BAYESIAN SUBSPACE ESTIMATION USING CS DECOMPOSITION

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.
olivier.besson@isae.fr, nicolas.dobigeon@enseeiht.fr, jean-yves.tourneret@enseeiht.fr

ABSTRACT

Subspace estimation using relatively few samples is a frequently encountered problem in numerous applications, including hyperspectral imagery the target application of this paper. We address this problem in a Bayesian framework assuming that some rough prior knowledge about the subspace is available. Our approach is based on the CS decomposition of an orthogonal matrix whose columns span the subspace of interest. This parametrization only involves mild assumptions about the distribution of the angles between the actual subspace and the prior subspace, and is intuitively appealing. We derive the posterior distribution for the matrices involved in the CS decomposition and the angles between subspaces, and we propose a Gibbs sampling scheme to compute the minimum mean-square distance estimator of the subspace of interest. The estimator accuracy is evaluated through numerical simulations and tested against real hyperspectral data.

1. MOTIVATION AND PROBLEM STATEMENT

Hyperspectral imagery is becoming increasingly used in earth observation systems due to its potentially more accurate analysis of earth surfaces [1]. It consists of collecting, in a large number of spectral bands, the electromagnetic signal reflected by the earth surface when illuminated by solar radiation. The spectral diversity of the various soil components can then be used for many hyperspectral imagery applications including target detection or classification. It is widely admitted that the response of a particular type of soil can be well approximated by a linear combination of a very few spectral components, referred to as endmembers. Therefore, the data mostly lies in a subspace, and the coordinates within this subspace provide information about the so-called abundances. A principal component analysis (PCA) is usually carried out on the whole image in order to estimate the endmembers subspace. This approach has the main merit of being relatively simple and computationally efficient. However, some recent studies have questioned the validity of the linear mixing model (LMM), and have instead investigated non-linear mixing models [2]. Although potentially more accurate, this approach suffers from a main drawback, namely its high computational cost. Moreover, if the linear model is not fully valid for the whole image, it can be appropriate for many parts of the image to be analyzed. Thus it can be interesting to detect the image areas where the LMM is questionable. To address this issue, we investigate a mixed approach where the subspace is estimated locally for each pixel of the image. The evolution of the estimated local subspaces is then used to characterize the existence and importance of non-linearities for each pixel of the image. Towards this end, we need to perform subspace estimation with a possibly very limited number of samples. Since the PCA of the whole image often provides quite an accurate approximation of the local subspaces, we intend to use this information in order to improve estimation in the low sample support regime. This paper studies a new Bayesian knowledge-aided subspace estimation method appropriate for datasets composed of few samples.

Since this problem is relevant in numerous applications, we recast it in a general framework. More precisely, let us consider an N-dimensional observation space with K available snapshots gathered column-wise in an $N \times K$ data matrix X, and let us assume that X can be written as

$$\mathsf{X} = \mathsf{H}\Psi + \mathsf{N}$$

where H is an $N \times p$ semi-orthogonal matrix ($\mathsf{H}^T\mathsf{H} = \mathsf{I}$) whose columns span the p-dimensional subspace of interest, Ψ is a $p \times K$ matrix whose columns contain the coordinates of the signal in the range space $\mathcal{R}(\mathsf{H})$ of H, and N denotes the additive noise. Assuming that the columns of N are independent and Gaussian distributed with zero mean and covariance matrix $\sigma^2\mathsf{I}_N$, the maximum likelihood (ML) estimator of $\mathcal{R}(\mathsf{H})$ is obtained from the p most significant left singular vectors of X. Therefore, the singular value decomposition (SVD) plays a central role in subspace estimation and is known to provide very accurate estimates of $\mathcal{R}(\mathsf{H})$ in most cases. However, it may yield inaccurate estimates either when the signal to noise ratio is low or when the sample support is small. Moreover, when the number of snapshots K is less than the subspace dimension p, X is at most of rank $K < p$. Therefore it becomes impossible to recover $\mathcal{R}(\mathsf{H})$ without any further information. For both cases, additional prior information about H may prove to be helpful. A natural way to introduce such knowledge is provided by the Bayesian framework which is advocated in the present paper.

