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 squared error based approach to infer the signal subspace in hyperspectral imagery. The method, termed *hyperspectral signal identification by minimum error* (HySime), is eigendecomposition based and adaptive (*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 proposed method is illustrated using simulated and real hyperspectral images.

Index Terms

Dimensionality Reduction; Subspace Identification; Hyperspectral Imagery; Linear Mixture; Hyperspectral Unmixing; Minimum Mean Squared Error.

I. INTRODUCTION

HYPERSPECTRAL sensors sample the reflected solar radiation from the Earth surface in the portion of the spectrum extending from the visible region through the near-infrared and mid-infrared (wavelengths between $0.3\mu m$ and $2.5\mu m$) in hundreds of narrow (on the order of $10nm$) contiguous bands [3]. This high spectral resolution yields large amounts of data. For example, AVIRIS collects a 512 (along track) $\times$ 614 (across track) $\times$ 224 (bands) $\times$ 12 (bits) data cube in 43 seconds, corresponding to more than 700 Mbits; Hyperion collects 4 Mbits in 3 seconds, corresponding to 366Kbytes/Km^2 [4]. Such huge data volumes put stringent requirements in what concerns communications, storage, and processing.

Lets us represent the spectral radiances read by a hyperspectral sensor from a given pixel as an L-dimensional 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]), the spectral vectors are a linear combination of the

This paper elaborates and extends works [1, 2].

This work was supported by the Fundação para a Ciência e Tecnologia, under the project PDCTE/CPS/49967/2003 and Instituto Politécnico de Lisboa under project IPL-5828/2004.

José M. Bioucas-Dias is with the Instituto de Telecomunicações and Instituto Superior Técnico, Techical University of Lisbon (email:bioucas@lx.it.pt)

José M. P. Nascimento is with the Instituto de Telecomunicações and Instituto Superior de Engenharia de Lisboa. (email: zen@isel.pt)

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 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, since the number of samples required to obtain accurate estimates increases drastically with the dimensionality of the data [6].

A. Hyperspectral Dimensionality Reduction Methods

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 [7]–[11]. Projection techniques seek for the best subspace to project data by minimizing an objective function. For example, principal component analysis (PCA) [12] computes the *Karhunen-Loéve* transform, which is the best data representation in the least squares sense; singular value decomposition (SVD) [13] provides the projection that best represents data in the maximum power sense; maximum noise fraction (MNF)[14] and noise adjusted principal components (NAPC)[15] seek for the projection that optimizes the ratio of noise power to signal power. NAPC is mathematically equivalent to MNF [15] and can be interpreted as a sequence of two principal component transforms: the first applies to the noise and the second applies to the transformed data set. Orthogonal subspace projection (OSP) [16] is an optimal (in the least squares sense) interference suppression process that projects the data set onto a subspace orthogonal to the undesired signatures.

Topological methods are local approaches that infer the manifold, usually of low dimension, where data set live [17]. For example, curvilinear component analysis [18], curvilinear distance analysis [19], and manifold learning [20] are non-linear projections based on the preservation of the local topology. Independent component analysis [21, 22], projection pursuit [23, 24], and wavelet decomposition [25, 26] have also been considered.

U. S. Naval Research Laboratory developed ORASIS [27] aiming at real-time implementations. This framework consists of several algorithms, where the exemplar selector module uses a non-statistical technique based on a squared error optimization to reduce the dimension of the hyperspectral data.

The identification of the signal subspace is a model order inference problem to which information theoretic criteria like the *minimum description length* (MDL) [28, 29] or the *Akaike information criterion* (AIC) [30] comes to mind. These criteria have in fact been used in hyperspectral applications [31] adopting the approach introduced by Wax and Kailath in [32].

Harsanyi, Farrand, and Chang [33] developed a Neyman-Pearson detection theory-based thresholding method (HFC) to determine the number of spectral endmembers in hyperspectral data, referred to in [31] 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, termed noise-whitened HFC (NWHFC), includes a noise-whitening step [31].

B. Proposed Approach

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. The signal 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 adaptive 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. The appendix summarizes the hyperspectral subspace estimation algorithm introduced in [1, 2].

II. Problem Formulation and Classical Dimensionality Reduction Methods

Assume that the observed spectral vectors, $\mathbf{y} \in \mathbb{R}^L$, are given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ and $\mathbf{n}$ are L-dimensional vectors standing for signal and additive noise, respectively. Furthermore, assume that signal vectors are in an unknown p-dimensional 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 [34], abundance fractions are subject nonnegativity and full additivity constraints, *i.e.*, $\mathbf{s} \in \Delta_p$, where

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

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 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, since the data is represented by vectors of smaller dimension, thus lightening the computational complexity of any posterior processing scheme.

Let us assume, for a while, that noise $\mathbf{n}$ is zero-mean Gaussian i.i.d., (*i.e.*, the components of $\mathbf{n}$ are independent and identical distributed) with variance σ_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 sample correlation matrix of $\mathbf{y}$ [13, Ch. 6]; *i.e.*, $\langle \mathbf{M} \rangle = \langle [\mathbf{e_1}, \ldots, \mathbf{e_p}] \rangle$, (the notation $\langle \mathbf{M} \rangle$ represents the subspace spanned by the columns of $\mathbf{M}$) where $\mathbf{e}_i$, for $i = 1, \ldots, p$, are the p-dominant eigenvectors of the sample correlation matrix $\widehat{\mathbf{R}}_y$.

A. Eigenanalysis of $\widehat{\mathbf{R}}_y$

The ML estimator of the subspace signal 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, 35]. It consists in assuming that the noise is zero-mean i.i.d.. Thus, the correlation matrix of the observed vectors may be written as $\mathbf{R}_y = \mathbf{E}(\mathbf{\Sigma} + \sigma_n^2 \mathbf{I}_L)\mathbf{E}^T$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix, $\mathbf{E}$ and $\mathbf{\Sigma}$ 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 $\mathbf{\Sigma}$ by decreasing magnitude, we have then $\langle \mathbf{M} \rangle = \langle [\mathbf{e}_1 \ldots \mathbf{e}_p] \rangle$; *i.e.*, the estimate of the signal subspace is the span of the eigenvectors of $\mathbf{R}_y$ whose respective eigenvalues values are larger than σ_n^2 [13].

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

i) 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.

ii) Even if the noise was i.i.d., the procedure described above to infer the subspace dimension would be prone to errors owing to random perturbations always present in the estimates of σ_n^2, $\mathbf{E}$, and $\mathbf{\Sigma}$.

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 USGS digital spectral library [36]. Abundance fractions are generated according to a Dirichlet distribution given by

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

where $\{s_1, \ldots, s_p\} \in \Delta_p$. The mean value of the jth endmember fraction is $\mathbb{E}[s_j] = \theta_j / \sum_{l=1}^{p} \theta_l$ ($\mathbb{E}[\cdot]$ denotes the expectation operator) [37]. The Dirichlet density, besides enforcing positivity and full additivity

constraints, displays a wide range of shapes, depending on the parameters of the distribution. On the other hand, as noted in [38], the Dirichlet density is suited to model fractions.

