A Comparative Study for Orthogonal Subspace Projection and Constrained Energy Minimization

Qian Du, *Member, IEEE*, Hsuan Ren, *Member, IEEE*, and Chein-I Chang, *Senior Member, IEEE*

***Abstract*—In this letter, we conduct a comparative study and investigate the relationship between two well-known techniques in hyperspectral image detection and classification: orthogonal subspace projection (OSP) and constrained energy minimization (CEM). It is shown that they are closely related and essentially equivalent provided that the noise is white with large SNR. Based on this relationship, the performance of OSP can be improved via data-whitening and noise-whitening processes.**

***Index Terms*—Classification, constrained energy minimization (CEM), detection, hyperspectral imagery, orthogonal subspace projection (OSP).**

I. Introduction

LINEAR UNMIXING has been widely used for hyperspectral image detection and classification [1]–[9]. It models a hyperspectral image pixel to be a linear mixture of a set of finite image endmembers that are assumed in the image data. Then, the detection and classification is performed by unmixing the pixel and finding the respective abundance fractions of these endmembers present in the pixel. Several approaches have been studied in the past, such as singular value decomposition [2], subspace projection [9], maximum likelihood (ML) [10], etc. The relationship between ML and subspace projection was investigated in [10]–[12] where ML-based linear unmixing was shown to be equivalent to the orthogonal subspace projection (OSP)-based linear unmixing, provided that the noise in the linear mixing model is white Gaussian noise. Unfortunately, such linear unmixing methods require the complete knowledge of the image endmembers. On many practical occasions, obtaining this prior knowledge may not be realistic. In order to resolve this issue, a method, referred to as Constrained Energy Minimization (CEM) was developed in [13] where the only required knowledge is the desired image endmember rather than the entire set of image endmembers. The relationship between CEM and linear unmixing was recently studied in [14]. They also are investigated as matched filter detector in [15]. Both of them have been successfully applied to hyperspectral image detection and classification because of their effectiveness and simplicity. Despite the fact that these two approaches require different levels of knowledge, it is interesting to find that they are indeed closely related, which is to be explored in this letter.

The OSP is based on the linear mixture model, which says a hyperspectral pixel vector $\mathbf{r}$ of size $K\times 1$ with K spectral bands can be represented as

$$\mathbf{r}=\mathbf{S}\boldsymbol{\alpha}+\mathbf{n} \tag{1}$$

where $\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_p]$ is a $K\times p$ signature matrix with p endmembers, and $\mathbf{s}_i$ is the ith endmember signature; $\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\ldots,\alpha_p]^T$ is a $p\times 1$ abundance fraction vector where the ith element α_i represents the abundance fraction of $\mathbf{s}_i$ present in that pixel; $\mathbf{n}$ is a $K\times 1$ vector that can be interpreted as noise term or model error. For the OSP, the signature matrix $\mathbf{S}$ in (1) is further divided into two parts, desired signature of interest $\mathbf{d}$ and undesired signature matrix $\mathbf{U}$. Without loss of generality, we assume $\mathbf{d}$ is the first endmember signature $\mathbf{s}_1$, and $\mathbf{U}$ is formed by the rest of signatures $[\mathbf{s}_2\cdots\mathbf{s}_p]$, i.e., $\mathbf{S}=[\mathbf{d}\mathbf{U}]$. Then, (1) can be rewritten as

$$\mathbf{r}=\mathbf{d}\alpha_{\mathbf{d}}+\mathbf{U}\boldsymbol{\alpha}_{\mathbf{U}}+\mathbf{n} \tag{2}$$

where $\alpha_{\mathbf{d}}$ is the abundance fraction of the desired signature $\mathbf{d}$, and $\boldsymbol{\alpha}_{\mathbf{U}}$ is a $(p-1)\times 1$ abundance fraction vector of the undesired signatures in $\mathbf{U}$ and $\boldsymbol{\alpha}=(\alpha_{\mathbf{d}}\boldsymbol{\alpha}_{\mathbf{U}})$. Under the white-noise assumption, the OSP classifier projector $\mathbf{P}_{\mathrm{OSP}}$ was derived as [9]