More precisely, we assume that H is a random matrix, with some prior distribution $\pi(\mathsf{H})$, and our goal is to estimate H from the posterior distribution $p(\mathsf{H}|\mathsf{X})$. Similarly to [3,4] we consider minimum mean square distance (MMSD) estimators of H, i.e., estimators $\hat{\mathsf{H}}$ of H that minimize the average squared Frobenius norm of the difference between the projection matrices, viz $\mathsf{E}\left\{\left\|\hat{\mathsf{H}}\hat{\mathsf{H}}^T - \mathsf{H}\mathsf{H}^T\right\|_F^2\right\}$. The rationale behind this approach is that this is a natural metric on the Stiefel manifold [5,6]. Using this distance, the MMSD estimator was shown to be given by [3,4]

$$\begin{aligned}\hat{\mathsf{H}}_{\text{mmsd}} &= \arg\max_{\hat{\mathsf{H}}} \mathsf{E}\left\{\text{Tr}\left\{\hat{\mathsf{H}}^T\mathsf{H}\mathsf{H}^T\hat{\mathsf{H}}\right\}\right\} \\ &= \mathcal{P}_p\left\{\int \mathsf{H}\mathsf{H}^T p(\mathsf{H}|\mathsf{X})\, d\mathsf{H}\right\} \qquad (1)\end{aligned}$$

where $\text{Tr}\{.\}$ denotes the trace of a matrix and $\mathcal{P}_p\{.\}$ stands for the p principal eigenvectors of the matrix between braces. The MMSD

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

estimator thus amounts to finding the principal subspace of the posterior mean of the projection matrix $\mathsf{P} = \mathsf{H}\mathsf{H}^T$. Depending on the prior distribution $\pi(\mathsf{H})$, it may or may not be an easy task to compute the MMSD estimator. In the sequel, we state our assumptions regarding H and derive its corresponding MMSD estimator.

2. DATA MODEL AND SUBSPACE ESTIMATION

Consider the linear model (1) and assume that N is Gaussian distributed with independent columns so that the probability density function of X, conditioned on H and $\mathsf{\Psi}$, is given by

$$p\left(\mathsf{X}|\mathsf{H},\mathsf{\Psi}\right) \propto \operatorname{etr}\left\{-\frac{1}{2\sigma^2}\left(\mathsf{X}-\mathsf{H}\mathsf{\Psi}\right)^T\left(\mathsf{X}-\mathsf{H}\mathsf{\Psi}\right)\right\} \qquad (2)$$

where $\operatorname{etr}\{.\}$ stands for the exponential of the trace of the matrix between braces and $\propto$ means proportional to. We assume here that the white noise level σ^2 is known. Since no knowledge about $\mathsf{\Psi}$ is generally available, we treat it as a random matrix with uniform prior distribution, i.e., $\pi\left(\mathsf{\Psi}\right) \propto 1$. Thus the distribution of X, conditioned on H, is obtained as

$$\begin{aligned} p\left(\mathsf{X}|\mathsf{H}\right) &= \int p\left(\mathsf{X}|\mathsf{H},\mathsf{\Psi}\right)\pi\left(\mathsf{\Psi}\right)d\mathsf{\Psi} \\ &\propto \operatorname{etr}\left\{-\frac{1}{2\sigma^2}\mathsf{X}^T\mathsf{X}+\frac{1}{2\sigma^2}\mathsf{X}^T\mathsf{H}\mathsf{H}^T\mathsf{X}\right\}. \end{aligned} \qquad (3)$$

Note that $p\left(\mathsf{X}|\mathsf{H}\right)$ depends on H only through $\mathsf{P} = \mathsf{H}\mathsf{H}^T$. Let us turn now to the hypotheses regarding H. We assume that we have some a priori knowledge about the subspace spanned by the columns of H: this knowledge can come from some available models or expertise. More precisely, we assume that the range space $\mathcal{R}\left(\mathsf{H}\right)$ of H is close to the range space of some semi-orthogonal matrix $\bar{\mathsf{H}}$. Without loss of generality, this paper assumes that $\bar{\mathsf{H}} = \begin{bmatrix}\mathsf{I}_p & \mathsf{0}\end{bmatrix}^T$. In [4], we assumed that H was either drawn from a Bingham – $\pi_{\mathrm{B}}\left(\mathsf{H}\right) \propto \operatorname{etr}\left\{\kappa\mathsf{H}^T\bar{\mathsf{H}}\bar{\mathsf{H}}^T\mathsf{H}\right\}$– or a von Mises Fisher (vMF) distribution –$\pi_{\mathrm{vMF}}\left(\mathsf{H}\right) \propto \operatorname{etr}\left\{\kappa\mathsf{H}^T\bar{\mathsf{H}}\right\}$– and we derived MMSD estimators of H. When H follows a Bingham prior distribution, the MMSD estimator can be obtained in closed-form while, for a vMF prior, a Gibbs sampling estimator has to be implemented to compute the MMSD estimator (see [4] for details).