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

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

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

$$\text{SNR} \equiv 10 \log_{10} \frac{\mathbb{E}\left[\mathbf{x}^T\mathbf{x}\right]}{\mathbb{E}\left[\mathbf{n}^T\mathbf{n}\right]}. \tag{5}$$

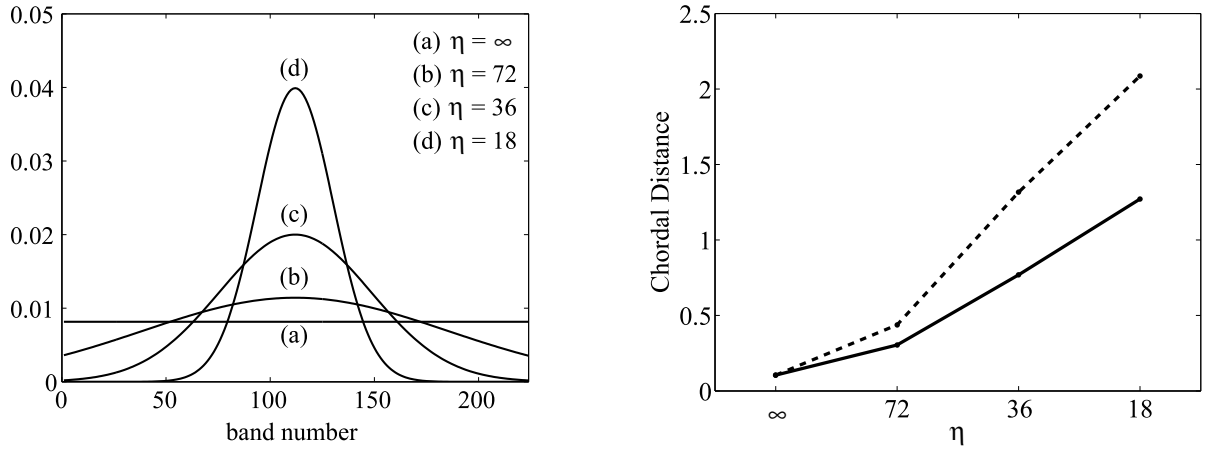
To measure the dissimilarity between the signal subspace and the subspace inferred by SVD, we adopt the *chordal distance* [39, 40], defined as

$$d = \frac{1}{\sqrt{2}} \|\mathbf{U}_p - \mathbf{U}_M\|_F, \tag{6}$$

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 by, respectively, the first p singular vectors of $\widehat{\mathbf{R}}_y$ 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 σ_i^2 for $\eta \in \{18, 38, 74, \infty\}$. Fig. 1(b) presents the chordal distance between the signal subspace and the subspace inferred by SVD, as a function of η. Notice the clear increasing of the chordal distance with η, *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 known. However, this dimension is unknown in most real applications and must be inferred from data as already referred to. This is a model-order inference problem that, if based only on the eigenvalues of the data correlation matrix, may lead to poor results. This aspect is illustrated in Fig. 2, where we have plotted the eigenvalues (circles) of $\widehat{\mathbf{R}}_y$ computed in the experiment above with SNR $= 12$ dB and $\eta = 18$. Note how difficult is to infer the number of signal eigenvalues for this SNR, because they are masked by the noise eigenvalues. We will see that HySime is able to infer the correct subspace dimension in the present scenario.



(a) (b)

Fig. 1. (a) Noise variance in each band given (4), for different values of η; (b) *Chordal distance* as a function of parameter η for SNR$=17$ dB. (line) subspace inferred by SVD; (dashed line) subspace inferred by MNF.

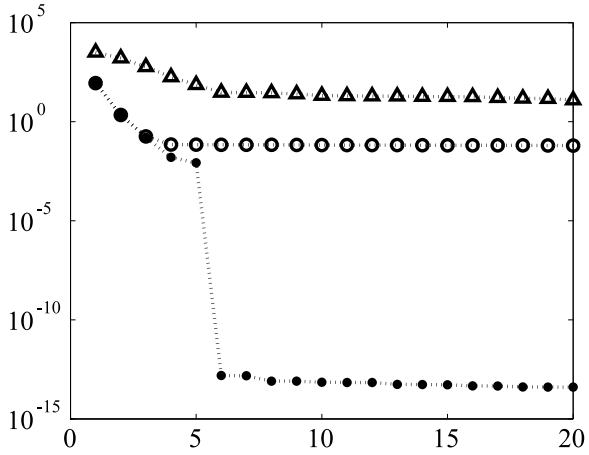


Fig. 2. (dots) Eigenvalues of $\widehat{\mathbf{R}}_{\mathbf{x}}$; (circles) Eigenvalues of $\widehat{\mathbf{R}}_y$; (triangles) decreasing ratio of $(\mathbf{v}_i^T\mathbf{R}_y\mathbf{v}_i)/(\mathbf{v}_i^T\mathbf{R}_n\mathbf{v}_i)$ [$\mathbf{v}_i$, are the left-hand eigenvectors o $\mathbf{R}_n\mathbf{R}_y^{-1}$]. Simulation parameters: SNR $= 12$ dB, $\eta = 18$, and $p = 5$.

B. Maximum Noise Fraction

Maximum noise fraction (MNF) is a another popular subspace inference tool in remote sensing that takes into account the noise statistics. Nonetheless, it has limitations similar to SVD-based approaches, as we illustrate below.

MNF finds 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 minimization consists in finding orthogonal directions minimizing the ratio

$$\frac{\mathbf{v}_i^T\mathbf{R}_n\mathbf{v}_i}{\mathbf{v}_i^T\mathbf{R}_y\mathbf{v}_i}, \tag{7}$$

with respect to $\mathbf{v}_i$. This problem is known as the generalized Rayleigh quotient and the solution is given by the left-hand eigenvectors $\mathbf{v}_i$, for $i = 1, \ldots, L$, of $\mathbf{R}_n\mathbf{R}_y^{-1}$ [41].

For i.i.d noise, we have $\mathbf{R}_n = \sigma_n^{-2}\mathbf{I_L}$ and $\mathbf{R}_y^{-1} = \mathbf{E}\left(\mathbf{\Sigma} + \sigma_\mathbf{n}^\mathbf{2}\mathbf{I_L}\right)^{-\mathbf{1}}\mathbf{E^T}$, and therefore 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 identify correctly 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 parameter η and assuming $p = 5$. The chordal distance exhibits a pattern similar to that of the SVD based approach being, however, larger for $\eta \neq \infty$ (white noise case). Fig. 2 (triangles) plots the ratio $(\mathbf{v}_i^T\mathbf{R}_y\mathbf{v}_i)/(\mathbf{v}_i^T\mathbf{R}_n\mathbf{v}_i)$ along direction $\mathbf{v}_i$ by decreasing order. As in the SVD-based approach, we face a model-order selection problem that is hard to solve without any further information or criterium.

III. Signal Subspace Estimation

This section introduces formally 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 squared error sense. The application of this criterium 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 particulary in remote sensing. Arguably, in hyperspectral imagery, the simplest noise estimation procedure is the *shift difference* method, also denominated as *nearest neighbor difference* [14]. 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 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, a supervised selection of homogeneous areas must be carried out.

Herein, we follow a multiple regression theory [31, 42] based approach, which, as shown in [31, 43], outperforms the shift difference method, used, for example, in the NAPC [15] algorithm. The high correlation exhibited by close 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 $\mathbf{Z} = \mathbf{Y}^T$, the $N \times 1$ vector $\mathbf{z}_i = [\mathbf{Z}]_{:,i}$, where $[\mathbf{Z}]_{:,i}$ stands for the ith column of $\mathbf{Z}$ (*i.e.*, $\mathbf{z}_i$ contains the data read by the hyperspectral sensor at the ith band for all image pixels), and the $N \times (L-1)$ matrix $\mathbf{Z}_{\partial_i} = [\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \ldots, \mathbf{z}_L]$.

