

# Hyperspectral Subspace Identification

José M. Bioucas-Dias, *Member, IEEE*, and José M. P. Nascimento, *Member, IEEE*

**Abstract**—Signal subspace identification is a crucial first step in many hyperspectral processing algorithms such as target detection, change detection, classification, and unmixing. The identification of this subspace enables a correct dimensionality reduction, yielding gains in algorithm performance and complexity and in data storage. This paper introduces a new minimum mean square error-based approach to infer the signal subspace in hyperspectral imagery. The method, which is termed hyperspectral signal identification by minimum error, is eigen decomposition based, unsupervised, and fully automatic (i.e., it does not depend on any tuning parameters). It first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense. State-of-the-art performance of the proposed method is illustrated by using simulated and real hyperspectral images.

**Index Terms**—Dimensionality reduction, hyperspectral imagery, hyperspectral signal subspace identification by minimum error (HySime), hyperspectral unmixing, linear mixture, minimum mean square error (mse), subspace identification.

## I. INTRODUCTION

**H**YPERSPECTRAL sensors sample the reflected solar radiation from the Earth's surface in the portion of the spectrum extending from the visible region through the near-infrared and midinfrared (wavelengths between 0.3 and 2.5  $\mu\text{m}$ ) in hundreds of narrow (on the order of 10 nm) contiguous bands [3]. This high spectral resolution yields large amounts of data. For example, the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) collects a 512 (along track)  $\times$  614 (across track)  $\times$  224 (bands)  $\times$  12 (bits) data cube in 43 s, corresponding to more than 700 Mb; Hyperion collects 4 Mb in 3 s, corresponding to 366 kB/km<sup>2</sup> [4]. Such huge data volumes put stringent requirements on communications, storage, and processing.

Lets us represent the spectral radiances read by a hyperspectral sensor from a given pixel as an  $L$ -D vector, where  $L$  is the number of bands and each channel is assigned to one axis of  $\mathbb{R}^L$ . Under the linear mixing scenario (see, e.g., [5]–[13]), the spectral vectors are a linear combination of the so-called endmember signatures. The number of endmembers

present in a given scene is, very often, much smaller than the number of bands  $L$ . Therefore, hyperspectral vectors lie in a low-dimensional linear subspace. The identification of this subspace enables the representation of spectral vectors in a low-dimensional subspace, thus yielding gains in computational time and complexity and in data storage. The computation of statistical estimates is a relevant example of the advantages of dimensionality reduction, considering that the number of samples required to obtain accurate estimates increases drastically with the dimensionality of the data [14].

## A. Unsupervised Hyperspectral Dimensionality Reduction Methods

Unsupervised dimensionality reduction has been approached in many ways. Band selection or band extraction, as the name suggests, exploits the high correlation existing between adjacent bands to select a few spectral components among those with higher signal-to-noise ratio (SNR) [15]–[20]. Projection techniques seek for the best subspace to project data by minimizing an objective function. For example, principal component analysis [21] computes the Karhunen–Loève transform, which is the best data representation in the least squares sense; singular value decomposition (SVD) [22] provides the projection that best represents data in the maximum power sense; maximum noise fraction (MNF) [23] and noise-adjusted principal components (NAPC) [24] seek the projection that optimizes the ratio of noise to signal powers. The NAPC are mathematically equivalent to the MNF [24] and can be interpreted as a sequence of two principal component transforms, where the first applies to the noise and the second applies to the transformed data set.

Topological methods are local approaches that infer the manifold, which is usually of low dimension, where data set lives [25]. For example, curvilinear component analysis [26], curvilinear distance analysis [27], and manifold learning [28]–[31] are nonlinear projections based on the preservation of the local topology. Independent component analysis [32], [33], projection pursuit [34], [35], and wavelet decomposition [36], [37] have also been considered.

The optical real-time adaptive spectral identification system (ORASIS) [38] is a hyperspectral framework developed by U.S. Naval Research Laboratory aiming at real-time data processing. This framework consists of several modules, where the dimension reduction is achieved by identifying a subset of exemplar pixels that convey the variability in a scene. Each new pixel collected from the scene is compared with each exemplar pixel by using an angle metric. The new pixel is added to the exemplar set if it is sufficiently different from each of the existing exemplars. An orthogonal basis is periodically

Manuscript received December 13, 2006; revised December 10, 2007. This work was supported by the Fundação para a Ciência e Tecnologia, under Project PDCTE/CPS/49967/2003, Project POSC/EEA-CPS/61271/2004, and Grant SFRH/BPD/39475/2007.

J. M. Bioucas-Dias is with the Instituto de Telecomunicações and the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

J. M. P. Nascimento is with the Instituto de Telecomunicações, Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisbon, Portugal, and also with the Department of Electronics, Telecommunications, and Computer Engineering, Instituto Superior de Engenharia de Lisboa, Polytechnical Institute of Lisbon, 1959-007 Lisbon, Portugal (e-mail: zen@isel.pt).

Digital Object Identifier 10.1109/TGRS.2008.918089

created from the current set of exemplars by using a modified Gram–Schmidt procedure. The algorithm performance depends on the value of the error-angle threshold (chosen by the user). The dimension reduction flowchart of ORASIS can be found in [39].

The identification of the signal subspace is a model order inference problem to which information theoretic criteria, such as the minimum description length (MDL) [40], [41] or the Akaike information criterion (AIC) [42], comes to mind. These criteria have, in fact, been used in hyperspectral applications [43], adopting the approach introduced by Wax and Kailath [44].

Harsanyi *et al.* [45] developed a Neyman–Pearson detection theory-based thresholding method (HFC) to determine the number of spectral endmembers in hyperspectral data, which is referred to in [43] as virtual dimensionality (VD). The HFC method is based on a detector built on the eigenvalues of the sample correlation and covariance matrices. A modified version, which is termed noise-whitened HFC (NWHFC), includes a noise-whitening step [43].

### B. Proposed Approach

Hyperspectral signal identification by minimum error (HySime) starts by estimating the signal and the noise correlation matrices, using multiple regression. A subset of eigenvectors of the signal correlation matrix is then used to represent the signal subspace. This subspace is inferred by minimizing the sum of the projection error power with the noise power, which are, respectively, decreasing and increasing functions of the subspace dimension. Therefore, if the subspace dimension is overestimated, the noise power term is dominant, whereas if the subspace dimension is underestimated, the projection error power term is the dominant. The overall scheme is computationally efficient, unsupervised, and fully automatic in the sense that it does not depend on any tuning parameters.

This paper is organized as follows. Section II formulates the signal subspace identification problem and reviews the SVD and MNF methods. Section III describes the fundamentals of the proposed method, including the noise estimation approach. Sections IV and V evaluate the proposed algorithm using simulated and real data, respectively. Section VI summarizes our contributions and presents some concluding remarks.

## II. PROBLEM FORMULATION AND CLASSICAL DIMENSIONALITY REDUCTION METHODS

Assume that the observed spectral vectors are given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathbf{x}$  and  $\mathbf{n}$  are  $L$ -D vectors standing for signal and additive noise, respectively. Furthermore, assume that signal vectors are in an unknown  $p$ -D subspace, i.e.,

$$\mathbf{x} = \mathbf{M}\mathbf{s}$$

with  $p < L$  and  $\mathbf{M}$  being a full-rank  $L \times p$  matrix. Under the linear mixing scenario, the columns of  $\mathbf{M}$  are the endmember signatures, and  $\mathbf{s}$  is the abundance fraction vector. To be physically meaningful [46], abundance fractions are subject

to nonnegativity and full additivity constraints, i.e.,  $\mathbf{s} \in \Delta_p$ , where

$$\Delta_p \equiv \left\{ \mathbf{s} \in \mathbb{R}^p : s_j \geq 0, \sum_{j=1}^p s_j = 1 \right\}. \quad (2)$$

Herein, we do not assume any special structure for the scattering mechanism, i.e., our approach works both under the linear and nonlinear scenarios. Even in the nonlinear mixing scenario, it often happens that the signal subspace dimension, although larger than the number of endmembers, is much smaller than the number of bands  $L$ . In these cases, it is still worthwhile, then, to estimate the signal subspace and represent the data on it. Note that this procedure does not preclude the application of future nonlinear projection techniques; on the contrary, it is an advantage, considering that the data is represented by vectors of a smaller dimension, thus lightening the computational complexity of any posterior processing scheme.

Let us assume, for a while, that the noise  $\mathbf{n}$  is zero-mean Gaussian independent and identically distributed (i.i.d.), with variance  $\sigma_n^2$  per band. Under these circumstances, the maximum-likelihood (ML) estimate of the signal subspace is spanned by the  $p$ -dominant eigenvectors of the observed data correlation matrix of  $\mathbf{y}$  [22, ch. 6], i.e.,  $\langle \mathbf{M} \rangle = \langle [\mathbf{e}_1, \dots, \mathbf{e}_p] \rangle$  (the notation  $\langle \mathbf{M} \rangle$  represents the subspace spanned by the columns of  $\mathbf{M}$ ), where  $\mathbf{e}_i$ , which is for  $i = 1, \dots, p$ , represents the  $p$ -dominant eigenvectors of the observed data correlation matrix  $\mathbf{R}_y$ .

