Bayesian estimation of a subspace

Olivier Besson*, Nicolas Dobigeon† and Jean-Yves Tourneret†
*University of Toulouse, ISAE, Department Electronics Optronics Signal, Toulouse, France
†University of Toulouse, IRIT/ENSEEIHT, Signal and Communications Group, Toulouse, France

***Abstract*—We consider the problem of subspace estimation in a Bayesian setting. First, we revisit the conventional minimum mean square error (MSE) estimator and explain why the MSE criterion may not be fully suitable when operating in the Grassmann manifold. As an alternative, we propose to carry out subspace estimation by minimizing the mean square distance between the true subspace U and its estimate, where the considered distance is a natural metric on the Grassmann manifold. We show that the resulting estimator is no longer the posterior mean of U but entails computing the principal eigenvectors of the posterior mean of UU^T. Illustrative examples involving a linear Gaussian model for the data and a Bingham or von Mises Fisher prior distribution for U are presented. In the former case the minimum mean square distance (MMSD) estimator is obtained in closed-form while, in the latter case, a Markov chain Monte Carlo method is used to approximate the MMSD estimator. The method is shown to provide accurate estimates even when the number of samples is lower than the dimension of U. Finally, an application to hyperspectral imagery is presented.**

***Index Terms*—Subspace estimation, Bayesian inference, minimum distance error.**

I. Problem statement

In many signal processing applications, the signals of interest are confined to a low-dimensional subspace of the entire observation space. Therefore subspace estimation plays a central role in recovering these signals with maximum accuracy. In a frequentist approach, and under the classical model $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{S} + \boldsymbol{N}$ (where $\boldsymbol{Y}$ stands for the $N \times K$ observation matrix, $\boldsymbol{U}$ is an $N \times p$ matrix, with $p < N$, whose columns span the p-dimensional subspace of interest, $\boldsymbol{S}$ is the $p \times K$ waveform matrix and $\boldsymbol{N}$ stands for noise), an almost inevitable solution to estimate the range $\mathcal{R}(\boldsymbol{U})$ is to resort to the singular value decomposition (SVD) of the data matrix $\boldsymbol{Y}$. The p principal left singular vectors of $\boldsymbol{Y}$ provide very accurate estimates of a basis for $\mathcal{R}(\boldsymbol{U})$, and have been used successfully, e.g., in estimating the frequencies of damped exponentials or the directions of arrival of multiple plane waves [1], [2]. However, the SVD can incur some performance loss in two main cases. The first case concerns the low signal-to-noise ratio (SNR) regime for which there exists a non negligible probability of subspace swap or subspace leakage [3], [4] leading to very inaccurate estimates. SVD-based estimates may also be unreliable when the number of samples K is small. Moreover, subspace estimation is no longer possible when K becomes lower than the subspace dimension p: indeed, $\boldsymbol{Y}$ is at most of rank K and information is lacking about how to complement $\mathcal{R}(\boldsymbol{Y})$ in order to estimate $\mathcal{R}(\boldsymbol{U})$.

Under such circumstances, a Bayesian approach might be helpful as it enables one to assist estimation by providing some statistical information about $\boldsymbol{U}$. We investigate such an approach herein by assigning an appropriate prior distribution to the unknown matrix $\boldsymbol{U}$.