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

$$\mathbf{z}_i = \mathbf{Z}_{\partial_i}\boldsymbol{\beta}_i + \boldsymbol{\xi}_i, \tag{8}$$

where $\mathbf{Z}_{\partial_i}$ is the explanatory data matrix, $\boldsymbol{\beta}_i$ is the regression vector of size $(L-1)\times 1$, and $\boldsymbol{\xi}_i$ is the modeling error vector of size $N\times 1$. For each $i\in\{1,\ldots,L\}$, the least squares estimator of the regression vector $\boldsymbol{\beta}_i$ is given by

$$\widehat{\boldsymbol{\beta}}_i = \left(\mathbf{Z}_{\partial_i}^T\mathbf{Z}_{\partial_i}\right)^{-1}\mathbf{Z}_{\partial_i}^T\mathbf{z}_i. \tag{9}$$

The noise is estimated by

$$\widehat{\boldsymbol{\xi}}_i = \mathbf{z}_i - \mathbf{Z}_{\partial_i}\widehat{\boldsymbol{\beta}}_i, \tag{10}$$

and the correlation matrix by $\widehat{\mathbf{R}}_n = \left[\widehat{\boldsymbol{\xi}}_1,\ldots,\widehat{\boldsymbol{\xi}}_N\right]^T\left[\widehat{\boldsymbol{\xi}}_1,\ldots,\widehat{\boldsymbol{\xi}}_N\right]/N$. Notice that the determination of each noise vector $\widehat{\boldsymbol{\xi}}_i$ implies the computation of the pseudo-inverse $\mathbf{Z}_{\partial_i}^{\#} = \left(\mathbf{Z}_{\partial_i}^T\mathbf{Z}_{\partial_i}\right)^{-1}\mathbf{Z}_{\partial_i}^T$, of size $(L-1)\times(L-1)$, for each $i=1,\ldots,L$. This huge 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 definite matrices $\mathbf{R}$ and $\mathbf{R}^{-1}$ be partitioned into block matrices as follows

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \\ \hline \mathbf{b}^T & c \end{array}\right], \quad \mathbf{R}^{-1} = \left[\begin{array}{c|c} \mathbf{A}' & \mathbf{b}' \\ \hline \mathbf{b}'^T & c' \end{array}\right], \tag{11}$$

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. Since $\mathbf{R}$, $\mathbf{R}^{-1}$, $\mathbf{A}$, $\mathbf{A}^{-1}$, and c' are positive definite, thus

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

$$\mathbf{A}\mathbf{b}' + \mathbf{b}c' = \mathbf{0}_{L-1}. \tag{13}$$

Replacing $\mathbf{A}^{-1}\mathbf{b}' = -\mathbf{b}'/c'$, derived from (13), into expression (12), we obtain

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

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

The pseudo-code for the noise estimation is shown in the Algorithm 1. Symbol $[\widehat{\mathbf{R}}]_{\partial_i,\partial_i}$ denotes the matrix obtained from $\widehat{\mathbf{R}}$ by deleting the ith row and the ith column, $[\widehat{\mathbf{R}}]_{i,\partial_i}$ denotes the ith row of $[\widehat{\mathbf{R}}]_{:,\partial_i}$, and $[\widehat{\mathbf{R}}]_{\partial_i,i}$ denotes $[\widehat{\mathbf{R}}]_{i,\partial_i}^T$. Steps 2 and 3 compute matrix $\widehat{\mathbf{R}} = \mathbf{Z}^T\mathbf{Z}$ and its inverse, respectively. Steps 5 and 6 estimate, respectively, the regression vector $\widehat{\boldsymbol{\beta}}_i$ and the noise $\widehat{\boldsymbol{\xi}}_i$, for each $i=1,\ldots,L$.

The main advantage of Algorithm 1 is that the computation of $\widehat{\mathbf{R}}$ and of $\mathbf{R}' = \widehat{\mathbf{R}}^{-1}$ are out of the loop *for*. Thus, the computational complexity, i.e., the number of floating point operations (flops), of Algorithm

Algorithm 1 : Noise estimation

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

TABLE I

COMPUTATIONAL COMPLEXITY OF THE NOISE ESTIMATION ALGORITHMS.

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

1 is substantially lower than that of an algorithm implementing the multiple regression without using the relation (14). Note that the computation of the sample correlation matrix and of its inversion demands, approximately, $2NL^2 + L^3$ flops, whereas the multiple regression algorithm without using the relation (14) has to compute L times the above matrices, thus demanding, approximately, $2NL^3 + L^4$ flops. Table I presents 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, what is a significant figure, since $L/2$ takes, in many applications, values of the order of 100.

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 parameter σ^2 of expression (4) is defined by the desired SNR.

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

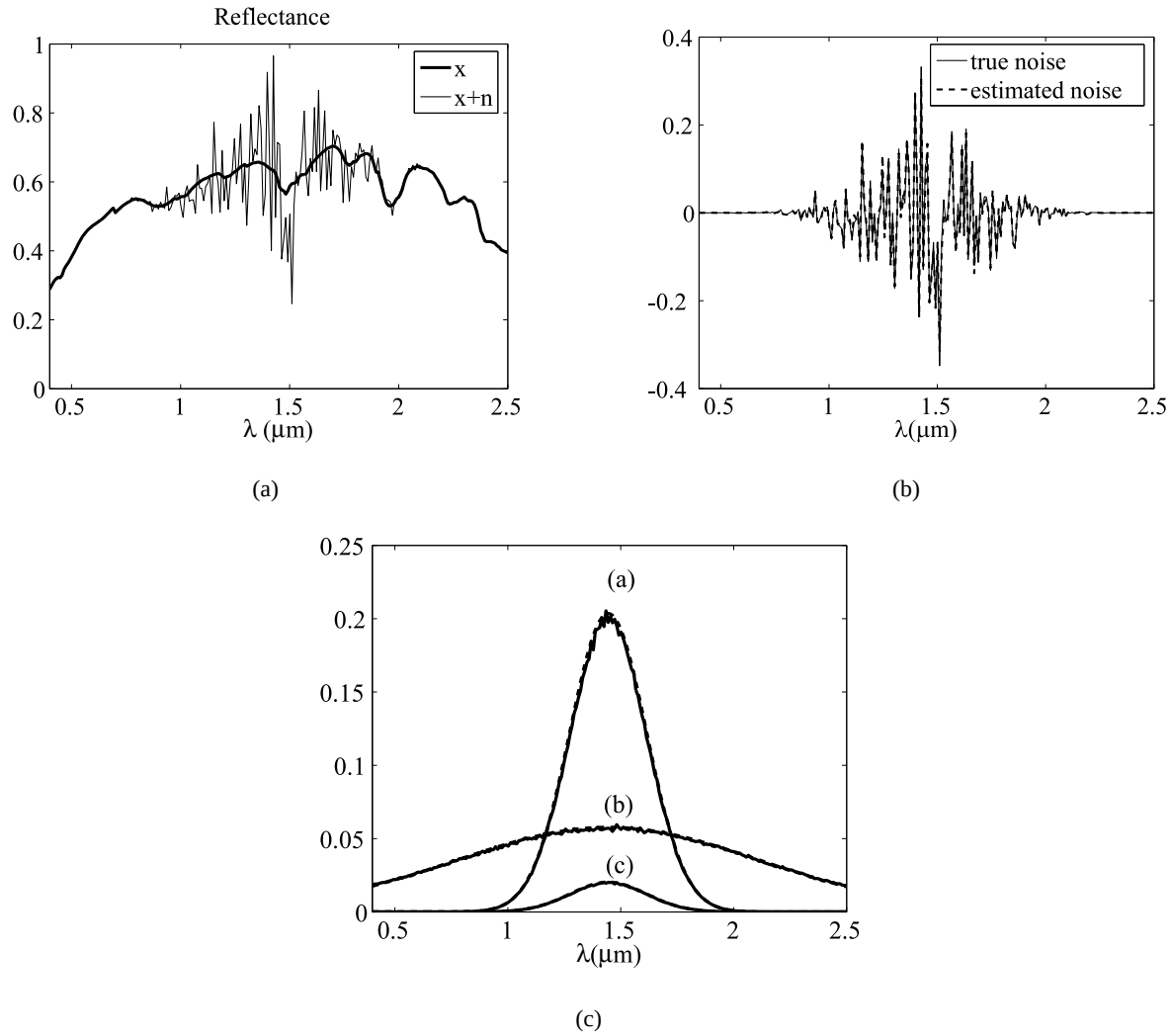


Fig. 3. Illustration of the noise estimation algorithm; (a) thick line: reflectance signal of a pixel; thin line: reflectance of the same pixel with noise; (b) solid line: true noise; dashed line: estimated noise. (c) solid line: diagonal of the estimated noise covariance matrix; dashed line: diagonal 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 $\eta = 18$;

B. Signal Subspace Estimation

This section presents the core structure of the proposed method. The first step, 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 mean squared error 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 $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \widehat{\mathbf{R}}_n)$, *i.e.,* the noise is zero-mean Gaussian distributed with covariance matrix $\widehat{\mathbf{R}}_n$.