In this paper, we also consider MMSD estimators, but we take a different route regarding the statistical modeling of H. More precisely, we look for a less constrained model for H and, moreover, a model that *only involves mild assumptions about the distribution of the angles between* H *and* $\bar{\mathsf{H}}$ (note that the distance between H and $\bar{\mathsf{H}}$ only depends on these angles [4]). Doing so, we purposely reduce the amount of prior information brought by the model and hence hopefully get an improved robustness. The model proposed herein is based on the CS decomposition of H, which writes [6]

$$\mathsf{H} = \begin{bmatrix}\mathsf{U}_1\mathsf{C}\\ \mathsf{U}_2\mathsf{S}\end{bmatrix}\mathsf{V}^T \qquad (4)$$

where U_1 and V are $p\times p$ orthogonal matrices, U_2 is an $(N-p)\times p$ semi-orthogonal matrix ($\mathsf{U}_2^T\mathsf{U}_2 = \mathsf{I}_p$), $\mathsf{C} = \operatorname{diag}\left(\cos\theta_1,\cdots,\cos\theta_p\right)$ and $\mathsf{S} = \operatorname{diag}\left(\sin\theta_1,\cdots,\sin\theta_p\right)$. The angles θ_k correspond to the principal angles between $\mathcal{R}\left(\mathsf{H}\right)$ and $\mathcal{R}\left(\bar{\mathsf{H}}\right)$ while the columns of $\begin{bmatrix}\mathsf{U}_1\\ \mathsf{0}\end{bmatrix}$ and $\mathsf{H}\mathsf{V}$ are the associated principal vectors. The advantage of such a representation is that the angles between $\mathcal{R}\left(\mathsf{H}\right)$ and $\mathcal{R}\left(\bar{\mathsf{H}}\right)$ are directly revealed, and do not depend on the matrices U_1, U_2 and V, which can be arbitrary.

Let us turn now to the prior distributions for the matrices U_1 and U_2 as well as the vector $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 & \cdots & \theta_p\end{bmatrix}^T$. We assume that U_1 and U_2 have uniform distributions on the orthogonal group $O(p)$ and the Stiefel manifold $\mathcal{S}_{p,N-p}$, i.e., the set of $(N-p)\times p$ matrices U_2 such that $\mathsf{U}_2^T\mathsf{U}_2 = \mathsf{I}_p$. Regarding $\boldsymbol{\theta}$, we assume that θ_k are independent and identically distributed (i.i.d.) random variables, with uniform distribution on $[0,\theta_{\max}]$, i.e., $\theta_k \sim \mathrm{U}\left([0,\theta_{\max}]\right)$.

It should be pointed out that the proposed model is rather loose: U_1 and U_2 are arbitrary matrices, while the angles θ_k are a priori uniformly distributed on $[0,\theta_{\max}]$ where $\theta_{\max}$ sets the maximum angle between $\mathcal{R}\left(\mathsf{H}\right)$ and $\mathcal{R}\left(\bar{\mathsf{H}}\right)$: therefore, the smaller $\theta_{\max}$, the closer these two subspaces. In contrast, when $\theta_{\max}$ increases, the two subspaces can be quite far apart. Consequently, for small $\theta_{\max}$ we can expect the MMSD estimator to strongly rely on $\bar{\mathsf{H}}$, while for large $\theta_{\max}$ the data X is likely to prevail. Moreover, $\theta_{\max}$, which is the only parameter of the proposed statistical model a user has to set, is a more intuitively meaningful quantity than the concentration parameter κ that ruled the Bingham or vMF distributions in [4]. Indeed, it is not always easy to select an appropriate value of κ and this value can have a strong influence on the shape of the distribution. Moreover, a choice for κ induces a given distribution for the angles θ_k. Therefore, compared to [4], the model presented here is less constrained and defined by a single parameter (viz $\theta_{\max}$) which is more intuitively appealing and easier to set.