II. Minimum mean square distance estimation

In this section, we consider the specificity of the problem (namely that the quantity we seek to estimate is a subspace) to introduce an alternative to the conventional minimum mean square error (MMSE) estimator. The MMSE estimator consists in minimizing the average Euclidian distance between the estimate $\hat{\boldsymbol{\theta}}$ of the actual parameter vector $\boldsymbol{\theta}$. This distance is natural in an Euclidian space, however it may not be the more natural metric in the Grassmann manifold $G_{N,p}$, that is the set of p-dimensional subspaces in $\mathbb{R}^N$ [5]. Since our primary goal is to estimate the range space $\mathcal{R}(\boldsymbol{U})$ of $\boldsymbol{U}$, it is thus logical to wonder whether we could turn to a more natural metric on the Grassman manifold. In fact, the distance between two subspaces $\mathcal{R}(\boldsymbol{U}_1)$ and $\mathcal{R}(\boldsymbol{U}_2)$ is given by $\left(\sum_{k=1}^{p} \theta_k^2\right)^{1/2}$ [5], where θ_k are the principal angles between these subspaces. The latter can be obtained by SVD of $\boldsymbol{U}_2^T\boldsymbol{U}_1$ which writes $\boldsymbol{U}_2^T\boldsymbol{U}_1 = \boldsymbol{X}\,\mathrm{diag}\left(\cos\theta_1, \cdots, \cos\theta_p\right)\boldsymbol{Z}^T$ where $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ denote orthonormal bases for the subspaces [6]. Therefore, it seems more adequate to minimize the natural distance between the subspaces spanned by $\hat{\boldsymbol{U}}$ and $\boldsymbol{U}$, rather than minimizing $\left\|\hat{\boldsymbol{U}} - \boldsymbol{U}\right\|_F^2$ as the MMSE estimator does. Although this is the most intuitively appealing method, it faces the drawback that the cosines of the angles and not the angles themselves emerge naturally from the SVD. In order to circumvent this pitfall, we instead propose to minimize the average (square) distance between the projection matrices $\hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^T$ and $\boldsymbol{U}\boldsymbol{U}^T$, viz $d^2\left(\hat{\boldsymbol{U}}, \boldsymbol{U}\right) = \left\|\hat{\boldsymbol{U}}\hat{\boldsymbol{U}}^T - \boldsymbol{U}\boldsymbol{U}^T\right\|_F^2$. This distance makes sense [5], [6] and is related to the angles θ_k between $\mathcal{R}\left(\hat{\boldsymbol{U}}\right)$ and $\mathcal{R}(\boldsymbol{U})$ through $d^2\left(\hat{\boldsymbol{U}}, \boldsymbol{U}\right) = 2\sum_{k=1}^{p} \sin^2\theta_k$. Since $d^2\left(\hat{\boldsymbol{U}}, \boldsymbol{U}\right) = 2p - 2\mathrm{Tr}\left\{\hat{\boldsymbol{U}}^T\boldsymbol{U}\boldsymbol{U}^T\hat{\boldsymbol{U}}\right\}$, we thus define the minimum mean-square distance (MMSD) estimator of $\boldsymbol{U}$ as

$$\hat{\boldsymbol{U}}_{\mathrm{mmsd}} \triangleq \arg\max_{\hat{\boldsymbol{U}}} \mathsf{E}\left\{\mathrm{Tr}\left\{\hat{\boldsymbol{U}}^T\boldsymbol{U}\boldsymbol{U}^T\hat{\boldsymbol{U}}\right\}\right\}. \qquad (1)$$

Observing that

$$\mathsf{E}\left\{\mathrm{Tr}\left\{\hat{\boldsymbol{U}}^T\boldsymbol{U}\boldsymbol{U}^T\hat{\boldsymbol{U}}\right\}\right\} = \int\left[\int \mathrm{Tr}\left\{\hat{\boldsymbol{U}}^T\boldsymbol{U}\boldsymbol{U}^T\hat{\boldsymbol{U}}\right\} p\left(\boldsymbol{U}|\boldsymbol{Y}\right) d\boldsymbol{U}\right] p\left(\boldsymbol{Y}\right) d\boldsymbol{Y} \quad (2)$$

the MMSD estimator can be rewritten as

$$\begin{aligned}\hat{\boldsymbol{U}}_{\text{mmsd}} &= \arg\max_{\hat{\boldsymbol{U}}} \int \mathrm{Tr}\left\{\hat{\boldsymbol{U}}^T\boldsymbol{U}\boldsymbol{U}^T\hat{\boldsymbol{U}}\right\} p\left(\boldsymbol{U}|\boldsymbol{Y}\right) d\boldsymbol{U} \\ &= \arg\max_{\hat{\boldsymbol{U}}} \mathrm{Tr}\left\{\hat{\boldsymbol{U}}^T\left[\int \boldsymbol{U}\boldsymbol{U}^T p\left(\boldsymbol{U}|\boldsymbol{Y}\right) d\boldsymbol{U}\right]\hat{\boldsymbol{U}}\right\}. \quad (3)\end{aligned}$$

It follows that the MMSD estimate of the subspace spanned by $\boldsymbol{U}$ is given by the p largest eigenvectors of the matrix $\int \boldsymbol{U}\boldsymbol{U}^T p\left(\boldsymbol{U}|\boldsymbol{Y}\right) d\boldsymbol{U}$, which we denote as

$$\hat{\boldsymbol{U}}_{\text{mmsd}} = \mathcal{P}_p\left\{\int \boldsymbol{U}\boldsymbol{U}^T p\left(\boldsymbol{U}|\boldsymbol{Y}\right) d\boldsymbol{U}\right\}. \quad (4)$$

MMSD estimation thus amounts to find the best rank-p approximation to the *posterior mean of the projection matrix* $\boldsymbol{U}\boldsymbol{U}^T$ on $\mathcal{R}\left(\boldsymbol{U}\right)$.