### A. Eigen Analysis of $\mathbf{R}_y$

The ML estimator of the subspace signal that was just presented assumes that the dimension of the subspace is known beforehand. However, this dimension is often *a priori* unknown. Nevertheless, a similar approach has been extensively used as a dimensionality reduction tool in hyperspectral image processing [5], [47]. It consists of assuming that the noise is zero-mean i.i.d. Thus, the correlation matrix of the observed vectors may be written as  $\mathbf{R}_y = \mathbf{E}(\Sigma + \sigma_n^2 \mathbf{I}_L)\mathbf{E}^T$ , where  $\mathbf{I}_L$  is the  $L \times L$  identity matrix and  $\mathbf{E}$  and  $\Sigma$  are the eigenvector and eigenvalue matrices of the signal correlation matrix  $\mathbf{R}_x$ , respectively. Assuming that  $\mathbf{R}_x$  has just  $p$  positive eigenvalues and that they are ordered along the diagonal of  $\Sigma$  by decreasing magnitude, we then have  $\langle \mathbf{M} \rangle = \langle [\mathbf{e}_1, \dots, \mathbf{e}_p] \rangle$ , i.e., the estimate of the signal subspace is the span of the eigenvectors of  $\mathbf{R}_y$  whose respective eigenvalues are larger than  $\sigma_n^2$  [22].

This is basically the idea behind SVD-based eigen-decomposition dimensionality reduction. Two limitations of this approach are the following.

- 1) The noise present in most hyperspectral data sets is not i.i.d., and thus, the signal subspace is no longer given by the span of the  $p$  eigenvectors corresponding to the largest eigenvalues nor by any other set of eigenvectors.
- 2) Even if the noise were i.i.d., the aforementioned procedure to infer the subspace dimension would be prone to errors, owing to random perturbations that are always present in the estimates of  $\sigma_n^2$ ,  $\mathbf{E}$ , and  $\Sigma$ .

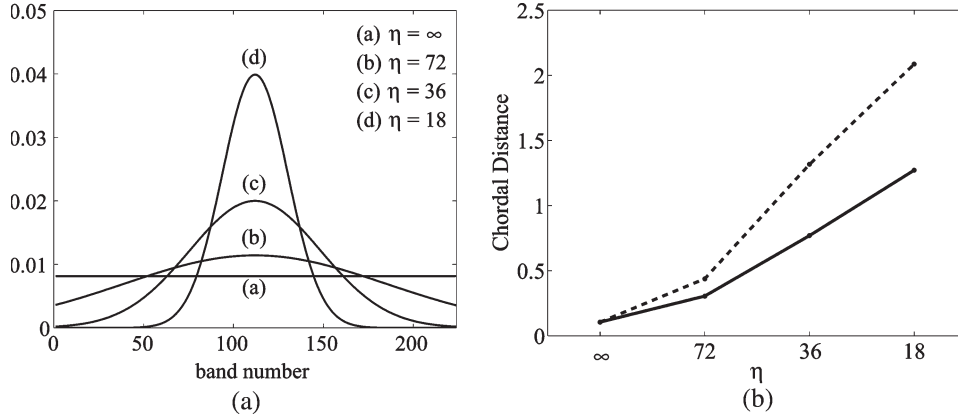


Fig. 1. (a) Noise variance in each band given by (4) for different values of  $\eta$ . (b) Chordal distance as a function of parameter  $\eta$  for SNR = 17 dB. (Solid Line) Subspace inferred by SVD. (Dashed line) Subspace inferred by MNF.

We illustrate these limitations with an experiment built on a simulated hyperspectral image composed of  $10^5$  pixels and generated according to the linear mixing scattering mechanism. Each pixel is a mixture of five endmembers' signatures ( $p = 5$ ) selected from the U.S. Geological Survey (USGS) digital spectral library [48]. Abundance fractions are generated according to a Dirichlet distribution given by

$$D(s_1, s_2, \dots, s_p | \theta_1, \theta_2, \dots, \theta_p) = \frac{\Gamma(\sum_{j=1}^p \theta_j)}{\prod_{j=1}^p \Gamma(\theta_j)} \prod_{j=1}^p s_j^{\theta_j-1} \quad (3)$$

where  $\Gamma(\cdot)$  denotes the Gamma function and  $\{s_1, \dots, s_p\} \in \Delta_p$ . The mean value of the  $j$ th endmember fraction is  $\mathbb{E}[s_j] = \theta_j / \sum_{l=1}^p \theta_l$  ( $\mathbb{E}[\cdot]$  denotes the expectation operator) [49]. The Dirichlet density, besides enforcing positivity and full additivity constraints, displays a wide range of shapes, depending on its parameters. On the other hand, as noted in [50], the Dirichlet density is suited to model fractions.

Consider that the noise correlation matrix is  $\mathbf{R}_n = \text{diag}(\sigma_1^2, \dots, \sigma_L^2)$  and that the diagonal elements follow a Gaussian shape centered at the band  $L/2$ , i.e.,

$$\sigma_i^2 = \sigma^2 \frac{e^{-\frac{(i-L/2)^2}{(2\eta^2)}}}{\sum_{j=1}^L e^{-\frac{(j-L/2)^2}{(2\eta^2)}}} \quad (4)$$

for  $i = 1, \dots, L$ . The parameter  $\eta$  plays the role of variance in the Gaussian shape ( $\eta \rightarrow \infty$  corresponds to white noise;  $\eta \rightarrow 0$  corresponds to one-band noise). The parameter  $\sigma^2$  controls the total noise power. We set  $\sigma^2 = 8.1 \times 10^{-3}$ , leading to SNR = 17 dB, where

$$\text{SNR} \equiv 10 \log_{10} \frac{\mathbb{E}[\mathbf{x}^T \mathbf{x}]}{\mathbb{E}[\mathbf{n}^T \mathbf{n}]} \quad (5)$$

To measure the dissimilarity between the signal subspace and the subspace inferred by SVD, we adopt the chordal distance [51], [52], which is defined as

$$d = \frac{1}{\sqrt{2}} \|\mathbf{U}_p - \mathbf{U}_M\|_F \quad (6)$$

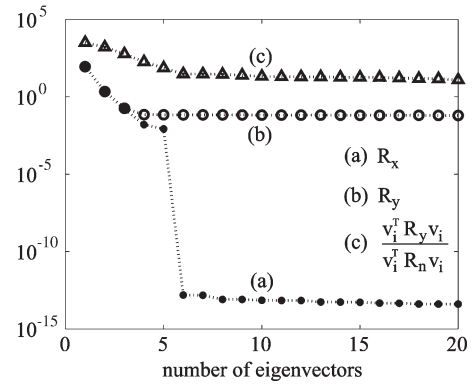


Fig. 2. (a) Eigenvalues of the signal correlation matrix  $\mathbf{R}_x$ . (b) Eigenvalues of observed data correlation matrix  $\mathbf{R}_y$ . (c) Decreasing ratio of  $(\mathbf{v}_i^T \mathbf{R}_y \mathbf{v}_i) / (\mathbf{v}_i^T \mathbf{R}_n \mathbf{v}_i)$  [ $\mathbf{v}_i$  represents the left-hand eigenvectors of  $\mathbf{R}_n \mathbf{R}_y^{-1}$ ]. Simulation parameters: SNR = 12 dB,  $\eta = 18$ , and  $p = 5$ .

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix,  $\mathbf{U}_p = \mathbf{E}_p \mathbf{E}_p^T$  and  $\mathbf{U}_M = \mathbf{E}_M \mathbf{E}_M^T$  are projection matrices onto the subspace of dimension  $p$  spanned, respectively, by the first  $p$  singular vectors of  $\mathbf{R}_y$  and by the columns of  $\mathbf{M}$ . We note that the chordal distance is a measure of the projection error norm, i.e., it is a measure of the errors  $(\mathbf{U}_p - \mathbf{U}_M)\mathbf{x}$  for  $\|\mathbf{x}\| = 1$  and  $\mathbf{x} \in \mathbb{R}^L$ . When this distance is zero, the two projections are equal.

Fig. 1(a) shows the shape of the noise variance  $\sigma_i^2$  for  $\eta \in \{18, 38, 74, \infty\}$ . Fig. 1(b) shows the chordal distance between the signal subspace and the subspace inferred by SVD as a function of  $\eta$ . Notice the clear increasing of the chordal distance with  $\eta$ , i.e., the chordal distance increases as the noise correlation shape becomes less flat. The degradation of the signal subspace estimate, owing to the violation of the white-noise assumption, is quite clear.

In the example just presented, the subspace dimension was assumed to be known. However, this dimension is unknown in most real applications and must be inferred from data, which is as already referred to. This is a model-order inference problem that, if based only on the eigenvalues of the observed data correlation matrix ( $\mathbf{R}_y$ ), may lead to poor results. This aspect is shown in Fig. 2, where we have plotted the eigenvalues of the signal correlation matrix  $\mathbf{R}_x$  [plot (a)] and the eigenvalues

of  $\mathbf{R}_y$  [plot (b)] computed in the aforementioned experiment with  $\text{SNR} = 12$  dB and  $\eta = 18$ . From plot (a), it is clear that the signal subspace dimension is  $p = 5$ . However, considering that  $\mathbf{R}_x$  is not known in real applications, we resort to  $\mathbf{R}_y$ . An estimate of the signal subspace dimension based on plot (b) would lead, most probably, to  $\hat{p} = 3$ . We will see that HySime is able to infer the correct subspace dimension in the similar scenarios.