Once the likelihood and the prior distributions have been set, we are in a position to derive the posterior distributions. Using (3)-(4) along with the partitioning $\mathsf{X} = \begin{bmatrix}\mathsf{X}_1^T & \mathsf{X}_2^T\end{bmatrix}^T$ of X, it is straightforward to show that the the joint posterior distribution of U_1, U_2 and $\boldsymbol{\theta}$ is given by

$$\begin{aligned} p\left(\mathsf{U}_1,\mathsf{U}_2,\boldsymbol{\theta}|\mathsf{X}\right) &\propto p\left(\mathsf{X}|\mathsf{H}\right)\pi\left(\mathsf{U}_1\right)\pi\left(\mathsf{U}_2\right)\pi\left(\boldsymbol{\theta}\right) \\ &\propto \operatorname{etr}\left\{\frac{1}{2\sigma^2}\left[\mathsf{C}^2\mathsf{U}_1^T\mathsf{X}_1\mathsf{X}_1^T\mathsf{U}_1+\mathsf{S}^2\mathsf{U}_2^T\mathsf{X}_2\mathsf{X}_2^T\mathsf{U}_2\right]\right\} \\ &\times \operatorname{etr}\left\{\frac{1}{\sigma^2}\mathsf{X}_2^T\mathsf{U}_2\mathsf{S}\mathsf{C}\mathsf{U}_1^T\mathsf{X}_1\right\}\pi(\mathsf{U}_1)\pi(\mathsf{U}_2)\pi(\boldsymbol{\theta}). \end{aligned} \qquad (5)$$

In order to obtain the MMSD estimator, we suggest, as in [4], to use a Gibbs sampler which enables one to iteratively draw samples from the full conditional distributions of (5). First, (5) implies that

$$p\left(\mathsf{U}_1|\mathsf{U}_2,\boldsymbol{\theta},\mathsf{X}\right) \propto \operatorname{etr}\left\{\frac{1}{2\sigma^2}\left[2\mathsf{U}_1^T\mathsf{X}_1\mathsf{X}_2^T\mathsf{U}_2\mathsf{S}\mathsf{C}+\mathsf{C}^2\mathsf{U}_1^T\mathsf{X}_1\mathsf{X}_1^T\mathsf{U}_1\right]\right\} \qquad (6)$$

which is recognized as a Bingham-von-Mises-Fisher (BMF) distribution with parameter matrices $\mathsf{X}_1\mathsf{X}_1^T$, $\frac{1}{2\sigma^2}\mathsf{C}^2$ and $\frac{1}{\sigma^2}\mathsf{X}_1\mathsf{X}_2^T\mathsf{U}_2\mathsf{S}\mathsf{C}$ respectively [7]. An efficient sampling scheme to generate random matrices drawn from a BMF distribution on the Stiefel manifold was proposed in [7] and can thus be used to draw matrices from (6). Similarly, looking carefully at (5) leads to

$$p\left(\mathsf{U}_2|\mathsf{U}_1,\boldsymbol{\theta},\mathsf{X}\right) \propto \operatorname{etr}\left\{\frac{1}{2\sigma^2}\left[2\mathsf{U}_2^T\mathsf{X}_2\mathsf{X}_1^T\mathsf{U}_1\mathsf{C}\mathsf{S}+\mathsf{S}^2\mathsf{U}_2^T\mathsf{X}_2\mathsf{X}_2^T\mathsf{U}_2\right]\right\} \qquad (7)$$

which is a BMF distribution with parameter matrices $\mathsf{X}_2\mathsf{X}_2^T$, $\frac{1}{2\sigma^2}\mathsf{S}^2$ and $\frac{1}{\sigma^2}\mathsf{X}_2\mathsf{X}_1^T\mathsf{U}_1\mathsf{C}\mathsf{S}$. Therefore, the sampling scheme of Hoff can also be used to draw matrices from the distribution in (7).