$$\mathbf{P}_{\mathrm{OSP}}=\mathbf{P}_{\mathbf{U}}^{\perp}\mathbf{d} \tag{3}$$

where $\mathbf{P}_{\mathbf{U}}^{\perp}=\mathbf{I}-\mathbf{U}\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T$ is the orthogonal complement projector that maps data onto a subspace orthogonal to the undesired signatures in $\mathbf{U}$. Here, $\mathbf{I}$ denotes a $K\times K$ identity matrix. According to [10] and [11], when the OSP is implemented as an abundance estimator a constant term, $(\mathbf{d}^T\mathbf{P}_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}$ should be included to account for estimation accuracy. Then (3) becomes

$$\mathbf{P}_{\mathrm{OSP}}=\left(\mathbf{d}^T\mathbf{P}_{\mathbf{U}}^{\perp}\mathbf{d}\right)^{-1}\mathbf{P}_{\mathbf{U}}^{\perp}\mathbf{d}. \tag{4}$$

In some situations we may be only interested in a certain object present in an unknown image scene and only its spectral signature is available. CEM was developed for such case. It designs a finite-impulse response filter in such a manner that the

Manuscript received May 22, 2002; revised April 14, 2003. The work of C.-I Chang was supported by a National Research Council Senior Research Associateship sponsored by the Edgewood Chemical Biological Center, U.S. Army Soldier and Biological Command, and from the TRW Foundation.

Q. Du is with the Department of Electrical Engineering and Computer Science, Texas A&M University, Kingsville, TX 78363 USA.

H. Ren is with the National Research Council, Washington, DC 20001 USA.

C.-I Chang is with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA.

Digital Object Identifier 10.1109/TGRS.2003.813704

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUN 2003		**00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Comparative Study for Orthogonal Subspace Projection and Constrained Energy Minimization	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Department of Computer Science and Electrical Engineering,Baltimore,MD,21250**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**5**	

filter output energy is minimized subject to a constraint imposed by desired signature of interest $\mathbf{d}$. It does not assume the linear mixture model or any noise characteristics. Let the filter be specified by the coefficients $\mathbf{w} = [w_1, w_2, \ldots, w_K]^T$. Then, the filter output for the input $\mathbf{r}_i$ is expressed by $y_i = \mathbf{w}^T\mathbf{r}_i$. The average output energy E is given by

$$E = \frac{1}{q}\sum_{i=1}^{q} y_i^2 = \frac{1}{q}\sum_{i=1}^{q} \mathbf{w}^T\mathbf{r}_i\mathbf{r}_i^T\mathbf{w} = \mathbf{w}^T\left(\frac{1}{q}\sum_{i=1}^{q}\mathbf{r}_i\mathbf{r}_i^T\right)\mathbf{w} = \mathbf{w}^T\mathbf{R}_\mathbf{r}\mathbf{w} \tag{5}$$

where $\mathbf{R}_\mathbf{r} = (1/q)\sum_{i=1}^{q}\mathbf{r}\mathbf{r}_i^T = \langle \mathbf{r}\mathbf{r}^T\rangle$ is the data sample correlation matrix. Here, $\langle\cdot\rangle$ denotes sample average over all pixels and q is the total number of pixels in the image. An optimal filter $\mathbf{w}$ should be the one that minimizes E in (5) subject to the constraint $\mathbf{w}^T\mathbf{d} = 1$. The solution of this constrained problem is [13]

$$\mathbf{w}_{\mathrm{CEM}} = \left(\mathbf{d}^T\mathbf{R}_\mathbf{r}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{r}^{-1}\mathbf{d}. \tag{6}$$

Using (6), a CEM-based filter can be designed to detect the desired target $\mathbf{d}$ while minimizing the filter output energy caused by unknown signal sources.