Remark 1 A few comments about the MMSE estimator and its difference compared to the MMSD estimator are in order. The MMSE approach which would entail calculating the posterior mean of $\boldsymbol{U}$, viz $\int \boldsymbol{U} p\left(\boldsymbol{U}|\boldsymbol{Y}\right) d\boldsymbol{U}$. First note that this matrix is not necessarily unitary: however, its range space can be used to estimate $\mathcal{R}\left(\boldsymbol{U}\right)$. A more important observation is that computing the posterior mean of $\boldsymbol{U}$ may not be meaningful, in particular when the posterior $p\left(\boldsymbol{U}|\boldsymbol{Y}\right)$ depends on $\boldsymbol{U}$ only through $\boldsymbol{U}\boldsymbol{U}^T$, see next section for an example. In such a case, $p\left(\boldsymbol{U}|\boldsymbol{Y}\right)$ is left unchanged if $\boldsymbol{U}$ is right-multiplied by any $p \times p$ unitary matrix $\boldsymbol{Q}$: therefore averaging $\boldsymbol{U}$ over $p\left(\boldsymbol{U}|\boldsymbol{Y}\right)$ does not make sense while computing (4) is relevant. On the other hand, if $p\left(\boldsymbol{U}|\boldsymbol{Y}\right)$ depends on $\boldsymbol{U}$ directly, then computing the posterior mean of $\boldsymbol{U}$ can be investigated. The two situations will be illustrated below.

III. MMSD estimation for the linear model

In order to illustrate how the previous theory can be used, we consider the example of a linear Gaussian model (conditioned on $\boldsymbol{U}$) and two different prior distributions for $\boldsymbol{U}$.

A. Likelihood

Let us consider that the data follow the model $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{S} + \boldsymbol{N}$ where the columns of $\boldsymbol{N}$ are independent and identically distributed (i.i.d.) Gaussian vectors with zero-mean and covariance matrix $\sigma_n^2\boldsymbol{I}$. We first assume that no knowledge about $\boldsymbol{S}$ is available and hence its prior distribution is set to $\pi\left(\boldsymbol{S}\right) \propto 1$ where $\propto$ means proportional to. Therefore, conditioned on $\boldsymbol{U}$, the probability density function of $\boldsymbol{Y}$ is given by

$$\begin{aligned}p\left(\boldsymbol{Y}|\boldsymbol{U}\right) &= \int p\left(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{S}\right)\pi\left(\boldsymbol{S}\right) d\boldsymbol{S} \\ &\propto \int \mathrm{etr}\left\{-\frac{1}{2\sigma_n^2}\left(\boldsymbol{Y}-\boldsymbol{U}\boldsymbol{S}\right)^T\left(\boldsymbol{Y}-\boldsymbol{U}\boldsymbol{S}\right)\right\} d\boldsymbol{S} \\ &\propto \mathrm{etr}\left\{-\frac{1}{2\sigma_n^2}\boldsymbol{Y}^T\boldsymbol{Y} + \frac{1}{2\sigma_n^2}\boldsymbol{Y}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{Y}\right\} \quad (5)\end{aligned}$$

where $\mathrm{etr}\left\{.\right\}$ stands for the exponential of the trace of the matrix between braces.

B. Prior distributions

In order to complete our statistical model we need now to set prior distributions for the matrix $\boldsymbol{U}$. There exist only a few distributions on the Stiefel or Grassmann manifolds [7], [8], the most popular being the Bingham or von Mises Fisher (vMF) distributions which have proven to be relevant in a number of applications including meteorology, biology, medicine, image analysis (see [7] and references therein), modeling of multipath communications channels [9] and shape analysis [10]. These distributions depend on a matrix whose range space is "close" to that of $\boldsymbol{U}$, along with a concentration parameter that rules the distance between the subspaces. More specifically, we assume that $\mathcal{R}\left(\boldsymbol{U}\right)$ is close to a given subspace spanned by the columns of an orthonormal matrix $\bar{\boldsymbol{U}}$ and we consider the following prior distributions for $\boldsymbol{U}$ [7], [8]:

$$\text{Bingham: } \pi_{\text{B}}\left(\boldsymbol{U}\right) \propto \mathrm{etr}\left\{\kappa\boldsymbol{U}^T\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T\boldsymbol{U}\right\} \quad (6)$$

$$\text{von Mises-Fisher: } \pi_{\text{vMF}}\left(\boldsymbol{U}\right) \propto \mathrm{etr}\left\{\kappa\boldsymbol{U}^T\bar{\boldsymbol{U}}\right\}. \quad (7)$$