Let us finally investigate the posterior distribution of $\boldsymbol{\theta}$

$$p\left(\boldsymbol{\theta}|\mathsf{U}_1,\mathsf{U}_2,\mathsf{X}\right) \propto \operatorname{etr}\left\{\frac{1}{2\sigma^2}\left[\mathsf{C}^2\mathsf{U}_1^T\mathsf{X}_1\mathsf{X}_1^T\mathsf{U}_1+\mathsf{S}^2\mathsf{U}_2^T\mathsf{X}_2\mathsf{X}_2^T\mathsf{U}_2\right]\right\} \times \operatorname{etr}\left\{\frac{1}{\sigma^2}\mathsf{X}_2^T\mathsf{U}_2\mathsf{S}\mathsf{C}\mathsf{U}_1^T\mathsf{X}_1\right\}\pi(\boldsymbol{\theta}) \propto \prod_{k=1}^{p}\exp\left\{\alpha_k\cos^2\theta_k+\gamma_k\sin^2\theta_k +2\beta_k\cos\theta_k\sin\theta_k\right\}\mathbb{I}_{[0,\theta_{\max}]}(\theta_k) \quad (8)$$

where $\mathbb{I}_{[0,\theta_{\max}]}(\theta_k)$ is the indicator function defined in the interval $[0,\theta_{\max}]$ and where α_k, β_k, γ_k are the k-th elements of the diagonal of $\frac{1}{2\sigma^2}\mathsf{U}_1^T\mathsf{X}_1\mathsf{X}_1^T\mathsf{U}_1$, $\frac{1}{2\sigma^2}\mathsf{U}_1^T\mathsf{X}_1\mathsf{X}_2^T\mathsf{U}_2$ and $\frac{1}{2\sigma^2}\mathsf{U}_2^T\mathsf{X}_2\mathsf{X}_2^T\mathsf{U}_2$, respectively. The first thing to be noted is that the variables θ_k, conditioned on U_1, U_2 and X, are independent and hence one needs to generate p independent random variables. Unfortunately, the distribution in (8) does not belong to any known class of distributions and, therefore, generating random variables drawn from $p\left(\boldsymbol{\theta}|\mathsf{U}_1,\mathsf{U}_2,\mathsf{X}\right)$ appears problematic. In order to overcome this problem, we propose to resort to a Metropolis-Hastings (MH) move [8] which consists of generating samples according to a proposal distribution and accepting them according to a certain probability. In our case, we make the change of variable $x_k=\sin^2\theta_k$ in (8), and come up with the equivalent problem of finding a proposal distribution for the conditional distribution of $x_k\in\left[0,x_{\max}=\sin^2\theta_{\max}\right]$, which is is given by

$$p\left(x_k|\mathsf{U}_1,\mathsf{U}_2,\mathsf{X}\right)\propto\exp\left\{-(\alpha_k-\gamma_k)x_k+2\beta_k x_k^{1/2}(1-x_k)^{1/2}\right\} \times x_k^{-1/2}(1-x_k)^{-1/2}\mathbb{I}_{[0,x_{\max}]}(x_k). \quad (9)$$

Through preliminary experiments, we found out that a scaled beta distribution $q(x_k)\propto\left(\frac{x_k}{x_{\max}}\right)^{a_k-1}(1-\frac{x_k}{x_{\max}})^{b_k-1}$, with a suitable choice of a_k and b_k, provides a good approximation to (9) with a high acceptance rate for the candidates.

The Gibbs sampling scheme allows one to generate a sequence of random matrices $\mathsf{H}^{(n)}=\begin{bmatrix}\mathsf{U}_1^{(n)}\mathsf{C}^{(n)}\\ \mathsf{U}_2^{(n)}\mathsf{S}^{(n)}\end{bmatrix}$ where $\mathsf{C}^{(n)}=\operatorname{diag}\left(\cos\boldsymbol{\theta}^{(n)}\right)$ and $\mathsf{S}^{(n)}=\operatorname{diag}\left(\sin\boldsymbol{\theta}^{(n)}\right)$, with $\mathsf{U}_1^{(n)}$, $\mathsf{U}_2^{(n)}$ and $\boldsymbol{\theta}^{(n)}$ drawn from their respective posterior distributions in (6), (7) and (8). These matrices can in turn be used to approximate the MMSD estimator as

$$\hat{\mathsf{H}}_{\text{mmsd}}=\mathcal{P}_p\left\{\frac{1}{N_r}\sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r}\mathsf{H}^{(n)}\left(\mathsf{H}^{(n)}\right)^T\right\} \quad (10)$$

where N_{bi} is the number of burn-in iterations and N_r is the number of samples used for estimating H.