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

$$\widehat{\mathbf{R}}_x = \mathbf{E}\mathbf{\Sigma}\mathbf{E}^T, \tag{15}$$

where $\mathbf{E} \equiv [\mathbf{e}_1, \ldots, \mathbf{e}_L]$ is a matrix with the eigenvectors of $\widehat{\mathbf{R}}_x$. Given a permutation $\pi = \{i_1, \ldots, i_L\}$ of indices $i = 1, \ldots, L$, let us decompose the space $\mathbb{R}^L$ into two orthogonal subspaces: the k-dimensional subspace $\langle \mathbf{E}_k \rangle$ spanned by $\mathbf{E}_k \equiv [\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}]$ and $\langle \mathbf{E}_k \rangle^\perp$ spanned by $\mathbf{E}_k^\perp \equiv [\mathbf{e}_{i_{k+1}}, \ldots, \mathbf{e}_{i_L}]$, i.e., the orthogonal complement of 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 $\widehat{\mathbf{x}}_k \equiv \mathbf{U}_k \mathbf{y}$ be the projection of the observed spectral vector $\mathbf{y}$ onto the subspace $\langle \mathbf{E}_k \rangle$. The first and the second-order moments of $\widehat{\mathbf{x}}_k$ given $\mathbf{x}$ are

$$\begin{aligned} \mathbb{E}\big[\widehat{\mathbf{x}}_k|\mathbf{x}\big] &= \mathbf{U}_k \mathbb{E}\big[\mathbf{y}|\mathbf{x}\big] \\ &= \mathbf{U}_k \mathbb{E}\big[\mathbf{x} + \mathbf{n}|\mathbf{x}\big] \\ &= \mathbf{U}_k \mathbf{x} \\ &\equiv \mathbf{x}_k, \end{aligned} \tag{16}$$

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

The mean squared error between $\mathbf{x}$ and $\widehat{\mathbf{x}}_k$ is

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

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

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

where c is an irrelevant constant. The criterium we propose to estimate the signal subspace, let us call it X, is the minimization of $\text{mse}(k)$ given by (19), with respect to all the permutations $\pi = \{i_1, \ldots, i_L\}$ of size L and to k, with the correlation matrix $\mathbf{R}_y$ replaced with the sample correlation matrix $\widehat{\mathbf{R}}_y = \mathbf{Y}\mathbf{Y}^T/N$; *i.e.*,

$$\widehat{X} = \langle [\mathbf{e}_{\widehat{i}_1}, \ldots, \mathbf{e}_{\widehat{i}_{\widehat{k}}}] \rangle \tag{20}$$

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

where the dependence on the permutation π is through $\mathbf{U}_k = \mathbf{E}_k\mathbf{E}_k^T$, since $\mathbf{E}_k \equiv [\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_k}]$. For a given permutation π, each term of expression (21) has a clear meaning: 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 the $\mathbf{U}_k$ is a projection matrix and that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$, provided that matrices $\mathbf{A}$ and $\mathbf{B}$ are compatible with respect to the multiplications $\mathbf{AB}$ and $\mathbf{BA}$, the minimization (21) can be rewritten as

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

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

$$p_{i_j} \quad = \quad \mathbf{e}_{i_j}^T \widehat{\mathbf{R}}_y \mathbf{e}_{i_j} \tag{23}$$

$$\sigma_{i_j}^2 \quad = \quad \mathbf{e}_{i_j}^T \widehat{\mathbf{R}}_n \mathbf{e}_{i_j}. \tag{24}$$

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

The pseudo-code for HySime is shown in the Algorithm 2. HySime inputs are the spectral observed vectors and the sample correlation matrix $\widehat{\mathbf{R}}_y$. Step 2 estimates the noise correlation matrix $\widehat{\mathbf{R}}_n$. Step 3 estimates the signal correlation matrix $\widehat{\mathbf{R}}_x$. Steps 4 and 5 calculate the eigenvectors of the signal correlation matrix and the terms δ_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 $\widehat{k}$ and $\widehat{\pi}$.

Algorithm 2 : HySime

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

An alternative to the projection $\mathbf{U}_k^{\perp}\widehat{\mathbf{R}}_y$, consists in projecting the sample mean $\overline{\mathbf{y}} \equiv \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i$ onto the subspace orthogonal to the signal subspace. This alternative method, called HySime$_m$ has been developed in [1, 2]. The underlying rational is that spectral vectors are nonnegative and then $\overline{\mathbf{y}}$ accumulates information about every spectral vectors in the data set. Of course, for this approach to work, the projection of $\overline{\mathbf{y}}$ onto any eigenvector vector $\mathbf{e}_j$, $j = 1, \ldots, k$ must be nonzero. Although we have not any proof of

this statement, we believe, supported on practical evidence, that the probability of $\overline{\mathbf{y}}^T\mathbf{e}_j = 0$ is practically zero in real data sets. Another difference between HySyme and HySime$_m$ is that the latter, as proposed in [1, 2], does not search the signal subspace over all the permutations π of the eigenvectors; instead it searches only over the sequence of nested subspaces $\mathbf{S}_k = [\mathbf{e}_1, \ldots, \mathbf{e}_k]$, where the eigenvectors are ordered by decreasing magnitude of the corresponding eigenvalues. We give evidence, in the next section, that HySime slightly outperforms HySime$_m$. Given the structural similarities between HySime$_m$ and HySime, we present in appendix a brief description of HySime$_m$ and its pseudocode.

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 [31]. As concluded in [31], these algorithms are the state-of-the-art in hyperspectral signal subspace identification, outperforming the information theoretical criteria approaches; namely, the *minimum description length* (MDL) [28, 29] and the *Akaike information criterion* [30]. HySime and HySime$_m$ algorithms are also compared, given their structural similarity.