### B. MNF

The MNF is a another popular subspace inference tool in remote sensing, which takes into account the noise statistics. Nonetheless, it has limitations that are similar to SVD-based approaches, as we illustrate in the following.

MNF finds non-orthogonal directions, minimizing the noise fraction (or, equivalently, maximizing the SNR). Assuming that the noise correlation matrix  $\mathbf{R}_n$  or an estimate is known, this procedure consists of finding non-orthogonal directions, minimizing the ratio

$$\frac{\mathbf{v}_i^T \mathbf{R}_n \mathbf{v}_i}{\mathbf{v}_i^T \mathbf{R}_y \mathbf{v}_i} \quad (7)$$

with respect to  $\mathbf{v}_i$ . This problem is known as the generalized Rayleigh quotient, and its solution is given by the left-hand eigenvectors  $\mathbf{v}_i$  of  $\mathbf{R}_n \mathbf{R}_y^{-1}$ , which are for  $i = 1, \dots, L$  [53].

For the i.i.d. noise, we have  $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_L$  and  $\mathbf{R}_y^{-1} = \mathbf{E}(\Sigma + \sigma_n^2 \mathbf{I}_L)^{-1} \mathbf{E}^T$ , and therefore, the MNF and SVD yield the same subspace estimate. However, if the noise is not i.i.d., the directions found by the MNF transform maximize the SNR but do not correctly identify the signal subspace. To illustrate this aspect, we apply the MNF transform to the data set generated in the previous section. The dashed line in Fig. 1(b) represents the chordal distance between the signal subspace and the subspace inferred by the MNF transform for different values of the parameter  $\eta$  and assuming that  $p = 5$ .<sup>1</sup> The chordal distance exhibits a pattern that is similar to that of the SVD-based approach, being, however, larger for  $\eta \neq \infty$  (white-noise case). If the space is unknown, it shall be inferred from the ratio  $(\mathbf{v}_i^T \mathbf{R}_y \mathbf{v}_i) / (\mathbf{v}_i^T \mathbf{R}_n \mathbf{v}_i)$ . Plot (c) in Fig. 2 shows a plot of this ratio, which is by decreasing order for  $p = 5$ ,  $\eta = 18$ , and  $\text{SNR} = 12$  dB. The inference of the signal subspace dimension based on this plot may be even worse than that obtained in the SVD case, considering that the curve does not display any clear drop nor any floor.

## III. SIGNAL SUBSPACE ESTIMATION

This section formally introduces the HySime algorithm. The method starts by estimating the signal and the noise correlation matrices, and then, it selects the subset of eigenvectors that best represents the signal subspace in the minimum mean square error (mse) sense. The application of this criterion leads to the minimization of a two-term objective function. One term corresponds to the power of the signal

projection error and is a decreasing function of the subspace dimension; the other term corresponds to the power of the noise projection and is an increasing function of subspace dimension.

### A. Noise Estimation

Noise estimation is a classical problem in data analysis and, particularly, in remote sensing. Arguably, in hyperspectral imagery, the simplest noise estimation procedure is the shift difference method, which is also denoted as the nearest neighbor difference [23]. This approach assumes that noise samples taken from adjacent pixels are independent and have the same statistics, but the signal component is practically equal. To obtain meaningful noise estimates, the shift difference method shall be applied in homogeneous areas rather than on the entire image. This method has the following two weaknesses. First, it assumes that adjacent pixels have the same signal information, which is not valid in most hyperspectral data sets; second, to improve the noise estimation, the selection of a homogeneous area must be carried out.

In this paper, we follow a multiple regression theory-based approach [43], [54], which outperforms the shift difference method, which is used, for example, in the NAPC [24] algorithm. The high correlation between neighboring spectral bands is the main reason underlying the good performance of the multiple regression theory in hyperspectral applications.

Let  $\mathbf{Y}$  denote an  $L \times N$  matrix holding the  $N$  spectral observed vectors of size  $L$ . Define the matrix  $\mathbf{Z} = \mathbf{Y}^T$  and the  $N \times 1$  vector  $\mathbf{z}_i = [\mathbf{Z}]_{:,i}$ , where  $[\mathbf{Z}]_{:,i}$  stands for the  $i$ th column of  $\mathbf{Z}$  (i.e.,  $\mathbf{z}_i$  contains the data read by the hyperspectral sensor at the  $i$ th band for all image pixels), and the  $N \times (L-1)$  matrix  $\mathbf{Z}_{\partial i} = [\mathbf{z}_1, \dots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \dots, \mathbf{z}_L]$ .

Assume that  $\mathbf{z}_i$  is explained by a linear combination of the remaining  $L-1$  bands. Formally, this consists in writing

$$\mathbf{z}_i = \mathbf{Z}_{\partial i} \beta_i + \xi_i \quad (8)$$

where  $\mathbf{Z}_{\partial i}$  is the explanatory data matrix,  $\beta_i$  is the regression vector of size  $(L-1) \times 1$ , and  $\xi_i$  is the modeling error vector of size  $N \times 1$ . For each  $i \in \{1, \dots, L\}$ , the least squares estimator of the regression vector  $\beta_i$  is given by

$$\hat{\beta}_i = (\mathbf{Z}_{\partial i}^T \mathbf{Z}_{\partial i})^{-1} \mathbf{Z}_{\partial i}^T \mathbf{z}_i. \quad (9)$$

The noise is estimated by

$$\hat{\xi}_i = \mathbf{z}_i - \mathbf{Z}_{\partial i} \hat{\beta}_i \quad (10)$$

and the correlation matrix by  $\hat{\mathbf{R}}_n = [\hat{\xi}_1, \dots, \hat{\xi}_N]^T [\hat{\xi}_1, \dots, \hat{\xi}_N] / N$ . Notice that the determination of each noise vector  $\hat{\xi}_i$  implies the computation of the pseudoinverse  $\mathbf{Z}_{\partial i}^\# = (\mathbf{Z}_{\partial i}^T \mathbf{Z}_{\partial i})^{-1} \mathbf{Z}_{\partial i}^T$ , of size  $(L-1) \times (L-1)$ , which is for each  $i = 1, \dots, L$ . This computational complexity can, however, be greatly reduced by taking advantage of the relation between  $\mathbf{Z}_{\partial i}^\#$  and  $\mathbf{Z}$ . Let the  $L \times L$  symmetric and positive

<sup>1</sup>The chordal distance is based on the projection matrix computed from directions  $\mathbf{v}_i$  obtained from (7).



definite matrices  $\mathbf{R}$  and  $\mathbf{R}^{-1}$  be partitioned into block matrices as follows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{bmatrix} \quad \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{A}' & \mathbf{b}' \\ \mathbf{b}'^T & c' \end{bmatrix} \quad (11)$$

where  $\mathbf{A}$  and  $\mathbf{A}'$  are  $(L-1) \times (L-1)$  matrices,  $\mathbf{b}$  and  $\mathbf{b}'$  are  $(L-1) \times 1$  vectors, and  $c$  and  $c'$  are scalars. As  $\mathbf{R}$ ,  $\mathbf{R}^{-1}$ ,  $\mathbf{A}$ ,  $\mathbf{A}^{-1}$ , and  $c'$  are positive definites, thus

$$\mathbf{A}\mathbf{A}' + \mathbf{b}\mathbf{b}'^T = \mathbf{I}_{L-1} \quad (12)$$

$$\mathbf{A}\mathbf{b}' + \mathbf{b}c' = \mathbf{0}_{L-1} \quad (13)$$

where  $\mathbf{0}_{L-1}$  is an  $(L-1) \times 1$  vector with null entries. By replacing  $\mathbf{A}^{-1}\mathbf{b} = -\mathbf{b}'/c'$ , which is derived from (13), into (12), we obtain

$$\mathbf{A}^{-1} = \mathbf{A}' - \mathbf{b}'\mathbf{b}'^T/c'. \quad (14)$$

Based on this relation, the inversion of the matrix  $\mathbf{Z}_{\partial_i}^T \mathbf{Z}_{\partial_i}$ , which is for  $i = 1, \dots, L$ , can be obtained by removing the  $i$ th row and the  $i$ th column of the matrix  $(\mathbf{Z}^T \mathbf{Z})^{-1}$  and implementing (14) with the necessary adjustments.

The pseudocode for the noise estimation is shown in the Algorithm 1. The symbol  $[\hat{\mathbf{R}}]_{\partial_i, \partial_i}$  denotes the matrix obtained from  $\hat{\mathbf{R}}$  by deleting the  $i$ th row and the  $i$ th column,  $[\hat{\mathbf{R}}]_{i, \partial_i}$  denotes the  $i$ th row of  $[\hat{\mathbf{R}}]_{\cdot, \partial_i}$ , and  $[\hat{\mathbf{R}}]_{\partial_i, i}$  denotes  $[\hat{\mathbf{R}}]_{i, \partial_i}^T$ . Steps 2) and 3) compute the matrix  $\hat{\mathbf{R}} = \mathbf{Z}^T \mathbf{Z}$  and its inverse, respectively. Steps 5) and 6) estimate, respectively, the regression vector  $\hat{\beta}_i$  and the noise  $\hat{\xi}_i$ , which are for each  $i = 1, \dots, L$ .