The prior knowledge $\bar{\boldsymbol{U}}$ can either stem from some available model or can be obtained from the data itself, as in the hyperspectral imagery case developed in section V. In (6) or (7), κ is a concentration parameter: the larger κ the more concentrated around $\bar{\boldsymbol{U}}$ are the subspaces $\boldsymbol{U}$. Observe that $\pi_{\text{B}}\left(\boldsymbol{U}\right)$ depends on the projection matrix $\boldsymbol{U}\boldsymbol{U}^T$ and not on $\boldsymbol{U}$ only: hence only $\mathcal{R}\left(\boldsymbol{U}\right)$ and $\mathcal{R}\left(\bar{\boldsymbol{U}}\right)$ are close, at least for large κ. On the other hand, with a vMF prior, the matrices $\boldsymbol{U}$ and $\bar{\boldsymbol{U}}$ themselves are close. Therefore, the vMF prior distribution is more conservative than the Bingham distribution. In order to illustrate the influence of κ, Figure 1 displays the average fraction of energy (AFE) of $\boldsymbol{U}$ in $\mathcal{R}\left(\bar{\boldsymbol{U}}\right)$ defined as

$$\mathrm{AFE}\left(\boldsymbol{U},\bar{\boldsymbol{U}}\right) = \mathsf{E}\left\{\mathrm{Tr}\left\{\boldsymbol{U}^T\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T\boldsymbol{U}\right\}/p\right\} \quad (8)$$

for both distributions. As can be observed, the two distributions allow the distance between $\boldsymbol{U}$ and $\bar{\boldsymbol{U}}$ to be set in a rather flexible way. However, although they may have the same AFE for small to moderate κ, the distributions of the angles between $\mathcal{R}\left(\boldsymbol{U}\right)$ and $\mathcal{R}\left(\bar{\boldsymbol{U}}\right)$ are different, see [11] for details.

C. Posterior distributions and MMSD estimation

We are now in a position to derive the posterior distribution $p\left(\boldsymbol{U}|\boldsymbol{Y}\right)$. Let us start with the Bingham prior. Using (5)

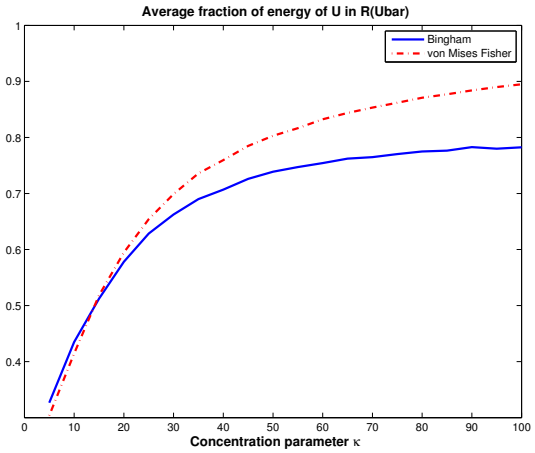

Fig. 1. Average fraction of energy of $\boldsymbol{U}$ in $\mathcal{R}\left(\bar{\boldsymbol{U}}\right)$ versus κ. $N = 20$, $p = 5$.

along with (6), it follows that the posterior distribution of $\boldsymbol{U}$, conditioned on $\boldsymbol{Y}$ is given by

$$p\left(\boldsymbol{U}|\boldsymbol{Y}\right) \propto \text{etr}\left\{\boldsymbol{U}^T\left[\kappa\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T + \frac{1}{2\sigma_n^2}\boldsymbol{Y}\boldsymbol{Y}^T\right]\boldsymbol{U}\right\} \quad (9)$$

which is recognized as a Bingham distribution with parameter matrix $\kappa\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T + \frac{1}{2\sigma_n^2}\boldsymbol{Y}\boldsymbol{Y}^T$. In this case, the MMSD estimator can be obtained in **closed-form** as [11]

$$\hat{\boldsymbol{U}}_{\text{mmsd-LM-B}} = \mathcal{P}_p\left\{\kappa\bar{\boldsymbol{U}}\bar{\boldsymbol{U}}^T + \frac{1}{2\sigma_n^2}\boldsymbol{Y}\boldsymbol{Y}^T\right\}. \quad (10)$$

The MMSD amounts to computing the principal subspace from a combination of the prior knowledge and the information brought by the data. Note also that in this case, it coincides with the maximum a posteriori (MAP) estimator of $\boldsymbol{U}$.