The CEM generally outperforms the OSP in terms of eliminating unidentified signal source and suppressing noise, but it has a poor generalization property since it is very sensitive to the knowledge of the desired signature $\mathbf{d}$ used in (6). This is because a pixel with slightly different signature from the desired signature $\mathbf{d}$ may be considered as undesired or unknown, therefore, will be eliminated. One way to mitigate this problem is to find a good representative of $\mathbf{d}$ based on a large number of samples. However, in some cases, a large number of samples may not be available. Another way is to use only part of eigenvalues and eigenvectors of $\mathbf{R}_\mathbf{r}$ to calculate $\mathbf{R}_\mathbf{r}^{-1}$ as did in [16]. But a problem associated with it is how to determine the number of eigenvalues and eigenvectors to be used. In this letter, we further investigate the relationship between the OSP and CEM, which may help us to better understand the strengths and weaknesses of both techniques and under what conditions they can perform well.

II. Relationship Between OSP and CEM

Assume the sample mean is removed. Let $\mathbf{m} = \mathbf{U}\boldsymbol{\alpha}_\mathbf{U} + \mathbf{n}$. Then (2) becomes

$$\mathbf{r} = \mathbf{d}\alpha_\mathbf{d} + \mathbf{m}. \tag{7}$$

If $\langle\alpha_\mathbf{d}\boldsymbol{\alpha}_\mathbf{U}^T\rangle = \mathbf{0}$, $\mathbf{R}_\mathbf{r}$ can be represented as [17]

$$\mathbf{R}_\mathbf{r} = \alpha_\mathbf{d}^2\mathbf{d}\mathbf{d}^T + \mathbf{R}_\mathbf{m} \tag{8}$$

where $\mathbf{R}_\mathbf{m} = \langle\mathbf{m}\mathbf{m}^T\rangle$ is the sample correlation matrix of $\mathbf{m}$.

The following matrix inversion lemma [18] is used to calculate $\mathbf{R}_\mathbf{r}^{-1}$ in (8).

Lemma: Let $\mathbf{A}$ and $\mathbf{B}$ be two positive-definite $M \times M$ matrices related by

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T \tag{9}$$

where $\mathbf{D}$ is another positive-definite $N \times N$ matrix, and $\mathbf{C}$ is an $M \times N$ matrix. $\mathbf{A}^{-1}$ can be calculated by

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}\left(\mathbf{D} + \mathbf{C}^T\mathbf{B}\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{B}. \tag{10}$$

Comparing (8) with (9), $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ correspond to $\mathbf{R}_\mathbf{r}$, $\mathbf{R}_\mathbf{m}^{-1}$, $\mathbf{d}$, and $\alpha_\mathbf{d}^{-2}$, respectively. Therefore

$$\begin{aligned}\mathbf{R}_\mathbf{r}^{-1} &= \mathbf{R}_\mathbf{m}^{-1} - \mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\left(\alpha_\mathbf{d}^{-2} + \mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1} \\ &= \mathbf{R}_\mathbf{m}^{-1} - \mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\alpha_\mathbf{d}^2\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\end{aligned} \tag{11}$$

and $\mathbf{w}_o$ is defined as

$$\begin{aligned}\mathbf{w}_o &= \mathbf{R}_\mathbf{r}^{-1}\mathbf{d} \\ &= \left[\mathbf{R}_\mathbf{m}^{-1}\mathbf{d} - \mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\alpha_\mathbf{d}^2\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right].\end{aligned} \tag{12}$$

The term $\alpha_\mathbf{d}^2\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}$ in (12) is a scalar, so it can be moved to the front of the term $\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}$ next to it. Then (12) becomes

$$\begin{aligned}\mathbf{w}_o &= \mathbf{R}_\mathbf{r}^{-1}\mathbf{d} \\ &= \left[\mathbf{R}_\mathbf{m}^{-1}\mathbf{d} - \alpha_\mathbf{d}^2\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right] \\ &= \left[1 - \alpha_\mathbf{d}^2\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right]\mathbf{R}_\mathbf{m}^{-1}\mathbf{d} \\ &= \left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}.\end{aligned} \tag{13}$$

Substituting the result in (13) into (12) and (6) and noticing $\alpha_\mathbf{d}^2\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}$ is a scalar

$$\begin{aligned}\mathbf{w}_{\mathrm{CEM}} &= \left(\mathbf{d}^T\mathbf{R}_\mathbf{r}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{r}^{-1}\mathbf{d} \\ &= \left[\mathbf{d}^T\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right]^{-1} \\ &\quad\cdot\left(1 + \alpha_\mathbf{d}^2\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{m}^{-1}\mathbf{d} \\ &= \left(\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}.\end{aligned} \tag{14}$$

