonar Navig.* 153 (6) (2006) 473–481.
- [18] G. Ginolhac, P. Forster, Approximate distribution of the low-rank adaptive normalized matched filter test statistic under the null hypothesis, *IEEE Trans. Aerosp. Electron.Syst.* 52 (4) (2016) 2016–2023.

- [19] M. Rangaswamy, F.C. Lin, K.R. Gerlach, Robust adaptive signal processing methods for heterogeneous RADAR clutter scenarios, *Signal Process.* 84 (9) (2004) 1653–1665.
- [20] K. Greenewald, E. Zelnio, A.O.H. III, Kronecker PCA based robust SAR STAP, *arXiv:1501.07481* (2015).
- [21] A. Breloy, G. Ginolhac, F. Pascal, P. Forster, Clutter subspace estimation in low rank heterogeneous noise context, *IEEE Trans. Signal Process.* 63 (9) (2015) 2173–2182.
- [22] Y. Sun, A. Breloy, P. Babu, D.P. Palomar, F. Pascal, G. Ginolhac, Low-complexity algorithms for low rank clutter parameters estimation in RADAR systems, *IEEE Trans. Signal Process.* 64 (8) (2016) 1986–1998.
- [23] O. Besson, Bounds for a mixture of low-rank compound-gaussian and white gaussian noises, *IEEE Trans. Signal Process.* 64 (21) (2016) 5723–5732.
- [24] K.V. Mardia, P.E. Jupp, Distributions on spheres, *Dir. Stat.* 898 (2000) 182.
- [25] P.D. Hoff, Simulation of the matrix Bingham–Von Mises–Fisher distribution with applications to multivariate and relational data, *J. Comput. Graph. Stat.* 18 (2) (2009) 438–456.
- [26] J.T. Kent, A.M. Ganeiber, K.V. Mardia, A new method to simulate the Bingham and related distributions in directional data analysis with applications, *arXiv:1310.8110* (2013).
- [27] A.T. Wood, Estimation of the concentration parameters of the fisher matrix distribution on  $S^1$  (3) and the Bingham distribution on  $S^1$ , *Aust. J. Stat.* 35 (1) (1993) 69–79.
- [28] N.A. Goodman, J.M. Stiles, On clutter rank observed by arbitrary arrays, *IEEE Trans. Signal Process.* 55 (1) (2007) 178–186.
- [29] P. Stoica, Y. Selen, Model-order selection: a review of information criterion rules, *IEEE Signal Process. Mag.* 21 (4) (2004) 36–47.
- [30] L. Huang, H.C. So, et al., Source enumeration via MDL criterion based on linear shrinkage estimation of noise subspace covariance matrix., *IEEE Trans. Signal Process.* 61 (19) (2013) 4806–4821.
- [31] E. Terreaux, J.P. Ovarlez, F. Pascal, Robust model order selection in large dimensional elliptically symmetric noise, *arXiv:1710.06735* (2017).
- [32] R.S. Raghavan, Statistical interpretation of a data adaptive clutter subspace estimation algorithm, *IEEE Trans. Aerosp. Electron.Syst.* 48 (2) (2012) 1370–1384.
- [33] J.H. Manton, Optimization algorithms exploiting unitary constraints, *IEEE Trans. Signal Process.* 50 (3) (2002) 635–650.
- [34] Y. Sun, P. Babu, D.P. Palomar, Majorization–minimization algorithms in signal processing communications and machine learning, *IEEE Trans. Signal Process.* 65 (2016) 794–816.
- [35] K. Benidis, Y. Sun, P. Babu, D.P. Palomar, Orthogonal sparse PCA and covariance estimation via procrustes reformulation, *IEEE Trans. Signal Process.* 64 (23) (2016) 6211–6226.