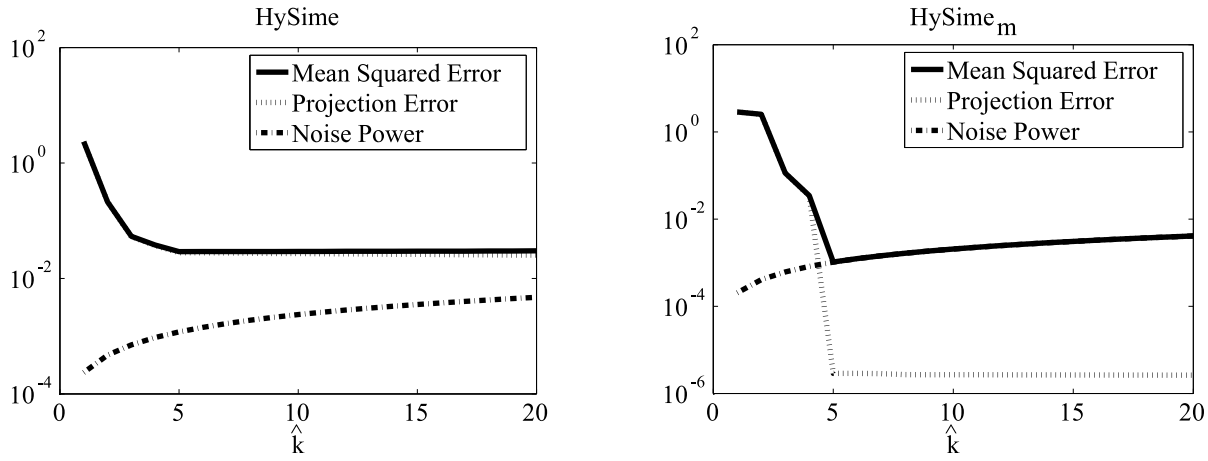
The spectral signatures are selected from the USGS digital spectral library [36]. The abundance fractions are generated according to a Dirichlet distribution defined in expression (3). The results presented here are organized into two experiments: in the first experiment, the method is evaluated with respect to the SNR [see definition (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. 4 shows the evolution of the mean squared error for HySime and HYSime$_m$ algorithms as a function of the parameter k, for SNR$= 35$ dB and $p = 5$. The minimum of the mean squared error occurs at $k = 5$, which is exactly the number of endmembers present in the image. As expected, the projection error power and of noise power display decreasing and increasing behaviors, respectively, as function of the subspace dimension k.

Table II presents the signal subspace order estimates yielded by HySime, HySime$_m$ algorithms, and the virtual dimensionality (VD) determined by the NWHFC and the HFC algorithms [31], as a function of the SNR, of the number of endmembers, p, and of the noise shape.

NWHFC algorithm is basically the HFC one [33] preceded by a noise-whitening step, 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 is low. For this reason, we have used both the true and estimated noise 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$_m$ methods



(a) (b)

Fig. 4. Mean squared error versus k, with $SNR = 35$ dB, $p = 5$; (a) HySime method; (b) HySime$_m$ method.

yield the same results, whether using the estimated or the true noise correlation matrices. Table II does not present results of the HFC method for noise colored scenarios because this method was designed only for the white noise scenarios.

Another central issue of NWHFC and HFC algorithms is the false-alarm probability P_f it is 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 [31] and in our own results, we choose $P_f \in \{10^{-3}, 10^{-4}, 10^{-5}\}$.

The figures shown in Table II, based on 50 Monte Carlo runs, have the following behavior:

a) HySime and HySime$_m$ algorithms display similar performance, with a small advantage for former, namely at small SNRs and colored noise;

b) HySime and NWHFC algorithms parameterized with $P_f = 10^{-3}$ display similar performances at low subspace dimension, say $p \leq 5$, and white noise. This is also true for colored noise and NWHFC working with known noise covariance matrix. However, if the noise statistics is unknown, NWHFC performs much worse than HySime;

c) HySime performs better that NWHFC for high space dimensions, say $p > 5$.

We conclude, therefore, that the HySime algorithm is slightly better than the HySime$_m$ one, yielding systematically equal or better results than NWHFC and HFC algorithms. Another advantage of HySime approach is its adaptiveness, *i.e.*, it does not depend on any tunable parameter.

A. Experiment 2

In this experiment, we set SNR $= 35$ dB and $p = 8$. The first five endmembers are mixed according to a Dirichlet distribution, as in the previous experiment, the sixth, the seventh, and the eighth endmembers are

TABLE II

SIGNAL SUBSPACE DIMENSION $\widehat{k}$, BASED ON 50 MONTE CARLO RUNS, AS FUNCTION OF SNR, p, AND η (NOISE SHAPE). FIGURES IN BRACKETS WERE COMPUTED BASED ON THE TRUE NOISE STATISTICS.

		White Noise ($\eta = 0$)				Gaussian shaped noise ($\eta = 1/18$)			
SNR	Method	$p = 3$	$p = 5$	$p = 10$	$p = 15$	$p = 3$	$p = 5$	$p = 10$	$p = 15$
50 dB	HySime	**3**	**5**	**10**	**15**	**3**	**5**	**10**	**15**
	HySime$_m$	3	5	10	15	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)	59 (3)	41 (5)	61 (10)	45 (10)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	5 (5)	7 (7)	8 (8)	48 (3)	33 (5)	54 (10)	34 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	7 (6)	8 (8)	43 (3)	28 (5)	41 (9)	27 (10)
35 dB	HySime	**3**	**5**	**10**	**15**	**3**	**5**	**10**	**15**
	HySime$_m$	3	5	10	15	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)	9 (3)	10 (5)	12 (10)	10 (10)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	4 (4)	7 (6)	8 (8)	9 (3)	9 (5)	11 (10)	8 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	6 (6)	8 (8)	7 (3)	7 (5)	10 (9)	8 (10)
25 dB	HySime	**3**	**5**	**10**	**14**	**3**	**5**	**10**	**15**
	HySime$_m$	3	5	9	12	3	5	10	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)	4 (3)	5 (5)	11 (10)	9 (11)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	5 (5)	6 (6)	7 (7)	4 (3)	5 (5)	11 (10)	9 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	5 (5)	7 (7)	4 (3)	5 (5)	11 (9)	8 (10)
15 dB	HySime	**3**	**5**	**8**	**12**	**3**	**5**	**8**	**12**
	HySime$_m$	3	3	6	8	3	3	5	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)	4 (3)	5 (5)	11 (10)	10 (10)
	NWHFC ($P_f = 10^{-4}$)	3 (3)	4 (4)	3 (3)	3 (2)	4 (3)	5 (5)	11 (10)	8 (10)
	NWHFC ($P_f = 10^{-5}$)	3 (3)	4 (4)	3 (3)	2 (2)	4 (3)	5 (5)	11 (9)	8 (10)

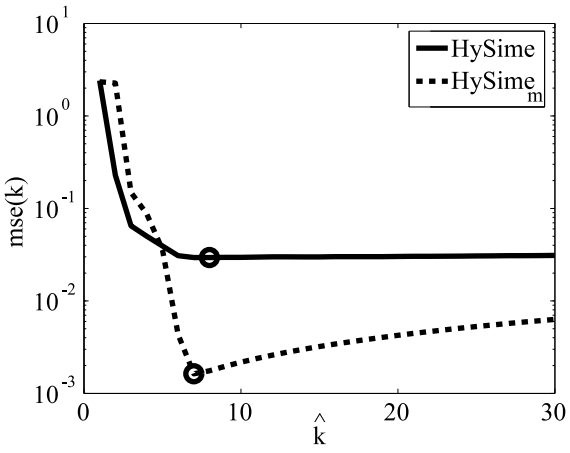


Fig. 5. Mean squared error versus k, with SNR $= 35$ dB, $p = 8$ (second experiment). Solid line: HySime method; dashed line: HySime$_m$ method; circles: estimated number of endmembers.

TABLE III

SIGNAL SUBSPACE DIMENSION $\widehat{k}$ AS FUNCTION PARAMETER η EXPERIMENT 2 (FEW PURE PIXELS) WITH SNR $= 35$ DB.

Method $(p = 8)$	$\eta = 0$	$\eta = 1/18$
HySime	8	7
HySime$_m$	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

present as pure in 8, 4, and 2 pixels, respectively. Fig. 5 shows the mean squared error versus the subspace dimension k for the HySime and HySime$_m$ methods. The respective curves achieve their minima at $k = 8$ and $k = 7$. Thus, the HySime algorithm infers the correct subspace dimension, whereas HySime$_m$ one underestimates it in one unit.