Let us now consider the vMF prior distribution. From (5) and (7), the posterior distribution becomes

$$p\left(\boldsymbol{U}|\boldsymbol{Y}\right) \propto \text{etr}\left\{\kappa\boldsymbol{U}^T\bar{\boldsymbol{U}} + \frac{1}{2\sigma_n^2}\boldsymbol{U}^T\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{U}\right\} \quad (11)$$

which is referred to as the Bingham-von-Mises-Fisher (BMF) distribution [12] denoted as $\boldsymbol{U}|\boldsymbol{Y} \sim \text{BMF}\left(\boldsymbol{Y}\boldsymbol{Y}^T, \frac{1}{2\sigma_n^2}\boldsymbol{I}, \kappa\bar{\boldsymbol{U}}\right)$. Despite the fact that this distribution is known, it appears that the integral in (4) needed to derive the MMSD estimator does not admit an analytic expression. Therefore, the MMSD estimator cannot be computed in closed-form. In order to remedy this problem, a Markov chain Monte Carlo simulation method can be advocated to generate a large number of matrices $\boldsymbol{U}^{(n)}$ drawn from (11), and to approximate (4) by

$$\hat{\boldsymbol{U}}_{\text{mmsd-LM-vMF}} \simeq \mathcal{P}_p\left\{\frac{1}{N_r}\sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r}\boldsymbol{U}^{(n)}\boldsymbol{U}^{(n)^T}\right\} \quad (12)$$

where N_{bi} is the number of burn-in samples and N_r is the number of samples used to approximate the estimator. Towards this end, we can make use of the efficient Gibbs sampling scheme proposed by Hoff [12] to generate random unitary matrices drawn from a BMF distribution, with appropriate modifications, see Appendix B of [11]. In summary, the MMSD estimator for the vMF prior can be implemented via the Gibbs sampling scheme described in Table I.

TABLE I
IMPLEMENTATION OF THE MMSD ESTIMATOR WITH A GIBBS SAMPLER FOR THE VMF PRIOR DISTRIBUTION.

Input: initial value $\boldsymbol{U}^{(0)}$
1: **for** $n = 1, \cdots, N_{bi} + N_r$ **do**
2: sample $\boldsymbol{U}^{(n)}$ from BMF $\left(\boldsymbol{Y}\boldsymbol{Y}^T, \frac{1}{2\sigma_n^2}\boldsymbol{I}, \kappa\bar{\boldsymbol{U}}\right)$ in (11).
3: **end for**
4: compute $\hat{\boldsymbol{U}}_{\text{mmsd-LM-vMF}}$ as in (12)

Remark 2 When $\boldsymbol{U}$ has a Bingham prior distribution, the MAP estimator of $\boldsymbol{U}$ and the MMSD estimator were shown to coincide. This is no longer the case when $\boldsymbol{U}$ has a vMF prior distribution, and hence a BMF posterior distribution. The mode of the latter is not known in closed-form either. Consequently, the MAP estimator can be approximated by selecting, among the matrices $\boldsymbol{U}^{(n)}$ generated by the Gibbs sampler of Table I, the matrix which results in the largest value of the posterior distribution in (11).

IV. SIMULATIONS

In this section we illustrate the performance of the approach developed above through Monte Carlo simulations. In all simulations $N = 20$, $p = 5$ and $\kappa = 20$. The matrix $\boldsymbol{S}$ is generated from a Gaussian distribution with zero-mean and covariance matrix $\sigma_s^2\boldsymbol{I}$ and the signal-to-noise ratio is defined as $SNR = 10\log_{10}\left(\sigma_s^2/\sigma_n^2\right)$. The matrix $\boldsymbol{U}$ is generated from the Bingham or vMF distributions with $\bar{\boldsymbol{U}} = \begin{bmatrix}\boldsymbol{I}_p & \boldsymbol{0}\end{bmatrix}^T$. The number of burn-in iterations in the Gibbs sampler is set to $N_{bi} = 10$ and $N_r = 1000$. The MMSD estimator (4) is compared with the MAP estimator, the MMSE estimator, the usual SVD-based estimator and the estimator $\hat{\boldsymbol{U}} = \bar{\boldsymbol{U}}$ (denoted Ubar in the figures) that discards the available data and use

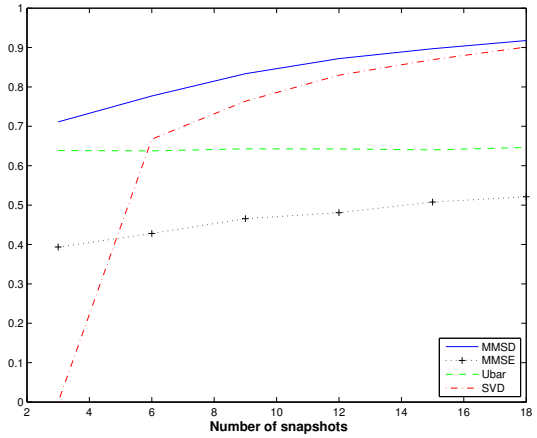

Fig. 2. Average fraction of energy of $\hat{\boldsymbol{U}}$ in $\mathcal{R}\left(\boldsymbol{U}\right)$ versus K for a Bingham prior. $N = 20$, $p = 5$, $\kappa = 20$ and $SNR = 5$dB.