If noise is white, $\mathbf{R}_\mathbf{m}$ in (8) becomes

$$\mathbf{R}_\mathbf{m} = \mathbf{U}\mathbf{R}_{\boldsymbol{\alpha}_\mathbf{U}}\mathbf{U}^T + \mathbf{R}_\mathbf{n} = \mathbf{U}\mathbf{R}_{\boldsymbol{\alpha}_\mathbf{U}}\mathbf{U}^T + \sigma^2\mathbf{I} \tag{15}$$

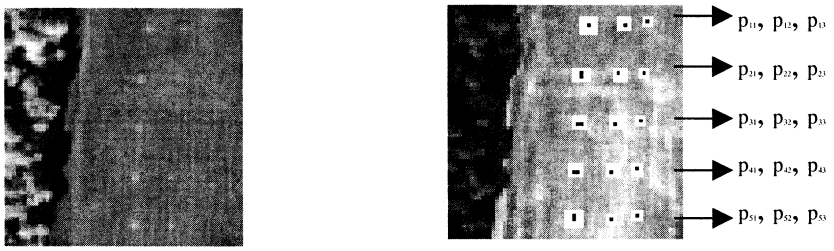
where $\mathbf{R}_{\boldsymbol{\alpha}_\mathbf{U}} = \langle\alpha_\mathbf{U}\boldsymbol{\alpha}_\mathbf{U}^T\rangle$. Using the matrix lemma again, and assuming that σ^2 is small enough compared to the undesired signatures $\mathbf{U}$, $\mathbf{R}_\mathbf{m}^{-1}$ can be computed as

$$\begin{aligned}\mathbf{R}_\mathbf{m}^{-1} &= \sigma^{-2}\mathbf{I} - \sigma^{-2}\mathbf{U}\left(\mathbf{R}_{\boldsymbol{\alpha}_\mathbf{U}}^{-1} + \sigma^{-2}\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\sigma^{-2} \\ &= \sigma^{-2}\left[\mathbf{I} - \mathbf{U}\left(\sigma^2\mathbf{R}_{\boldsymbol{\alpha}_\mathbf{U}}^{-1} + \mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\right] \\ &\approx \sigma^{-2}\mathbf{P}_\mathbf{U}^\perp.\end{aligned} \tag{16}$$

Plugging the result of (16) into (14) gives

$$\begin{aligned}\mathbf{w}_{\mathrm{CEM}} &= \left(\mathbf{d}^T\mathbf{R}_\mathbf{r}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{r}^{-1}\mathbf{d} = \left(\mathbf{d}^T\mathbf{R}_\mathbf{m}^{-1}\mathbf{d}\right)^{-1}\mathbf{R}_\mathbf{m}^{-1}\mathbf{d} \\ &\approx \left(\mathbf{d}^T\mathbf{P}_\mathbf{U}^\perp\mathbf{d}\right)^{-1}\mathbf{P}_\mathbf{U}^\perp\mathbf{d} = \mathbf{P}_{\mathrm{OSP}}.\end{aligned} \tag{17}$$

Equation (17) implies that the OSP and the CEM are essentially the same filter as long as the noise is white and as long as its variance is negligible compared to the signals, i.e., SNR



(a) (b)

Fig. 1. HYDICE image scene used in the experiment. (a) Image scene. (b) Panel arrangement.

is sufficiently high, which is generally true in hyperspectral imagery.

III. Performance Improvement for OSP

The relationship between OSP and CEM derived in Section II can be used to improve the performance of OSP. In Section II $\langle \alpha_{\mathbf{d}} \boldsymbol{\alpha}_{\mathbf{U}}^T \rangle = \mathbf{0}$ is assumed, and white noise with large SNR is required. In this section, we will show that if these two conditions are satisfied, the performance of OSP can be improved.