Remark 1. We would like to point out that a minimum mean-square error (MMSE) estimator of H, which would entail approximating the integral $\int\mathsf{H}\,p\left(\mathsf{H}|\mathsf{X}\right)d\mathsf{H}$ by the arithmetic mean of the set of matrices $\mathsf{H}^{(n)}$, may not be meaningful here. Indeed, the range space of H is given up to right multiplication by an orthogonal matrix and therefore, $\mathcal{R}\left(\mathsf{H}^{(n)}\right)$ could be close to $\mathcal{R}\left(\mathsf{H}\right)$ even if the actual matrices $\mathsf{H}^{(n)}$ and H are not close. It ensues that the arithmetic mean of the matrices $\mathsf{H}^{(n)}$ could result in a poor subspace estimate despite the fact that, individually, the subspaces spanned by each matrix $\mathsf{H}^{(n)}$ might be accurate. A maximum a posteriori (MAP) approach could also be advocated where the MAP estimator would be obtained as the matrix $\mathsf{H}^{(n)}$ which maximizes $p\left(\mathsf{U}_1^{(n)},\mathsf{U}_2^{(n)},\boldsymbol{\theta}^{(n)}|\mathsf{X}\right)$.

3. SIMULATIONS

In this section we assess the performance of the above estimators through Monte-Carlo simulations. The performance measure will be the average distance between the projection matrices $\hat{\mathsf{H}}\hat{\mathsf{H}}^T$ and $\mathsf{H}\mathsf{H}^T$ where $\hat{\mathsf{H}}$ stands for one of the estimators. More precisely, we will display the average fraction of energy (AFE) of $\hat{\mathsf{H}}$ in $\mathcal{R}\left(\mathsf{H}\right)$ which is defined as

$$\text{AFE}\left(\hat{\mathsf{H}},\mathsf{H}\right)=\mathsf{E}\left\{\text{Tr}\left\{\hat{\mathsf{H}}^T\mathsf{H}\mathsf{H}^T\hat{\mathsf{H}}\right\}/p\right\}. \quad (11)$$

In all simulations $N=20$, $p=5$ and $\bar{\mathsf{H}}=\begin{bmatrix}\mathsf{I}_p & \mathsf{0}\end{bmatrix}^T$. The matrix $\boldsymbol{\Psi}$ is generated from a Gaussian distribution with zero-mean and covariance matrix I_p and the signal-to-noise ratio is defined as $\text{SNR}=10\log_{10}\left(\frac{p}{N\sigma^2}\right)$. The angles between $\mathcal{R}\left(\mathsf{H}\right)$ and $\mathcal{R}\left(\bar{\mathsf{H}}\right)$ are fixed over all simulations and set to $\boldsymbol{\theta}=\begin{bmatrix}15^\circ & 25^\circ & 35^\circ & 45^\circ & 55^\circ\end{bmatrix}^T$ which results in $\text{AFE}\left(\mathsf{H},\bar{\mathsf{H}}\right)=0.6509$. The matrices U_1 and U_2 are drawn randomly at each Monte-Carlo run and $\theta_{\max}$ is set to $\theta_{\max}=60^\circ$. The number of burn-in iterations in the Gibbs sampler is set to $N_{\text{bi}}=10$ and $N_r=1000$. The MMSD estimator (1) is compared with the usual SVD-based estimator and the estimator $\hat{\mathsf{H}}=\bar{\mathsf{H}}$ that only uses the a priori knowledge and does not make use of the data. We successively investigate the influence of K and SNR in figures 1 and 2.