Table III displays the results of this experiment computed by HySime, HySime$_m$, HFC and NWHFC algorithms. We observe the same pattern of behavior shown in Table II, with HySime method yielding the best performance.

V. EXPERIMENTS WITH REAL HYPERSPECTRAL DATA

In this section, the proposed method is applied to real hyperspectral data collected by the AVIRIS [44] sensor over Cuprite, Nevada[1]. Cuprite is a mining area in southern Nevada with mineral and little vegetation [45]. The Cuprite test site, located approximately 200 Km northwest of Las Vegas is a relatively

[1]Available at http://aviris.jpl.nasa.gov/html/aviris.freedata.html

TABLE IV

MINERALS DETECTED BY AVIRIS [50].

Mineral name
K-Alunite
Na-Alunite
Buddingtonite
Calcite
Fe-Chlorite
Dickite
Goethite
Halloysite
Hematite
NH4 Illite/Smectite
Jarosite
Kaolinite (wxl)
Kaolinite (pxl)
Na-Montmorillonite
Ca-Montmorillonite
Nontronite
Chalcedony
K-Muscovite

undisturbed acid-sulfate hidrothermal system near highway 95. The geology and alteration were previously mapped in detail [46, 47]. A geologic summary and a mineral map can be found in [45] which is resumed in Table IV. This site has been extensively used for remote sensing experiments over the past years [48, 49] and it has become a standard test site for comparison of unmixing and endmember extraction algorithms. This study is based on a subimage (350×350 pixels and 224 bands) of a data set acquired on the AVIRIS flight of June 19, 1997 [see Fig. 6]. AVIRIS instrument covers the spectral region from $0.41\mu m$ to $2.45\mu m$ in 224 bands with a $10nm$ band width. Flying at an altitude of $20km$, it has an IFOV of $20m$ and views a swath over $10km$ wide.

Table V shows the subspace dimension inferred by HySime, Hysime$_m$, HFC, and NWHFC. The values $\{10^{-5}, 10^{-4}, 10^{-3}\}$ were considered in the false alam probabilities, P_f, parameterizing the HFC and the NWHFC methods. Fig. 7 plots the evolution of the mean squared error (over all subspaces spanned by k eigenvectors of the signal correlation matrix estimate $\widehat{\mathbf{R}}_x$) as a function of the subspace dimension $\widehat{k}$. The minimum is obtained for $k = 20$. HySime$_m$, NWHFC ($P_f = 10^{-3}$) and HFC ($P_f = 10^{-3}$) yield $\widehat{k} = 25, 23, 22$, respectively.

According to the ground truth presented in [45], HySime, Hysime$_m$, NWHFC and HFC methods overestimate the number of endmembers in the Cuprite data set. This difference is due to a) the presence of rare pixels not accounted for in [45] and b) spectral variability. The bulk of spectral energy is explained by a small number of eigenvectors. This can be observed from Fig. 8(a), where the accumulated signal energy is

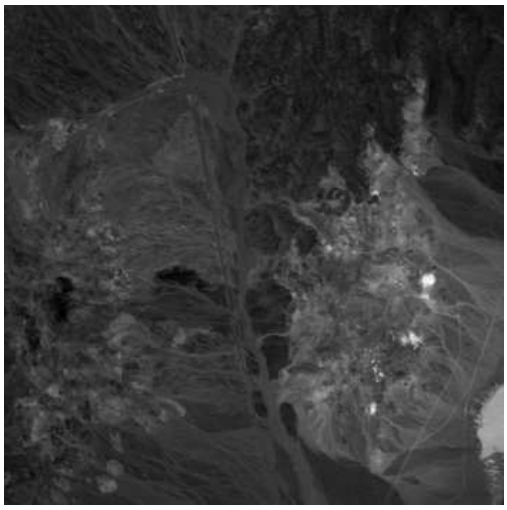


Fig. 6. Band 30 (wavelength $\lambda = 655.8nm$) of the subimage of AVIRIS cuprite Nevada data set;

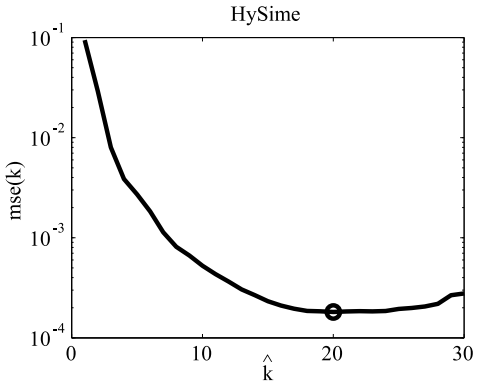


Fig. 7. Mean squared error versus k for Cuprite data set. Solid line: HySime method; circle: estimated number of endmembers.

plotted as function of the signal eigenvalue index ordered by decreasing magnitude. The energy contained in the first 8 eigenvalues is 99.95% of the total. This fact is further confirmed in Fig. 8(b), which shows, in gray level and for each pixel, the percentage of energy contained in the subspace $\langle \mathbf{E}_{9:19} \rangle = \langle [\mathbf{e}_9, \ldots, \mathbf{e}_{19}] \rangle$. Note that only a few (rare) pixels contain energy in this subspace. Furthermore, these energies are a very small percentage of the corresponding spectral vector energies (less than 0.3%).

An objective comparison of the relative performance of the different algorithms is not straightforward due to the following:

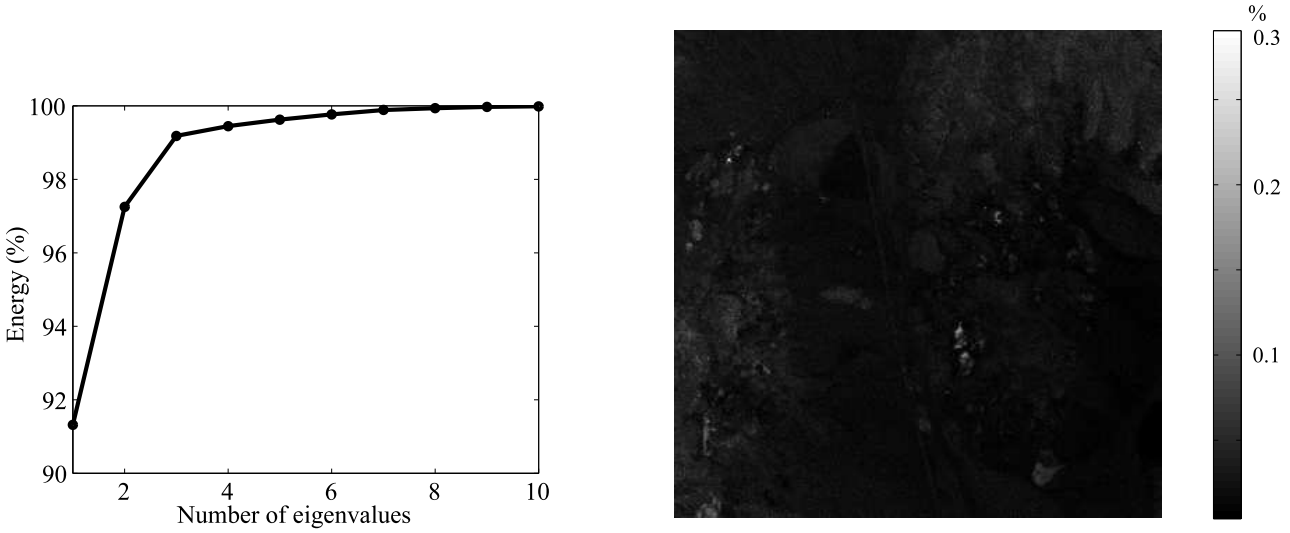
1) The rare pixels present in the scene are not all accounted for in the ground truth
2) HFC and NWHFC depend on a tunable parameter, what is not the case with the HySime method.
3) The noise present in real hyperspectral data sets is colored, scenario for which HFC was not conceived.