#### Algorithm 1: Noise estimation

- 1) INPUT  $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$
- 2)  $\mathbf{Z} := \mathbf{Y}^T$ ,  $\hat{\mathbf{R}} := (\mathbf{Z}^T \mathbf{Z})$ ;
- 3)  $\mathbf{R}' := \hat{\mathbf{R}}^{-1}$ ;
- 4) for  $i := 1$  to  $L$  do
  - 5)  $\hat{\beta}_i := ([\mathbf{R}']_{\partial_i, \partial_i} - [\mathbf{R}']_{\partial_i, i}[\mathbf{R}']_{i, \partial_i}/[\mathbf{R}']_{i, i})[\hat{\mathbf{R}}]_{\partial_i, i}$ ;  
 {Note that  $\partial_i = 1, \dots, i-1, i+1, \dots, L$ }
  - 6)  $\hat{\xi}_i := \mathbf{z}_i - \mathbf{Z}_{\partial_i} \hat{\beta}_i$ ;
- 7) end for
- 8) OUTPUT  $\hat{\xi}$ ;  
 { $\hat{\xi}$  is an  $N \times L$  matrix with the estimated noise}

The main advantage of Algorithm 1 is that the computation of  $\hat{\mathbf{R}}$  and of  $\mathbf{R}' = \hat{\mathbf{R}}^{-1}$  are out of the *for* loop. Thus, the computational complexity, i.e., the number of floating-point operations (flops), of Algorithm 1 is substantially lower than that of an algorithm implementing the multiple regression without using the relation (14). Note that the computation of the observed data correlation matrix and of its inversion demands, approximately,  $2NL^2 + L^3$  flops, whereas the multiple regression algorithm, without using (14), has to compute the aforementioned matrices  $L$  times, thus demanding, approximately,  $2NL^3 + L^4$  flops. Table I presents the approximated expressions for the number of floating-point operations used by each algorithm. For  $N \gg L$ , Algorithm 1 demands approximately  $L/2$  less flops, which

TABLE I  
COMPUTATIONAL COMPLEXITY OF THE NOISE ESTIMATION ALGORITHMS

Algorithm 1	$4NL^2 + 6L^3$
Algorithm without relation (14)	$4NL^2 + 2NL^3 + L^4$

is a significant figure, considering that  $L/2$  takes, in many applications, values on the order of 100. For example, the computational complexity of Algorithm 1, when applied to a subimage ( $145 \times 145$  pixels and 224 bands) of the Indian Pines test site acquired by the AVIRIS instrument is 2.8-G flops, whereas the algorithm implemented without using (14) has a computational complexity of 253.1-G flops.

The next experiments illustrate Algorithm 1 working. The input data is a simulated hyperspectral image composed of  $10^4$  spectral vectors, each one following the linear mixing model (1). The abundance fractions are generated according to a Dirichlet distribution; the number of endmembers is set to  $p = 5$ , and their signatures are selected from the USGS digital spectral library. The noise is zero-mean independent with variances along the bands, following the Gaussian shape (4) [see Fig. 1(a)]. The value of the parameter  $\sigma^2$  in (4) is defined by the desired SNR.

Fig. 3(a) shows the noiseless spectral vector  $\mathbf{x}$  and the noisy version  $\mathbf{x} + \mathbf{n}$ , which are for SNR = 20 dB and  $\eta = 18$ . Fig. 3(b) shows the true and the estimated noises for the same SNR. The improvement in the SNR (i.e.,  $E[\|\mathbf{n}\|^2]/E[\|\hat{\mathbf{x}} - \mathbf{x}\|^2]$ ) is about 13 dB. Fig. 4 shows the plots of three noise covariances curves and the respective estimates, which are as a function of the band, for the three experiments. The SNR and covariance parameter  $\eta$  are the following: (a) SNR = 10 dB, and  $\eta = 18$ ; (b) SNR = 10 dB, and  $\eta = 72$ ; and (c) SNR = 20 dB, and  $\eta = 18$ . Notice that the three estimates are very accurate.

#### B. Signal Subspace Estimation

This section presents the core structure of the proposed method. The first step, which is based on the noise estimation procedure introduced in the previous section, identifies a set of orthogonal directions, of which an unknown subset spans the signal subspace. This subset is then determined by seeking the minimum mse between  $\mathbf{x}$ , the original signal, and a noisy projection of it obtained from the vector  $\mathbf{y} = \mathbf{x} + \mathbf{n}$ . In the following, we assume that  $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{R}}_n)$ , i.e., the noise is zero-mean Gaussian distributed with the covariance matrix  $\hat{\mathbf{R}}_n$ .

Let the eigen decomposition of the signal sample correlation matrix  $\hat{\mathbf{R}}_x = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N][\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N]^T/N$  be written as

$$\hat{\mathbf{R}}_x = \mathbf{E}\mathbf{\Sigma}\mathbf{E}^T \quad (15)$$

where  $\mathbf{E} \equiv [\mathbf{e}_1, \dots, \mathbf{e}_L]$  is a matrix with the eigenvectors of  $\hat{\mathbf{R}}_x$ . Given a permutation  $\pi = \{i_1, \dots, i_L\}$  of indices  $i = 1, \dots, L$ , let us decompose the space  $\mathbb{R}^L$  into two orthogonal subspaces, namely: the  $k$ -D subspace  $\langle \mathbf{E}_k \rangle$  spanned by  $\mathbf{E}_k \equiv [\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k}]$  and  $\langle \mathbf{E}_k \rangle^\perp$  spanned by  $\mathbf{E}_k^\perp \equiv [\mathbf{e}_{i_{k+1}}, \dots, \mathbf{e}_{i_L}]$ , i.e., the orthogonal complement of the subspace  $\mathbf{E}_k$ .

Let  $\mathbf{U}_k = \mathbf{E}_k \mathbf{E}_k^T$  be the projection matrix onto  $\langle \mathbf{E}_k \rangle$  and  $\hat{\mathbf{x}}_k \equiv \mathbf{U}_k \mathbf{y}$  be the projection of the observed spectral vector

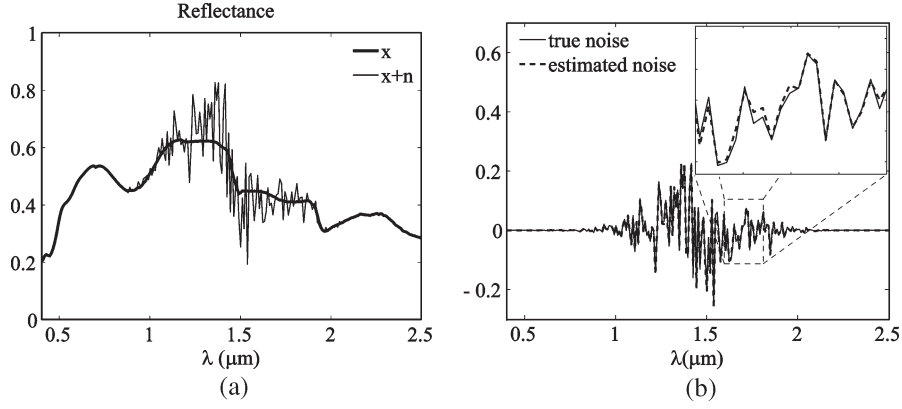


Fig. 3. Illustration of the noise estimation algorithm. (a) (Thick line) Reflectance signal of a pixel. (Thin line) Reflectance of the same pixel with noise. (b) (Solid line) True noise. (Dashed line) Estimated noise.

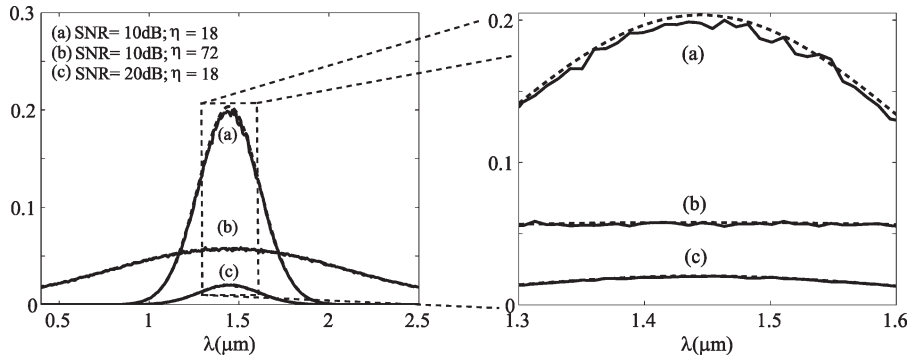


Fig. 4. Illustration of the noise estimation algorithm. (Solid line) Diagonal of the estimated noise covariance matrix. (Dashed line) Diagonal of the noise covariance matrix for three scenarios. (a) SNR = 10 dB and  $\eta = 18$ . (b) SNR = 10 dB and  $\eta = 72$ . (c) SNR = 20 dB and  $\eta = 18$ .