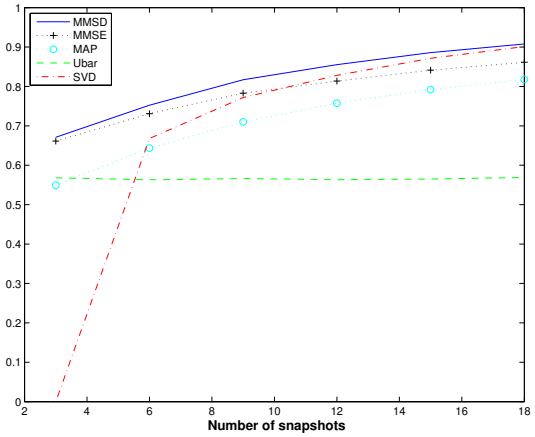

Fig. 3. Average fraction of energy of $\hat{\boldsymbol{U}}$ in $\mathcal{R}(\boldsymbol{U})$ versus K for a vMF prior. $N = 20$, $p = 5$, $\kappa = 20$ and $SNR = 5$dB.

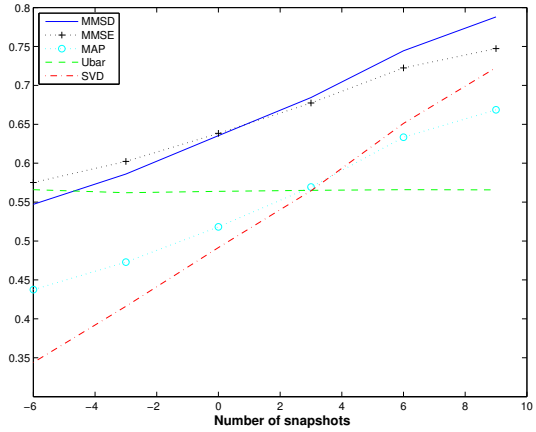

Fig. 5. Average fraction of energy of $\hat{\boldsymbol{U}}$ in $\mathcal{R}(\boldsymbol{U})$ versus SNR for vMF prior. $N = 20$, $p = 5$, $\kappa = 20$ and $K = 5$.

only the a priori knowledge. The estimators are evaluated in terms of the AFE of $\hat{\boldsymbol{U}}$ in $\mathcal{R}(\boldsymbol{U})$, i.e., AFE $\left(\hat{\boldsymbol{U}}, \boldsymbol{U}\right)$. The influence of K and SNR is displayed in Figures 2 to 5.

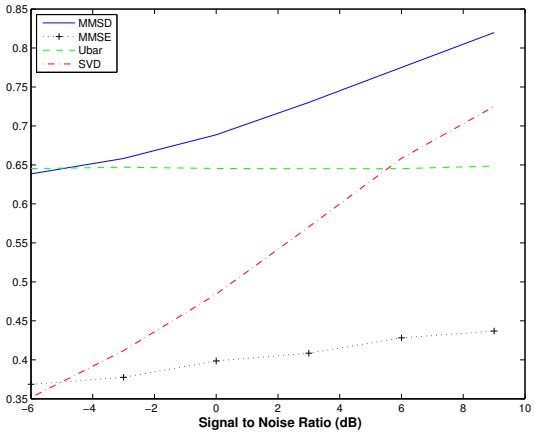

Fig. 4. Average fraction of energy of $\hat{\boldsymbol{U}}$ in $\mathcal{R}(\boldsymbol{U})$ versus SNR for a Bingham prior. $N = 20$, $p = 5$, $\kappa = 20$ and $K = 5$.

The following comments can be made. The MMSD approach enables one to improve over the estimator that uses $\bar{\boldsymbol{U}}$ even when the SNR is very low and the improvement is more pronounced when K increases. It also performs better than the SVD, especially at low SNR. Note also that the MMSD estimator can be computed when $K < p$, which is not the case with the SVD. The MMSE estimator does not perform well with the Bingham prior while it provides good performance with the vMF prior. This corroborates the discussion of remark 1, as the posterior distribution $p(\boldsymbol{U}|\boldsymbol{Y})$ in (9) depends on $\boldsymbol{U}\boldsymbol{U}^T$ only, while $p(\boldsymbol{U}|\boldsymbol{Y})$ in (11) also depends on $\boldsymbol{U}$.