A deeper assessment and comparison of the methods would demand field measurements beyond the

TABLE V

$\widehat{k}$ FOR THE CUPRITE DATA SET.

Method	$\widehat{k}$
HySime	20
HySime$_m$	25
HFC ($P_f = 10^{-3}$)	22
HFC ($P_f = 10^{-4}$)	19
HFC ($P_f = 10^{-5}$)	19
NWHFC ($P_f = 10^{-3}$)	23
NWHFC ($P_f = 10^{-4}$)	22
NWHFC ($P_f = 10^{-5}$)	20



(a) (b)

Fig. 8. (a) Percentage of signal energy as function of the number of eigenvalues;(b) percentage of energy in the subspace $\langle \mathbf{E}_{9:15} \rangle$.

scope of this paper. Nevertheless, the paper provides clear evidence that HySime yields state-of-the-art performance, without the need of any tunable parameters.

VI. CONCLUSIONS

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 introduces the HySime algorithm, a new approach to estimate the signal subspace in hyperspectral imagery. HySime algorithm estimates the signal and the noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the minimum mean squared error sense. A set of experiments with simulated and real data leads to the conclusion that the HySime algorithm is an effective and useful tool, yielding comparable or better results than the state-of-the-art algorithms.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. C.I-Chang and Wei-min Liu for the valuable comments and suggestions about the implementation of HFC and NWHFC algorithms.

APPENDIX

A. Signal Subspace Order Estimation Using the Mean Value of Spectral Vectors HySime$_m$

Give the structural similarities similarities between HySime and HySime$_m$ algorithms and for self-containedness, we summarize in this appendix the main concepts behind the HySime$_m$ algorithm.

According to the linear observation model $\mathbf{y} = \mathbf{x} + \mathbf{n} = \mathbf{M}\mathbf{s} + \mathbf{n}$, the sample mean of the spectral vectors is

$$\begin{aligned} \overline{\mathbf{y}} &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_i \\ &= \frac{1}{N}\mathbf{M}\sum_{i=1}^{N}\mathbf{s}_i + \frac{1}{N}\sum_{i=1}^{N}\mathbf{n}_i \\ &= \mathbf{c} + \boldsymbol{\omega}, \end{aligned} \tag{25}$$

where $\mathbf{c}$ is in the signal subspace and $\boldsymbol{\omega} \sim \mathcal{N}(0, \mathbf{R}_n/N)$. Let $\mathbf{c}_k$ denote the projection of $\mathbf{c}$ onto $\langle \mathbf{E}_k \rangle$. The estimation of $\mathbf{c}_k$ can be obtained by projecting $\overline{\mathbf{y}}$ onto the signal subspace $\langle \mathbf{E}_k \rangle$, i.e., $\widehat{\mathbf{c}}_k = \mathbf{U}_k\overline{\mathbf{y}}$. The first and the second-order moments of the estimated error $\mathbf{c} - \widehat{\mathbf{c}}_k$ are respectively,

$$\begin{aligned} \mathbb{E}\big[\mathbf{c} - \widehat{\mathbf{c}}_k\big] &= \mathbf{c} - \mathbb{E}\big[\widehat{\mathbf{c}}_k\big] \\ &= \mathbf{c} - \mathbf{U}_k\mathbf{c} \\ &= \mathbf{c} - \mathbf{c}_k \\ &\equiv \mathbf{b}_k, \end{aligned} \tag{26}$$

$$\mathbb{E}\big[(\mathbf{c} - \widehat{\mathbf{c}}_k)(\mathbf{c} - \widehat{\mathbf{c}}_k)^T\big] = \mathbf{b}_k\mathbf{b}_k^T + \mathbf{U}_k\widehat{\mathbf{R}}_n\mathbf{U}_k^T/N, \tag{27}$$

where the bias $\mathbf{b}_k = \mathbf{U}_k^{\perp}\mathbf{c}$ is the projection of $\mathbf{c}$ onto the space $\langle \mathbf{E}_k \rangle^{\perp}$. Therefore the density of the estimated error $\mathbf{c} - \widehat{\mathbf{c}}_k$ is $\mathcal{N}(\mathbf{b}_k, \mathbf{b}_k^T\mathbf{b}_k + \mathbf{U}_k\widehat{\mathbf{R}}_n\mathbf{U}_k^T/N)$,

The mean squared error between $\mathbf{c}$ and $\widehat{\mathbf{c}}_k$ is given by

$$\begin{aligned} \text{mse}(k) &= \mathbb{E}\big[(\mathbf{c} - \widehat{\mathbf{c}}_k)^T(\mathbf{c} - \widehat{\mathbf{c}}_k)\big] \\ &= \text{tr}\{\mathbb{E}\big[(\mathbf{c} - \widehat{\mathbf{c}}_k)(\mathbf{c} - \widehat{\mathbf{c}}_k)^T\big]\} \\ &= \mathbf{b}_k^T\mathbf{b}_k + \text{tr}(\mathbf{U}_k\widehat{\mathbf{R}}_n\mathbf{U}_k^T/N), \end{aligned} \tag{28}$$

As in the HySime approach, the bias $\mathbf{b}_k$ is not known and an approximation of expression (28) is obtained by using the bias estimate $\widehat{\mathbf{b}}_k = \mathbf{U}_k^{\perp}\overline{\mathbf{y}}$. However, $\mathbb{E}\big[\widehat{\mathbf{b}}_k\big] = \mathbf{b}_k$ and $\mathbb{E}\big[\widehat{\mathbf{b}}_k^T\widehat{\mathbf{b}}_k\big] = \mathbf{b}_k^T\mathbf{b}_k + \text{tr}(\mathbf{U}_k^{\perp}\widehat{\mathbf{R}}_n{\mathbf{U}_k^{\perp}}^T/N)$,

that is, an unbiased estimate of $\mathbf{b}_k^T\mathbf{b}_k$ is $\widehat{\mathbf{b}}_k^T\widehat{\mathbf{b}}_k - \mathrm{tr}(\mathbf{U}_k^{\perp}\widehat{\mathbf{R}}_n\mathbf{U}_k^{\perp^T}/N)$. The criteria for the signal subspace order determination is then

$$\begin{aligned}\widehat{k} &= \arg\min_k\Big\{\widehat{\mathbf{b}}_k^T\widehat{\mathbf{b}}_k + \mathrm{tr}(\mathbf{U}_k\widehat{\mathbf{R}}_n\mathbf{U}_k^T/N) - \mathrm{tr}(\mathbf{U}_k^{\perp}\widehat{\mathbf{R}}_n\mathbf{U}_k^{\perp^T}/N)\Big\} \\ &= \arg\min_k\Big\{\overline{\mathbf{y}}^T\mathbf{U}_k^{\perp}\overline{\mathbf{y}} + 2\mathrm{tr}(\mathbf{U}_k\widehat{\mathbf{R}}_n/N)\Big\}. \end{aligned} \quad (29)$$

The pseudo-code for the HySime$_m$ (*m* stands for mean), is shown in the Algorithm 3. The core of this algorithm is similar to the Algorithm 2, excepts for Step 4 which selects k eigenvectors corresponding to the k largest eigenvalues, Step 5, which calculates the sample mean of the data set $\overline{\mathbf{y}}$, and step 6 which refers to expression (29).