$\mathbf{y}$  onto the subspace  $\langle \mathbf{E}_k \rangle$ . The first- and the second-order moments of  $\hat{\mathbf{x}}_k$  given  $\mathbf{x}$  are

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{x}}_k | \mathbf{x}] &= \mathbf{U}_k \mathbb{E}[\mathbf{y} | \mathbf{x}] \\ &= \mathbf{U}_k \mathbb{E}[\mathbf{x} + \mathbf{n} | \mathbf{x}] \\ &= \mathbf{U}_k \mathbf{x} \\ &\equiv \mathbf{x}_k \end{aligned} \quad (16)$$

$$\begin{aligned} \mathbb{E}[(\hat{\mathbf{x}}_k - \mathbf{x}_k)(\hat{\mathbf{x}}_k - \mathbf{x}_k)^T | \mathbf{x}] &= \mathbb{E}[(\mathbf{U}_k \mathbf{y} - \mathbf{U}_k \mathbf{x}) \\ &\quad \times (\mathbf{U}_k \mathbf{y} - \mathbf{U}_k \mathbf{x})^T | \mathbf{x}] \\ &= \mathbb{E}[(\mathbf{U}_k \mathbf{n} \mathbf{n}^T \mathbf{U}_k^T) | \mathbf{x}] \\ &= \mathbf{U}_k \hat{\mathbf{R}}_n \mathbf{U}_k^T. \end{aligned} \quad (17)$$

The mse between  $\mathbf{x}$  and  $\hat{\mathbf{x}}_k$  is

$$\begin{aligned} \text{mse}(k | \mathbf{x}) &= \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}_k)^T (\mathbf{x} - \hat{\mathbf{x}}_k) | \mathbf{x}] \\ &= \mathbb{E} \left[ \underbrace{(\mathbf{x} - \mathbf{x}_k)}_{\mathbf{b}_k} - \mathbf{U}_k \mathbf{n} \right)^T \left( \underbrace{(\mathbf{x} - \mathbf{x}_k)}_{\mathbf{b}_k} - \mathbf{U}_k \mathbf{n} \right) | \mathbf{x} \right] \\ &= \mathbf{b}_k^T \mathbf{b}_k + \text{tr}(\mathbf{U}_k \hat{\mathbf{R}}_n \mathbf{U}_k^T). \end{aligned} \quad (18)$$

Computing the mean of (18) with respect to  $\mathbf{x}$ , noting that  $\mathbf{b}_k = \mathbf{x} - \mathbf{x}_k = \mathbf{U}_k^\perp \mathbf{x}$  and using the properties  $\mathbf{U} = \mathbf{U}^T$ ,  $\mathbf{U}^2 = \mathbf{U}$ ,

and  $\mathbf{U}^\perp = \mathbf{I} - \mathbf{U}$  of the projection matrices, we get

$$\begin{aligned} \text{mse}(k) &= \mathbb{E}[(\mathbf{U}_k^\perp \mathbf{x})^T (\mathbf{U}_k^\perp \mathbf{x})] + \text{tr}(\mathbf{U}_k \hat{\mathbf{R}}_n \mathbf{U}_k^T) \\ &= \text{tr}(\mathbf{U}_k^\perp \mathbf{R}_x) + \text{tr}(\mathbf{U}_k \hat{\mathbf{R}}_n) \\ &= \text{tr}(\mathbf{U}_k^\perp \mathbf{R}_y) + 2\text{tr}(\mathbf{U}_k \hat{\mathbf{R}}_n) + c \end{aligned} \quad (19)$$

where  $c$  is an irrelevant constant. The criterion that we propose to estimate the signal subspace, which is called  $X$ , is the minimization of  $\text{mse}(k)$  given by (19), with respect to all the permutations  $\pi = \{i_1, \dots, i_L\}$  of size  $L$  and to  $k$ , with the correlation matrix  $\mathbf{R}_y$  replaced with the sample correlation matrix  $\hat{\mathbf{R}}_y = \mathbf{Y}\mathbf{Y}^T/N$ , i.e.,

$$\hat{X} = \left\langle \left[ \mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k} \right] \right\rangle \quad (20)$$

$$(\hat{k}, \hat{\pi}) = \arg \min_{k, \pi} \left\{ \text{tr}(\mathbf{U}_k^\perp \hat{\mathbf{R}}_y) + 2\text{tr}(\mathbf{U}_k \hat{\mathbf{R}}_n) \right\} \quad (21)$$

where the dependence on the permutation  $\pi$  is through  $\mathbf{U}_k = \mathbf{E}_k \mathbf{E}_k^T$ , considering that  $\mathbf{E}_k \equiv [\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_k}]$ . For a given permutation  $\pi$ , each term of (21) has a clear meaning, namely: the first term accounts for the projection error power and is a decreasing function of  $k$ ; the second term accounts for the noise power and is an increasing function of  $k$ .

By exploiting, again, the fact that the  $\mathbf{U}_k$  is a projection matrix and that  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ , which is for  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{L \times L}$ , the minimization (21) can be rewritten as

$$(\hat{k}, \hat{\pi}) = \arg \min_{k, \pi} \left\{ c + \sum_{j=1}^k \underbrace{-p_{i_j} + 2\sigma_{i_j}^2}_{\delta_{i_j}} \right\} \quad (22)$$

where  $c$  is an irrelevant constant and  $p_{i_j}$  and  $\sigma_{i_j}^2$  are quadratic forms given by

$$p_{i_j} = \mathbf{e}_{i_j}^T \hat{\mathbf{R}}_y \mathbf{e}_{i_j} \quad (23)$$

$$\sigma_{i_j}^2 = \mathbf{e}_{i_j}^T \hat{\mathbf{R}}_n \mathbf{e}_{i_j}. \quad (24)$$

Based on the right-hand side of (22), it follows that the corresponding minimization is achieved simply by including all the negative terms  $\delta_i$ , which are for  $i = 1, \dots, L$ , and only these, in the sum.

The pseudocode for HySime is shown in Algorithm 2. HySime inputs are the spectral observed vectors and the sample correlation matrix  $\hat{\mathbf{R}}_y$ . Step 2) estimates the noise correlation matrix  $\hat{\mathbf{R}}_n$ . Step 3) estimates the signal correlation matrix  $\hat{\mathbf{R}}_x$ . Steps 4) and 5) calculate the eigenvectors of the signal correlation matrix and the terms  $\delta_i$  based on the quadratic forms (23) and (24). Steps 6) and 7) implement the minimization (22). Finally, Step 8) retrieves the signal subspace from  $\hat{k}$  and  $\hat{\pi}$ .

#### Algorithm 2: HySime

- 1) INPUT  $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$ ,  $\hat{\mathbf{R}}_y \equiv (\mathbf{Y}\mathbf{Y}^T)/N$ ;
- 2)  $\hat{\mathbf{R}}_n := 1/N \sum_i (\hat{\xi}_i \hat{\xi}_i^T)$ ;  $\{\hat{\xi}_i\}$  is given by (10)
- 3)  $\hat{\mathbf{R}}_x := 1/N \sum_i ((\mathbf{y}_i - \hat{\xi}_i)(\mathbf{y}_i - \hat{\xi}_i^T))$ ; {estimate of  $\hat{\mathbf{R}}_x$ }
- 4)  $\mathbf{E} := [\mathbf{e}_1, \dots, \mathbf{e}_L]$ ;  $\{\mathbf{e}_i\}$  are the eigenvectors of  $\hat{\mathbf{R}}_x$ }
- 5)  $\delta := [\delta_1, \dots, \delta_L]$ ;  $\{\delta_i\}$  is given by (22)
- 6)  $(\hat{\delta}, \hat{\pi}) := \text{sort}(\delta)$ ;  
 {sort  $\delta_i$  by ascending order; save the permutation  $\hat{\pi}$ }
- 7)  $\hat{k} := \text{number of terms } \hat{\delta}_i < 0$ ;
- 8)  $\hat{\mathbf{X}} = \langle [\mathbf{e}_{\hat{\pi}_1}, \dots, \mathbf{e}_{\hat{\pi}_k}] \rangle$ ; {signal subspace}

An alternative to the projection  $\mathbf{U}_k^\perp \hat{\mathbf{R}}_y$  consists of projecting the sample mean  $\bar{\mathbf{y}} \equiv 1/N \sum_{i=1}^N \mathbf{y}_i$  onto the subspace that is orthogonal to the signal subspace. This alternative method, which is called HySime<sub>m</sub>, has been developed in [1] and [2]. The underlying rationale is that spectral vectors are nonnegative, and then,  $\bar{\mathbf{y}}$  accumulates information about every spectral vector in the data set. Of course, for this approach to work, the projection of  $\bar{\mathbf{y}}$  onto any eigenvector vector  $\mathbf{e}_j$ ,  $j = 1, \dots, k$  must be nonzero. Although we have no proof of this statement, we believe, supported by practical evidence, that the probability of  $\bar{\mathbf{y}}^T \mathbf{e}_j = 0$  is practically zero in real data sets. Another difference between HySime and HySime<sub>m</sub> is that the latter, which is as proposed in [1] and [2], does not search the signal subspace over all the permutations  $\pi$  of the eigenvectors; instead, it searches only over the sequence of nested subspaces  $\mathbf{S}_k = [\mathbf{e}_1, \dots, \mathbf{e}_k]$ , where the eigenvectors are ordered by decreasing magnitude of the corresponding eigenvalues. We give

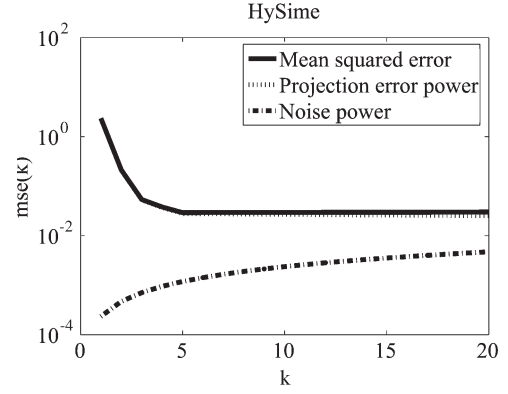


Fig. 5. Plot of mse versus  $k$ , with SNR = 35 dB and  $p = 5$  for HySime method.

evidence, in the next section, that HySime slightly outperforms HySime<sub>m</sub>.