V. Application to hyperspectral data

In this section, we apply the proposed subspace estimation procedure to multi-band image analysis. Hyperspectral imagery has received considerable attention because of its great interest for various purposes, e.g., agriculture monitoring, mineral mapping, military concerns. A primordial issue when analyzing such image is spectral unmixing [13] which aims at decomposing an observed pixel $\boldsymbol{y}_\ell$ into a set of $R = p + 1$ reference signatures, $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R$ (called *endmembers*) and to retrieve the respective proportions of these signatures (or *abundances*) $a_{1,\ell}, \ldots, a_{R,\ell}$ in the pixel [13]. So far, the most widely admitted mixing model is linear

$$\boldsymbol{y}_\ell = \sum_{r=1}^{R} a_{r,\ell} \boldsymbol{m}_r \tag{13}$$

where $\boldsymbol{y}_\ell \in \mathbb{R}^N$ is the pixel spectrum measured in N spectral bands, $\boldsymbol{m}_r \in \mathbb{R}^N$ $(r = 1, \ldots, R)$ are the R endmember spectra and $a_{r,\ell}$ $(r = 1, \ldots, R)$ are their corresponding abundances which satisfy $a_{r,\ell} \geq 0$ and $\sum_{r=1}^{R} a_{r,\ell} = 1$. If we consider L pixels $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L$ of an hyperspectral image following the linear mixing model (LMM) in (13) then the dataset formed by these L pixels lies in a p-dimensional polytope of $\mathbb{R}^N$ whose vertices are the endmembers $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R$ to be recovered [14]. Therefore, the centered data matrix belongs to a p-dimensional subspace whose estimation is generally achieved through a standard dimension reduction technique such as principal component analysis (PCA) [14]. However, it is well known that linearity is a simplifying assumption and does not hold anymore in several contexts, for instance with scenes including mixtures of minerals or vegetation. It thus becomes of interest to assess the validity of the linear model. Towards this end, we propose to estimate the subspace locally, i.e., within a few pixels: studying the evolution of these subspaces along the whole image can give an indication of the degree of non-linearity. The MMSD estimator, which performs subspace estimation with a limited number of samples, can thus be used to achieve this goal. The matrix $\bar{\boldsymbol{U}}$, which summarizes the prior knowledge in the MMSD approach, can be obtained from a PCA on the whole image. Then, for each pixel $\boldsymbol{y}_\ell$, we compute the MMSD estimator of the $N \times p$ matrix $\boldsymbol{U}_\ell$, whose columns are supposed to span the subspace containing $\boldsymbol{y}_\ell$ and its $K - 1$-nearest neighbors ($K = 4$ in the examples below).

The distance $d^2\left(\hat{\boldsymbol{U}}_\ell, \bar{\boldsymbol{U}}\right)$ is then evaluated. If the linear model was in force and in the noise-free case, all pixels should lie in $\mathcal{R}\left(\bar{\boldsymbol{U}}\right)$ and this distance should be zero: therefore examining this distance enables one to reveal possible deviations from the linear assumption.

In a first experiment, we investigate the estimation of $\boldsymbol{U}_\ell$ when the image pixels are non-linear functions of the abundances, i.e., the data is generated as $\boldsymbol{y}_\ell = \sum_{r=1}^{3} a_{r,\ell}\boldsymbol{m}_r + \gamma_{1,2,\ell} a_{1,\ell} a_{2,\ell} \boldsymbol{m}_1 \odot \boldsymbol{m}_2$, where $\gamma_{1,2,\ell} = 0$ for 75% of the image and $\gamma_{1,2,\ell}$ is varied from 0 to 1 in 25% of the image, see Figure 6. The MMSD estimator, with varying value of $\eta = 2\sigma_n^2\kappa$, is compared with the usual SVD-based estimator which corresponds to $\eta = 0$. As can be seen, the distance between $\hat{\boldsymbol{U}}_\ell$ and $\bar{\boldsymbol{U}}$ enables one to quantify the degree of non linearity.

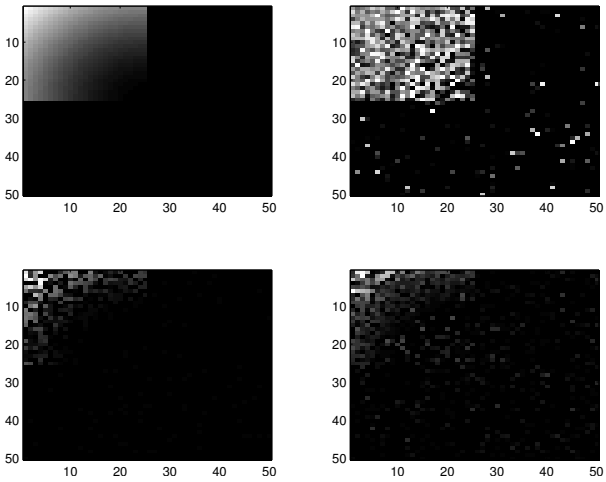

Fig. 6. Top, left: non-linearity coefficients $\gamma_{1,2}$. Top, right: distance between $\bar{\boldsymbol{U}}$ and $\hat{\boldsymbol{U}}_\ell$ estimated with $\eta = 2\sigma_n^2\kappa = 0$ (SVD). Bottom: distance between $\bar{\boldsymbol{U}}$ and $\hat{\boldsymbol{U}}_\ell$ estimated with $\eta = 0.5$ (left) and $\eta = 50$ (right).