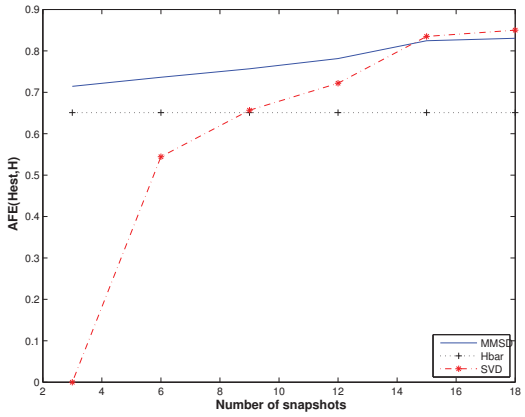

Fig. 1. Fraction of energy of $\hat{\mathsf{U}}$ in $\mathcal{R}\left(\mathsf{U}\right)$ versus K. $N=20$, $p=5$, SNR $=3$dB and $\theta_{\max}=60^\circ$.

It can be observed that the MMSD estimator outperforms the usual SVD-based estimator, for small K and small SNR: under these conditions, it makes a sound use of the prior information and provides more accurate estimates. Note also that it performs better than the estimate $\hat{\mathsf{H}}=\bar{\mathsf{H}}$, and hence the prior by itself is not sufficient.

4. APPLICATION TO HYPERSPECTRAL DATA

We now apply the subspace estimation scheme developed above to a real hyperspectral image, acquired by the NASA spectro-imager AVIRIS over Moffett Field, CA, in 1997. We consider a 50×50 sub-image which contains partly a lake (upper part of the sub-image) and partly a coastal area (lower part of the sub-image) composed of soil and vegetation, see [9] for a more detailed description. The data is collected in $N=183$ spectral bands. Under the linear

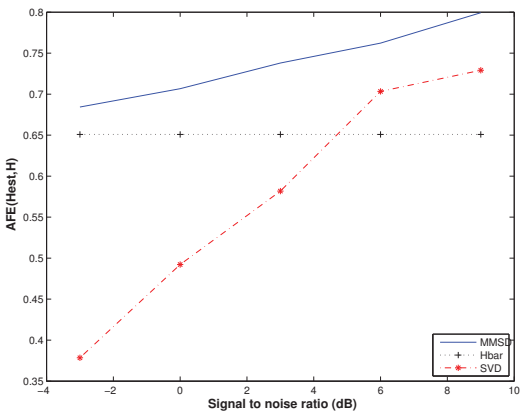

Fig. 2. Fraction of energy of $\hat{\mathsf{U}}$ in $\mathcal{R}(\mathsf{U})$ versus SNR. $N = 20$, $p = 5$, $K = 5$ and $\theta_{\max} = 60°$.

mixing model and in the absence of noise, the data matrix $\mathsf{Y} = \begin{bmatrix}\boldsymbol{y}_1 & \cdots & \boldsymbol{y}_L\end{bmatrix}$ where $\boldsymbol{y}_\ell \in \mathbb{R}^N$ stands for the ℓ-th pixel, can be decomposed as $\mathsf{Y} = \mathsf{MA}$ where $\mathsf{M} = \begin{bmatrix}\boldsymbol{m}_1 & \cdots \boldsymbol{m}_R\end{bmatrix}$. The vectors $\boldsymbol{m}_r$, $r = 1, \cdots, R$ stand for the endmembers, i.e., the spectral signatures which best describe the soil components. The columns $\boldsymbol{a}_\ell = \begin{bmatrix}a_{\ell,1} & \cdots a_{\ell,R}\end{bmatrix}^T$ of the matrix $\mathsf{A} = \begin{bmatrix}\boldsymbol{a}_1 & \cdots \boldsymbol{a}_L\end{bmatrix}$ are the so-called abundances: they satisfy the positivity constraint $a_{\ell,r} \geq 0$ and the sum-to-one property, i.e., $\boldsymbol{a}_\ell^T \mathbf{1}_R = 1$ where $\mathbf{1}_R$ is the $R \times 1$ vector whose elements are all equal to 1. The pixels $\boldsymbol{y}_\ell$ thus belong to a simplex whose vertices are the R endmembers $\boldsymbol{m}_r$ [9]. If $\boldsymbol{\mu} = L^{-1} \sum_{\ell=1}^{L} \boldsymbol{y}_\ell$ denotes the mean value of the pixels, then the centered data matrix $\mathsf{X} = \mathsf{Y} - \boldsymbol{\mu}\mathbf{1}_L^T$ belongs to a p-dimensional subspace (with $p = R - 1$) which can be estimated by a number of techniques, including PCA [9].