#### IV. EVALUATION OF HYSIME WITH SIMULATED DATA

In this section, we apply the proposed HySime algorithm to simulated scenes and compare it with HFC and NWHFC eigen-based Neyman–Pearson detectors [43]. As concluded in [43], these algorithms are the state of the art in hyperspectral signal subspace identification, outperforming the information theoretical criteria approaches, namely, the MDL [40], [41] and the AIC [42]. HySime and HySime<sub>m</sub> algorithms are also compared, given their structural similarity.

The spectral signatures are selected from the USGS digital spectral library [48]. The abundance fractions are generated according to a Dirichlet distribution defined in (3). The results presented here are organized into the following two experiments. In the first experiment, the method is evaluated with respect to the SNR [see (5)] to the number of endmembers  $p$ , and to the spectral noise shape (white and nonwhite). In the second experiment, the methods are evaluated with respect to their ability to detect rare pixels.

*Experiment 1:* Fig. 5 shows the evolution of the mse for the HySime method as a function of the parameter  $k$ , which is for SNR = 35 dB and  $p = 5$ . The minimum of the mse occurs at  $k = 5$ , which is exactly the number of endmembers present in the image. As expected, the projection of the error and noise powers displays decreasing and increasing behaviors, respectively, as a function of the subspace dimension  $k$ .

Tables II and III present the signal subspace order estimates yielded by the HySime and HySime<sub>m</sub> algorithms and the VD determined by the NWHFC and the HFC algorithms [43] as a function of the SNR and of the number of endmembers  $p$  for  $\eta = 0$  and  $\eta = 1/18$ , respectively.

The NWHFC algorithm is basically the HFC one [45] preceded by a noise-whitening step, which is based on the estimated noise correlation matrix. In implementing this step, we got poor results in very high SNRs and colored-noise scenarios. This is basically because the noise estimation step in NWHFC needs to invert the noise correlation matrix, which gives inaccurate results when the noise power is small. For this reason, we have used both the true and estimated noise

TABLE II  
SIGNAL SUBSPACE DIMENSION  $k$ , WHICH IS BASED ON 50 MONTE CARLO RUNS, AS A FUNCTION OF SNR AND  $p$  FOR  $\eta = 0$  (NOISE SHAPE). FIGURES IN BRACKETS WERE COMPUTED BASED ON THE TRUE NOISE STATISTICS

SNR	Method	White Noise ( $\eta = 0$ )			
		$p = 3$	$p = 5$	$p = 10$	$p = 15$
50 dB	HySime	<b>3</b>	<b>5</b>	<b>10</b>	<b>15</b>
	HySime <sub>m</sub>	3	5	10	15
	HFC ( $P_f = 10^{-3}$ )	3	5	7	11
	HFC ( $P_f = 10^{-4}$ )	3	5	7	8
	HFC ( $P_f = 10^{-5}$ )	3	4	6	8
	NWHFC ( $P_f = 10^{-3}$ )	3 (3)	5 (5)	7 (7)	10 (11)
	NWHFC ( $P_f = 10^{-4}$ )	3 (3)	5 (5)	7 (7)	8 (8)
	NWHFC ( $P_f = 10^{-5}$ )	3 (3)	4 (4)	7 (6)	8 (8)
35 dB	HySime	<b>3</b>	<b>5</b>	<b>10</b>	<b>15</b>
	HySime <sub>m</sub>	3	5	10	15
	HFC ( $P_f = 10^{-3}$ )	3	4	7	9
	HFC ( $P_f = 10^{-4}$ )	3	4	6	8
	HFC ( $P_f = 10^{-5}$ )	3	4	6	8
	NWHFC ( $P_f = 10^{-3}$ )	3 (3)	4 (4)	7 (7)	9 (9)
	NWHFC ( $P_f = 10^{-4}$ )	3 (3)	4 (4)	7 (6)	8 (8)
	NWHFC ( $P_f = 10^{-5}$ )	3 (3)	4 (4)	6 (6)	8 (8)
25 dB	HySime	<b>3</b>	<b>5</b>	<b>10</b>	<b>14</b>
	HySime <sub>m</sub>	3	5	9	12
	HFC ( $P_f = 10^{-3}$ )	3	5	6	8
	HFC ( $P_f = 10^{-4}$ )	3	5	6	7
	HFC ( $P_f = 10^{-5}$ )	3	4	5	7
	NWHFC ( $P_f = 10^{-3}$ )	3 (3)	5 (5)	6 (6)	9 (8)
	NWHFC ( $P_f = 10^{-4}$ )	3 (3)	5 (5)	6 (6)	7 (7)
	NWHFC ( $P_f = 10^{-5}$ )	3 (3)	4 (4)	5 (5)	7 (7)
15 dB	HySime	<b>3</b>	<b>5</b>	<b>8</b>	<b>12</b>
	HySime <sub>m</sub>	3	3	6	8
	HFC ( $P_f = 10^{-3}$ )	3	5	4	5
	HFC ( $P_f = 10^{-4}$ )	3	4	3	2
	HFC ( $P_f = 10^{-5}$ )	3	4	3	2
	NWHFC ( $P_f = 10^{-3}$ )	3 (3)	5 (5)	5 (4)	5 (5)
	NWHFC ( $P_f = 10^{-4}$ )	3 (3)	4 (4)	3 (3)	3 (2)
	NWHFC ( $P_f = 10^{-5}$ )	3 (3)	4 (4)	3 (3)	2 (2)

correlation matrices. The results based on the true correlation matrix are in brackets. We stress that for the setting of this experiment both HySime and HySime<sub>m</sub> methods yield the same results, whether using the estimated or the true noise correlation matrices. Table III (noise-colored scenario,  $\eta = 1/18$ ) does not present results for the HFC method because this method was designed only for white noise.

Another central issue of NWHFC and HFC algorithms is the false-alarm probability  $P_f$  they are parameterized with. This probability is used in a series of Neyman–Pearson tests, each one designed to detect a different orthogonal signal subspace direction. There is the need, therefore, to specify the false-alarm probability  $P_f$  of the tests. Based on the hints given in [43] and in our own results, we choose  $P_f \in \{10^{-3}, 10^{-4}, 10^{-5}\}$ .

TABLE III  
SIGNAL SUBSPACE DIMENSION  $k$ , WHICH IS BASED ON 50 MONTE CARLO RUNS, AS A FUNCTION OF SNR AND  $p$  FOR  $\eta = 1/18$  (NOISE SHAPE). FIGURES IN BRACKETS WERE COMPUTED BASED ON THE TRUE NOISE STATISTICS

SNR	Method	Gaussian shaped noise ( $\eta = 1/18$ )			
		$p = 3$	$p = 5$	$p = 10$	$p = 15$
50 dB	HySime	<b>3</b>	<b>5</b>	<b>10</b>	<b>15</b>
	HySime <sub>m</sub>	3	5	10	15
	NWHFC ( $P_f = 10^{-3}$ )	59 (3)	41 (5)	61 (10)	45 (10)
	NWHFC ( $P_f = 10^{-4}$ )	48 (3)	33 (5)	54 (10)	34 (10)
	NWHFC ( $P_f = 10^{-5}$ )	43 (3)	28 (5)	41 (9)	27 (10)
35 dB	HySime	<b>3</b>	<b>5</b>	<b>10</b>	<b>15</b>
	HySime <sub>m</sub>	3	5	10	15
	NWHFC ( $P_f = 10^{-3}$ )	9 (3)	10 (5)	12 (10)	10 (10)
	NWHFC ( $P_f = 10^{-4}$ )	9 (3)	9 (5)	11 (10)	8 (10)
25 dB	HySime	<b>3</b>	<b>5</b>	<b>10</b>	<b>15</b>
	HySime <sub>m</sub>	3	5	10	12
	NWHFC ( $P_f = 10^{-3}$ )	4 (3)	5 (5)	11 (10)	9 (11)
	NWHFC ( $P_f = 10^{-4}$ )	4 (3)	5 (5)	11 (10)	9 (10)
15 dB	HySime	<b>3</b>	<b>5</b>	<b>8</b>	<b>12</b>
	HySime <sub>m</sub>	3	3	5	8
	NWHFC ( $P_f = 10^{-3}$ )	4 (3)	5 (5)	11 (10)	10 (10)
	NWHFC ( $P_f = 10^{-4}$ )	4 (3)	5 (5)	11 (10)	8 (10)
	NWHFC ( $P_f = 10^{-5}$ )	4 (3)	5 (5)	11 (9)	8 (10)

The figures shown in Tables II and III, which are based on 50 Monte Carlo runs, display the following behavior.