This property is now exploited on a real image acquired in 1997 by the NASA spectro-imager AVIRIS over Moffett Field, CA. The scene consists of a large part of a lake (black pixels, top) and a coastal area (bottom) composed of soil (brown pixels) and vegetation (green pixels), leading to $R = 3$ endmembers whose spectra and abundance maps can be found in [14]. The same procedure was applied to this image and the result is shown in Figure 7. As can be observed, a simple local SVD is unable to locate possible non-linearities in the scene. However, for two non-zero values $\eta = 0.5$ and $\eta = 50$ (bottom left and right panels, respectively), the distances between the $\bar{\boldsymbol{U}}$ and the MMSD-based subspace $\hat{\boldsymbol{U}}_\ell$ clearly indicate that some non-linear effects occur in specific parts of the image, especially in the lake shore. This shows the accuracy of the proposed MMSD estimator to localize the non-linearities occurring in the scene, which is interesting for the analysis of hyperspectral images.

ACKNOWLEDGEMENT

The work of O. Besson was partly supported by DGA-MRIS under grant no. 2009.60.033.00.470.75.01.

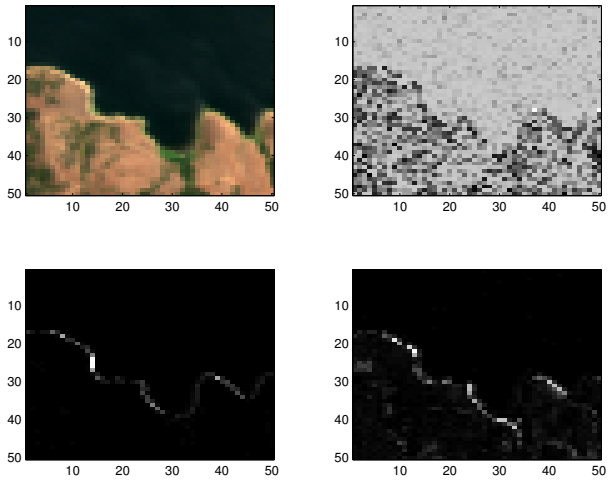

Fig. 7. Top, left: The Moffett Field scene as composite true colors. Top, right: distance between $\bar{\boldsymbol{U}}$ and $\hat{\boldsymbol{U}}_\ell$ estimated with $\eta = 0$ (SVD). Bottom: distance between $\bar{\boldsymbol{U}}$ and $\hat{\boldsymbol{U}}_\ell$ estimated with $\eta = 0.5$ (left) and $\eta = 50$ (right).

REFERENCES

[1] R. Kumaresan and D. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," *IEEE Transactions Acoustics Speech Signal Processing*, vol. 30, no. 6, pp. 833–840, December 1982.

[2] ——, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions Aerospace Electronic Systems*, vol. 19, no. 1, pp. 134–139, January 1983.

[3] J. Thomas, L. Scharf, and D. Tufts, "The probability of a subspace swap in the SVD," *IEEE Transactions Signal Processing*, vol. 43, no. 3, pp. 730–736, March 1995.

[4] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: prediction and cure," in *Proceedings ICASSP*, May 2001, pp. 4005–4008.

[5] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal Matrix Analysis Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[6] G. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Baltimore: John Hopkins University Press, 1996.

[7] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 1999.

[8] Y. Chikuse, *Statistics on special manifolds*. New York: Springer Verlag, 2003.

[9] K. Mammasis and R. W. Stewart, "The Fisher-Bingham spatial correlation model for multielement antenna systems," *IEEE Transactions Vehicular Technology*, vol. 58, no. 5, pp. 2130–2136, June 2009.

[10] O. Hamsici and A. Martinez, "Rotation invariant kernels and their application to shape analysis," *IEEE Transactions Pattern Analysis Machine Intelligence*, vol. 31, no. 11, pp. 1985–1999, November 2009.

[11] O. Besson, N. Dobigeon, and J.-Y. Tourneret, "Minimum mean square distance estimation of a subspace," IRIT/ENSEEIHT, Tech. Rep., 2011. [Online]. Available: http://dobigeon.perso.enseeiht.fr/app_MMSD.html

[12] P. D. Hoff, "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data," *Journal of Computational and Graphical Statistics*, vol. 18, no. 2, pp. 438–456, June 2009.

[13] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, January 2002.

[14] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions Signal Processing*, vol. 57, no. 11, pp. 4355–4368, November 2009.