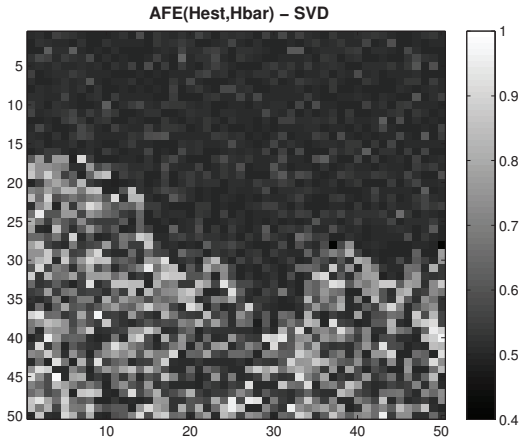

Fig. 3. Moffett image. $\mathrm{AFE}\left(\hat{\mathsf{H}}_\ell, \bar{\mathsf{H}}\right)$ obtained with the SVD. $N = 183$, $p = 2$ and $K = 4$.

Usually, PCA is performed on the whole set of L pixels with a view to obtain a subspace that fits the whole image. However, the LMM may not be valid on the whole image and we intend to use the MMSD estimator to detect the zones in the image where the LMM might be questioned. Towards this end, we conduct subspace

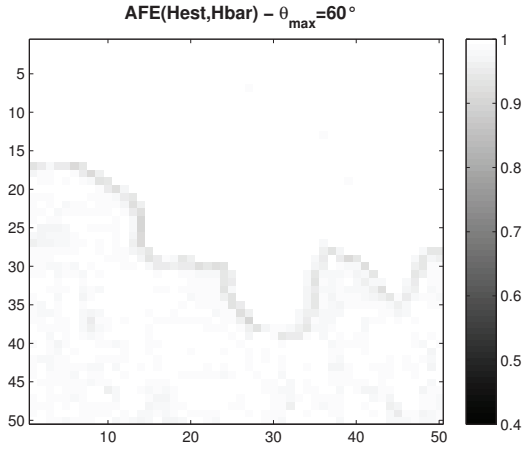

Fig. 4. Moffett image. $\mathrm{AFE}\left(\hat{\mathsf{H}}_\ell, \bar{\mathsf{H}}\right)$ obtained with the CS-based Bayesian MMSD estimator. $N = 183$, $p = 2$, $K = 4$ and $\theta_{\max} = 60°$.

estimation at the pixel level, i.e., the MMSD estimator is used to estimate the subspace spanned by $\boldsymbol{y}_\ell$ and its $K-1$ spectrally nearest pixels: this local estimate is then compared to the global estimate $\bar{\mathsf{H}}$ obtained from PCA over the whole image. A very small distance between these two subspaces means that the linear model described by $\bar{\mathsf{H}}$ is rather accurate. On the other hand, a large distance can result from a matrix $\bar{\mathsf{H}}$ that does not describe accurately the scene around pixel ℓ, or from some non-linear mixing effects affecting the pixel of interest. Figures 3 and 4 display $\mathrm{AFE}\left(\hat{\mathsf{H}}_\ell, \bar{\mathsf{H}}\right)$ for the usual SVD-based estimator and the proposed MMSD estimator. It can be observed that a local SVD would predict rather large differences between the local subspaces and $\bar{\mathsf{H}}$, especially for pixels in the lake area. In contrast, the Bayesian MMSD estimator shows that $\bar{\mathsf{H}}$ is a rather accurate subspace for the whole image, except for the pixels along the shore. This seems logical as non-linear mixing effects are more likely to occur in these zones, while the linear model is likely to apply well elsewhere. Therefore, the MMSD estimator can serve as a good indicator of the image regions where the LMM can be questioned.

5. REFERENCES

[1] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York: Kluwer Academic, 2003.

[2] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Transactions Geoscience Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, November 2011.

[3] A. Srivastava, "A Bayesian approach to geometric subspace estimation," *IEEE Transactions Signal Processing*, vol. 48, no. 5, pp. 1390–1400, May 2000.

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

[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] 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.

[8] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer Verlag, 2004.

[9] 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.