- 1) HySime and HySime<sub>m</sub> algorithms show similar performance, which is with a small advantage for the former, namely, at small SNRs and colored noise.
- 2) HySime and NWHFC algorithms parameterized with  $P_f = 10^{-3}$  show similar performances at low subspace dimension, for example  $p \leq 5$ , and white noise. This is also true for colored noise and NWHFC working with known a noise covariance matrix. However, if the noise statistics are unknown, NWHFC performs much worse than HySime.
- 3) HySime performs better than NWHFC for high space dimensions, for example  $p > 5$ .

We conclude, therefore, that the HySime algorithm is slightly better than the HySime<sub>m</sub> one, yielding systematically equal or better results than the NWHFC and HFC algorithms. Another advantage of the HySime approach is that it does not depend on any tunable parameter.

*Experiment 2:* In this experiment, we set SNR = 35 dB and  $p = 8$ . The first five endmembers are mixed according to a Dirichlet distribution, where, as in the previous experiment, the sixth, the seventh, and the eighth endmembers are present as pure in 8, 4, and 2 pixels, respectively. Fig. 6 shows the mse versus the subspace dimension  $k$  for the HySime method. The minimum of the mse occurs at  $k = 8$ . Thus, the HySime algorithm infers the correct subspace dimension.



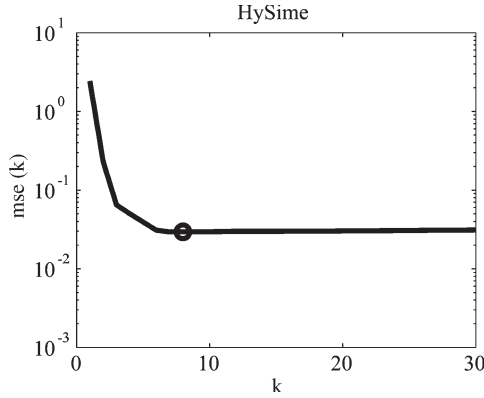


Fig. 6. Plot of mse versus  $k$ , with SNR = 35 dB and  $p = 8$  (second experiment).

TABLE IV  
SIGNAL SUBSPACE DIMENSION  $k$  AS A FUNCTION OF PARAMETER  $\eta$   
IN EXPERIMENT 2 (FEW PURE PIXELS), WITH SNR = 35 dB

Method ( $p = 8$ )	$\eta = 0$	$\eta = 1/18$
HySime	8	7
HySime <sub>m</sub>	7	7
HFC ( $P_f = 10^{-3}$ )	6	-
HFC ( $P_f = 10^{-4}$ )	6	-
HFC ( $P_f = 10^{-5}$ )	6	-
NWHFC ( $P_f = 10^{-3}$ )	6	13
NWHFC ( $P_f = 10^{-4}$ )	6	11
NWHFC ( $P_f = 10^{-5}$ )	6	11

Table IV displays the results of this experiment computed by HySime, HySime<sub>m</sub>, HFC, and NWHFC algorithms. The HySime method systematically yields the best performance. We observe the same pattern of behavior shown in Tables II and III.

## V. EXPERIMENTS WITH REAL HYPERSPECTRAL DATA

In this section, the proposed method is applied to real hyperspectral data collected by the AVIRIS [55]. A subset of the Indian Pines test site in northwestern Indiana<sup>2</sup> acquired by AVIRIS in June 1992 is considered. The data set is composed of 185 spectral bands with 10-nm bandwidth acquired in the 0.4–2.5  $\mu\text{m}$  region (noisy and water absorption bands {1–4, 103–113, and 148–166} were removed). The data set contains  $145 \times 145$  pixels (21 025 pixels) with a ground resolution of 17 m [56]. This observed region contains a mixture of agriculture and forestry. There is a major dual-lane highway (U.S. 52 and U.S. 231), a rail line crossing near the top, a major secondary road (Jackson Highway) near the middle, several other county roads, and houses (Fig. 7 shows band 29 of the data set). A detailed ground truth map for this region is published in [57]. Table V lists the 16 ground truth materials considered in [57].

Table VI shows the signal subspace dimension inferred by HySime, HySime<sub>m</sub>, HFC, and NWHFC. The value obtained by HySime coincides with the number of ground truth materials

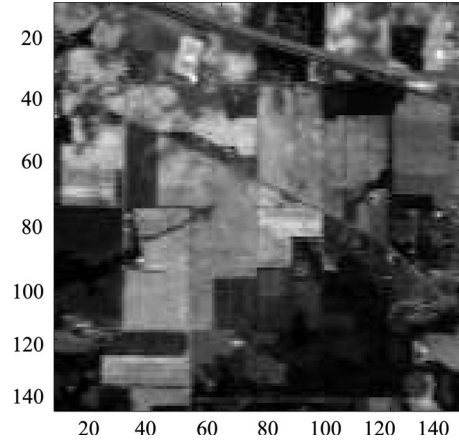


Fig. 7. Band 29 ( $\lambda = 0.666 \mu\text{m}$ ) of the Indian Pines hyperspectral data set.

TABLE V  
SUBSTANCES DETECTED ON THE INDIAN PINES TEST SITE

Alfalfa	Oats
Corn-notill	Soybeans-notill
Corn-min	Soybeans-min
Corn	Soybeans-clean
Grass/Pasture	Wheat
Grass/Trees	Woods
Grass/pasture-mowed	Bldg-Grass-Tree-drives
Hay-windrowed	Stone-steel towers

TABLE VI  
INFERRED SIGNAL SUBSPACE DIMENSION  $\hat{k}$   
FOR THE INDIAN PINES DATA SET

Method	$\hat{k}$
HySime	16
HySime <sub>m</sub>	12
HFC ( $P_f = 10^{-3}$ )	25
HFC ( $P_f = 10^{-4}$ )	22
HFC ( $P_f = 10^{-5}$ )	21
NWHFC ( $P_f = 10^{-3}$ )	18
NWHFC ( $P_f = 10^{-4}$ )	18
NWHFC ( $P_f = 10^{-5}$ )	18

listed in Table V. Although these results are an indicator of the HySime competitiveness, a better assessment of each algorithm performance is obtained by computing  $(\|\mathbf{E}_k^\perp \mathbf{x}\|^2) / (\|\mathbf{x}\|^2)$ , i.e., the relative power of the signal component that is orthogonal to the identified subspace and, therefore, not modeled by this subspace. Fig. 8(a)–(d) shows this ratio in gray level and for each pixel; we have used  $\hat{\mathbf{x}} = \mathbf{y} - \hat{\mathbf{n}}$ , with  $\hat{\mathbf{n}}$  given by (10), as an estimate of  $\mathbf{x}$ . HySime yielded the lowest values followed by HySime<sub>m</sub>, HFC, and NWHFC. The relative advantage of HySime over HySime<sub>m</sub>, HFC, and NWHFC is on the order of 10, 20, and  $10^3$ , respectively. Furthermore, there is no perceptible structure in the HySime relative error, which is not the case, at least with NWHFC, where the agriculture fields are clearly perceived, indicating the presence of signal where it should not be.

<sup>2</sup> Available at <http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/>.

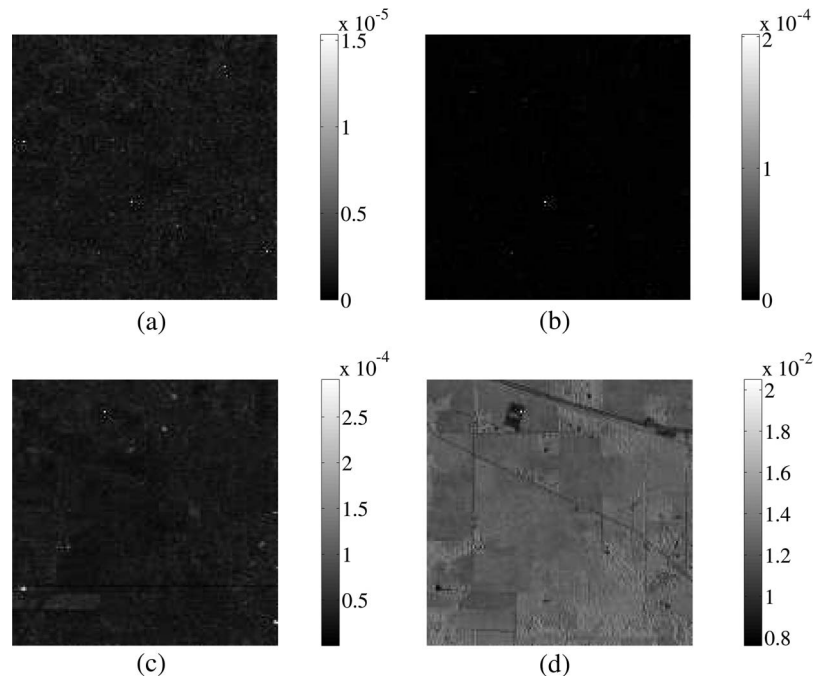


Fig. 8. Relative power of the signal component that is orthogonal to the identified subspace  $\langle \mathbf{E}_k \rangle$  for the methods (a) HySime, (b) HySime<sub>m</sub>, (c) HFC, and (d) NWHFC.

## VI. CONCLUSION

The huge volumes and rates of data generated by hyperspectral sensors demand expensive processors with very high performance and memory capacities. Dimensionality reduction is, therefore, a relevant first step in the hyperspectral data processing chain. This paper introduced the HySime algorithm, which is a new computationally efficient and automatic (in the sense that it dispenses with any tunable parameter) approach to infer the signal subspace in hyperspectral imagery. HySime estimates the signal and the noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the minimum mse sense. A set of experiments with simulated and real data led to the conclusion that the HySime is an effective and useful tool, yielding comparable or better results than the state-of-the-art algorithms.