HySime inputs are the spectral observed vectors and the sample correlation matrix $\widehat{\mathbf{R}}_y$. Step 2 estimates the noise correlation matrix $\widehat{\mathbf{R}}_n$. Step 3 estimates the signal correlation matrix $\widehat{\mathbf{R}}_x$. Steps 4 and 5 calculate the eigenvectors of the signal correlation matrix and the terms δ_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 $\widehat{k}$ and $\widehat{\pi}$.

Algorithm 3 HySime$_m$

1: INPUT $\mathbf{Y} \equiv [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$
2: $\widehat{\mathbf{R}}_n := \frac{1}{N}\sum_i \left(\widehat{\mathbf{n}}_i\widehat{\mathbf{n}}_i^T\right)$; { $\widehat{\mathbf{R}}_n$ is the noise correlation matrix estimates}
3: $\widehat{\mathbf{R}}_x := \frac{1}{N}\sum_i \left((\mathbf{y}_i - \widehat{\mathbf{n}}_i)(\mathbf{y}_i - \widehat{\mathbf{n}}_i^T)\right)$; { $\widehat{\mathbf{R}}_x$ is the signal correlation matrix estimates}
4: $\mathbf{U}_k := \mathbf{E}_k\mathbf{E}_k^T$; { where $\mathbf{E}_k$ are singular vectors of $\widehat{\mathbf{R}}_x$}
5: $\overline{\mathbf{y}} := \frac{1}{N}\sum_{i=1}^N \mathbf{y}_i$; { $\overline{\mathbf{y}}$ is the sample mean of the data set}
6: $\widehat{k} := \arg\min_k\Big\{\overline{\mathbf{y}}^T\mathbf{U}_k^{\perp}\overline{\mathbf{y}} + 2\mathrm{tr}(\mathbf{U}_k\mathbf{K}_n/N)\Big\}$;

REFERENCES

[1] J. M. Bioucas-Dias and J. M. P. Nascimento, "Estimation of signal subspace on hyperspectral data," in *Proc. of SPIE conference on Image and Signal Processing for Remote Sensing XI*, L. Bruzzone, Ed., vol. 5982, 2005, 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*, ser. Lecture Notes in Computer Science, J. S. Marques, N. P. de la Blanca, and P. Pina, Eds., vol. 3523, no. 2. Springer-Verlag, 2005, pp. 207–214.

[3] T. M. Lillesand, R. W. Kiefer, and J. W. Chipman, *Remote Sensing and Image Interpretation*, 5th ed. John Wiley & Sons, Inc., 2004.

[4] J. P. Kerekes and J. E. Baum, "Spectral imaging system analytical model for subpixel object detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, no. 5, pp. 1088–1101, 2002.

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

[6] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 17–28, 2002.

[7] S. S. Shen and E. M. Bassett, "Information-theory-based band selection and utility evaluation for reflective spectral systems," in *Proc. of the SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, vol. 4725, 2002, pp. 18–29.

[8] L. Sun and W. Gao, "Method of selecting the best classification bands from hyperspectral images based on genetic algorithm and rough set," in *Proc. of the SPIE conference on Hyperspectral Remote Sensing and Application*, R. O. Green and Q. Tong, Eds., vol. 3502, 1998, pp. 179–184.

[9] A. Ifarraguerri and M. W. Prairie, "Visual method for spectral band selection," *IEEE Geosci. Remote Sensing Let.*, vol. 1, no. 2, pp. 101–106, 2004.

[10] S. D. Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, "A band selection technique for spectral classification," *IEEE Geosci. Remote Sensing Let.*, vol. 2, no. 3, pp. 319–323, 2005.

[11] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sensing Let.*, vol. 2, no. 2, pp. 156–159, 2005.

[12] I. T. Jolliffe, *Principal Component Analysis*. New York: Spriger Verlag, 1986.

[13] L. L. Scharf, *Statistical Signal Processing, Detection Estimation and Time Series Analysis*. Addison-Wesley Pub. Comp., 1991.

[14] 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 Sensing*, vol. 26, no. 1, pp. 65–74, 1988.

[15] J. B. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 3, pp. 295–304, 1990.

[16] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 4, pp. 779–785, 1994.

[17] J. Bruske and G. Sommer, "Intrinsic dimensionality estimation with optimaly topologic preserving maps," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 5, pp. 572–575, 1998.

[18] P. Demartines and J. Hérault, "Curvilinear component analysis : A self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 148–154, 1997.

[19] M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy, "Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images," in *Proc. of the SPIE Symp. on Remote Sensing Conference on Image and Signal Processing for Remote Sensing VII*, vol. 4541, 2001, pp. 157–169.

[20] 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. of the SPIE conferece on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, S. S. Shen and P. E. Lewis, Eds., vol. 5806, 2005, pp. 342–351.

[21] 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. of the IEEE Int. Geosci. and Remote Sensing Symp.*, 2001.

[22] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.

[23] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 6, pp. 2653–2664, 1999.

[24] A. Ifarraguerri and C.-I. Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 6, pp. 127–143, 2000.

[25] S. Kaewpijit, J. L. Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 4, pp. 863– 871, 2003.

[26] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 2, pp. 397–408, 2006.

[27] J. H. Bowles, J. A. Antoniades, M. M. Baumback, 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. of the SPIE Conference on Imaging Spectrometry III*, vol. 3118, 1997, pp. 38–45.

[28] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[29] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[30] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.

[31] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 42, no. 3, pp. 608–619, 2004.

[32] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[33] 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 Sensing*, 1993.

[34] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using linear mixing model," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, 2001.

[35] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 12–16, 2002.

[36] 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 μm," U.S. Geological Survey, Open File Report 93-592, 1993.

[37] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. Chapman and Hall, 2004.

[38] T. Minka, "Estimating a dirichlet distribution," M.I.T., Tech. Rep., 2000.

[39] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed., ser. Mathematical Sciences. John Hopkins University Press, 1996.

[40] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*. Academic Press Inc., 1990.

[41] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, Inc., 2001.

[42] R. Roger and J. Arnold, "Reliably estimating the noise in aviris hyperspectral imagers," *Int. J. Rem. Sens.*, vol. 17, no. 10, pp. 1951–1962, 1996.

[43] C.-I. Chang and Q. Du, "Interference and noise adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 9, pp. 2387–2396, 1999.

[44] G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Rem. Sens. of the Environ.*, vol. 44, pp. 127–143, 1993.

[45] G. Swayze, R. Clark, S. Sutley, and A. Gallagher, "Ground-truthing aviris mineral mapping at cuprite, nevada,," in *Summaries of the Third Annual JPL Airborne Geosciences Workshop*, 1992, pp. 47–49.

[46] R. Ashley and M. Abrams, "Alteration mapping using multispectral images - cuprite mining district, esmeralda county,," U.S. Geological Survey, Open File Report 80-367, 1980.

[47] M. Abrams, R. Ashley, L. Rowan, A. Goetz, and A. Kahle, "Mapping of hydrothermal alteration in the cuprite mining district, nevada, using aircraft scanner images for the spectral region 0.46 to 2.36 μm," *Geology*, vol. 5, pp. 713–718, 1977.

[48] A. Goetz and V. Strivastava, "Mineralogical mapping in the cuprite mining district," in *Proc. of the Airborne Imaging Spectrometer Data Analysis Workshop, JPL Publication 85-41*, 1985, pp. 22–29.

[49] F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne and satellite hyperspectral data for geologic mapping," in *Proc. of the SPIE Aerospace Conference*, vol. 4725, 2002, pp. 128–139.

[50] G. Swayze, "The hydrothermal and structural history of the cuprite mining district, southwestern nevada: An integrated geological and geophysical approach," Ph.D. dissertation, Purdue University, University of Colorado, 1997.