## ACKNOWLEDGMENT

The authors would like to thank Dr. C.-I Chang and W.-M. Liu for the valuable comments and suggestions about the implementation of HFC and NWHFC algorithms. This paper elaborates on and extends works [1] and [2].

## REFERENCES

- [1] J. M. Bioucas-Dias and J. M. P. Nascimento, "Estimation of signal subspace on hyperspectral data," in *Proc. SPIE—Conf. Image and Signal Processing for Remote Sensing XI*, L. Bruzzone, Ed., 2005, vol. 5982, pp. 191–198.
- [2] J. M. P. Nascimento and J. M. Bioucas-Dias, "Signal subspace identification in hyperspectral linear mixtures," in *Pattern Recognition and Image Analysis*, vol. 3523, J. S. Marques, N. P. de la Blanca, and P. Pina, Eds. New York: Springer-Verlag, 2005, pp. 207–214.
- [3] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed. Hoboken, NJ: Wiley, 2004.
- [4] J. P. Kerekes and J. E. Baum, "Spectral imaging system analytical model for subpixel object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1088–1101, May 2002.
- [5] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [6] M. E. Winter, "Fast autonomous spectral endmember determination in hyperspectral data," in *Proc. 13th Int. Conf. Appl. Geologic Remote Sens.*, 1999, vol. II, pp. 337–344.
- [7] F. V. D. Meer, "Iterative spectral unmixing," *Int. J. Remote Sens.*, vol. 20, no. 17, pp. 3431–3436, Nov. 1999.
- [8] J. H. Bowles, P. J. Palmadesso, J. A. Antoniadis, M. M. Baumbach, and L. J. Rickard, "Use of filter vectors in hyperspectral data analysis," in *Proc. SPIE—Conf. Infrared Spaceborne Remote Sensing III*, 1995, vol. 2553, pp. 148–157.
- [9] J. W. Boardman, "Analysis, understanding, and visualization of hyperspectral data as convex sets in n-space," in *Proc. SPIE—Conf. Imaging Spectrometry*, M. R. Descour, J. M. Mooney, D. L. Perry, and L. R. Illing, Eds., 1995, vol. 2480, pp. 14–22.
- [10] M. O. Smith, J. B. Adams, and D. E. Sabol, *Spectral Mixture Analysis—New Strategies For the Analysis of Multispectral Data*. Brussels, Belgium: Kluwer, 1994, pp. 125–143.
- [11] J. B. Adams, M. Smith, and A. R. Gillespie, *Imaging Spectroscopy: Interpretation Based on Spectral Mixture Analysis*. New York: Cambridge Univ. Press, 1993, pp. 145–166.
- [12] J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop, AVIRIS Workshop*, 1993, vol. 1, pp. 11–14. JPL Pub. 93-26.
- [13] J. B. Adams and M. O. Smith, "Spectral mixing model: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res.*, vol. 91, no. B8, pp. 8098–8112, 1986.
- [14] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
- [15] C. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
- [16] S. D. Backer, P. Kempeneers, W. Debruyne, and P. Scheunders, "A band selection technique for spectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 319–323, Jul. 2005.
- [17] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 156–159, Apr. 2005.
- [18] S. S. Shen and E. M. Bassett, "Information-theory-based band selection and utility evaluation for reflective spectral systems," in *Proc. SPIE—Conf. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, 2002, vol. 4725, pp. 18–29.

- [19] C. Chang, Q. Du, T. Sun, and M. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 7, no. 6, pp. 2631–2641, Nov. 1999.
- [20] L. Sun and W. Gao, "Method of selecting the best classification bands from hyperspectral images based on genetic algorithm and rough set," in *Proc. SPIE—Conf. Hyperspectral Remote Sensing and Application*, R. O. Green and Q. Tong, Eds., 1998, vol. 3502, pp. 179–184.
- [21] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [22] L. L. Scharf, *Statistical Signal Processing, Detection Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [23] A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
- [24] J. B. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 3, pp. 295–304, May 1990.
- [25] J. Bruske and G. Sommer, "Intrinsic dimensionality estimation with optimally topology preserving maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 5, pp. 572–575, May 1998.
- [26] P. Demartines and J. Hérault, "Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 148–154, Jan. 1997.
- [27] M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy, "Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images," in *Proc. SPIE—Symp. Remote Sensing Conf. Image and Signal Processing for Remote Sensing VII*, 2001, vol. 4541, pp. 157–169.
- [28] C. Bachmann, T. Ainsworth, and R. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2786–2803, Oct. 2006.
- [29] C. Bachmann, T. Ainsworth, and R. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.
- [30] C. Yangchi, M. Crawford, and J. Ghosh, "Applying nonlinear manifold learning to hyperspectral data for land cover classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2005, vol. 6, pp. 4311–4314.
- [31] D. Gillis, J. Bowles, G. M. Lamela, W. J. Rhea, C. M. Bachmann, M. Montes, and T. Ainsworth, "Manifold learning techniques for the analysis of hyperspectral ocean data," in *Proc. SPIE—Conf. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, S. S. Shen and P. E. Lewis, Eds., 2005, vol. 5806, pp. 342–351.
- [32] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
- [33] M. Lennon, M. Mouchot, G. Mercier, and L. Hubert-Moy, "Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2001, pp. 2893–2895.
- [34] A. Ifarraguerri and C.-I. Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 127–143, Nov. 2000.
- [35] C. Bachmann and T. Donato, "An information theoretic comparison of projection pursuit and principal component features for classification of Landsat TM imagery of central Colorado," *Int. J. Remote Sens.*, vol. 21, no. 15, pp. 2927–2935, Oct. 2000.
- [36] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.
- [37] S. Kaewpajit, J. L. Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 863–871, Apr. 2003.
- [38] J. H. Bowles, J. A. Antoniadis, M. M. Baumbach, J. M. Grossmann, D. Haas, P. J. Palmadesso, and J. Stracka, "Real-time analysis of hyperspectral data sets using NRL's ORASIS algorithm," in *Proc. SPIE—Conf. Imaging Spectrometry III*, 1997, vol. 3118, pp. 38–45.
- [39] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 55–78, 2003.
- [40] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.
- [41] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.
- [42] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.
- [43] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
- [44] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.
- [45] J. Harsanyi, W. Farrand, and C.-I. Chang, "Determining the number and identity of spectral endmembers: An integrated approach using Neyman-Pearson eigenthresholding and iterative constrained rms error minimization," in *Proc. 9th Thematic Conf. Geologic Remote Sens.*, 1993.
- [46] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.
- [47] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.
- [48] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. Geological Survey digital spectral library: Version 1: 0.2 to 3.0  $\mu\text{m}$ ," *U.S. Geol. Survey*, 1993. Open File Report 93-592.
- [49] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. London, U.K.: Chapman & Hall, 2004.
- [50] T. Minka, "Estimating a Dirichlet distribution," MIT, Cambridge, MA, 2000. Tech. Rep.
- [51] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed., ser. Mathematical Sciences. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [52] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*. New York: Academic, 1990.
- [53] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ: Wiley, 2001.
- [54] R. Roger and J. Arnold, "Reliably estimating the noise in AVIRIS hyperspectral imagers," *Int. J. Remote Sens.*, vol. 17, no. 10, pp. 1951–1962, 1996.
- [55] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 44, no. 2/3, pp. 127–143, May/Jun. 1993.
- [56] P.-F. Hsieh and D. Landgrebe, "Classification of high dimensional data," Ph.D. dissertation, School Electr. Comput. Eng., Purdue Univ., West Lafayette, IN, 1998. Tech. Rep. TR-ECE 98-4.
- [57] D. Landgrebe, "Multispectral data analysis: A signal theory perspective," Purdue Univ., West Lafayette, IN, 1998. Tech. Rep.



**José M. Biucas-Dias** (S'87–M'95) received the E.E., M.Sc., Ph.D., and Agregado degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he is a Senior Researcher with the Communication Theory and Pattern Recognition Group, Institute of Telecommunications. His research interests include signal and image processing, pattern recognition, optimization, and synthetic aperture radar. He is a Researcher of several national and international research projects and networks including the Marie Curie Actions "Hyperspectral Imaging Network" and the "European Doctoral Program in Signal Processing."

Dr. Biucas-Dias was an Associate Editor of the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS II. He has been a member of program committees of several international conferences, including the IEEE Conference on Computer Vision and Pattern Recognition, IEEE International Geoscience and Remote Sensing Symposium, and the International Symposium on Visual Computing.



**José M. P. Nascimento** (S'03–M'06) received the B.S. degree in electronics and telecommunications engineering and the E.E. degree from the Instituto Superior de Engenharia de Lisboa, Polytechnical Institute of Lisbon, Lisbon, Portugal, in 1993 and 1995, respectively, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico, Technical University of Lisbon, Lisbon, in 2000 and 2006, respectively.

He is currently a Professor with the Department of Electronics, Telecommunications, and Computer Engineering, Instituto Superior de Engenharia de Lisboa, Polytechnical Institute of Lisbon. He is also a Researcher with the Institute of Telecommunications, Instituto Superior Técnico, Technical University of Lisbon. His research interests include remote sensing and signal and image processing.