A. Data Whitening

When applying the OSP to classify each class, its corresponding signature in $\mathbf{S} = [\mathbf{dU}]$ is treated as $\mathbf{d}$ while making the rest $p-1$ signatures as $\mathbf{U}$. So $\langle \alpha_{\mathbf{d}} \boldsymbol{\alpha}_{\mathbf{U}}^T \rangle = \mathbf{0}$ for each pair of $\mathbf{d} : \mathbf{U}$ means $\mathbf{R}_{\boldsymbol{\alpha}} = \langle \boldsymbol{\alpha}\boldsymbol{\alpha}^T \rangle$ is a diagonal matrix. According to [10] and [11] the abundance estimation using the OSP operator in (4) can be expressed as

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}\mathbf{r}. \tag{18}$$

So $\mathbf{R}_{\boldsymbol{\alpha}}$ can be approximated as

$$\begin{aligned}\mathbf{R}_{\boldsymbol{\alpha}} &\approx \left\langle \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}}^T \right\rangle = \left\langle \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}\mathbf{r}\mathbf{r}^T\mathbf{S}^T\left(\mathbf{S}^T\mathbf{S}\right)^{-1} \right\rangle \\ &= \left(\mathbf{S}^T\mathbf{S}\right)^{-1}\mathbf{S}\mathbf{R}_{\mathbf{r}}\mathbf{S}^T\left(\mathbf{S}^T\mathbf{S}\right)^{-1}. \end{aligned} \tag{19}$$

If $\mathbf{R}_{\mathbf{r}}$ is whitened to be the identity matrix and the signatures in $\mathbf{S}$ are orthogonal to each other, i.e., $\left(\mathbf{S}^T\mathbf{S}\right)^{-1} = \mathbf{I}$, then $\mathbf{R}_{\boldsymbol{\alpha}} = \mathbf{I}$ is a diagonal matrix.

The whitening of $\mathbf{R}_{\mathbf{r}}$ can be achieved by generating a data-whitening operator $\mathbf{P}_w$ as

$$\mathbf{P}_w^T = \mathbf{V}_2^{-1/2}\mathbf{V}_1^T. \tag{20}$$

Then, $\mathbf{R}_{\mathbf{r}}$ can be whitened by applying $\mathbf{P}_w^T\mathbf{r}$ to all the pixels. Here, $\mathbf{V}_1$ and $\mathbf{V}_2$ are eigenvector and eigenvalue matrices of $\mathbf{R}_{\mathbf{r}}$ respectively. They can be determined by eigendecomposition $\mathbf{R}_{\mathbf{r}} = \mathbf{V}_1\mathbf{V}_2\mathbf{V}_1^T$, where $\mathbf{V}_1 = [\mathbf{v}_1, \ldots, \mathbf{v}_k, \ldots, \mathbf{v}_K]$ with $\mathbf{v}_k$ being the kth eigenvector and $\mathbf{V}_2$ is a diagonal matrix with the kth diagonal item λ_k being the kth eigenvalue corresponding to $\mathbf{v}_k$, i.e., $\mathbf{V}_2 = \operatorname{diag}\{\lambda_1, \ldots, \lambda_k, \ldots, \lambda_K\}$.

As for the orthogonality among the signatures in $\mathbf{S}$, the Gram–Schmidt orthogonaliztion process can be used. But it becomes null when signature subspace of $\mathbf{U}$ in $\mathbf{P}_{\mathbf{U}}^{\perp}$ is constructed using $\mathbf{U}\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T$. So in practice the orthogonalization is unnecessary.

B. Noise Whitening

The assumption about the white noise may not be true in practice. And the noise in the whitened data generally is not white. A noise-whitening process is needed via noise estimation. The noise variance can be estimated by exploiting the interband correlation such as residual-based estimation [19] and the intra/interband correlation such as linear regression model-based prediction [20]. A noise covariance matrix can be estimated using neighborhood difference method [21] and Laplacian operator [22]. We find that an accurate estimate of band-to-band noise correlation is generally difficult to achieve. So here we only estimate noise variance and construct a diagonal noise covariance matrix using the method in [20] because of its relative efficiency and simplicity. After the estimated noise covariance matrix is constructed as $\boldsymbol{\Sigma}_n = \operatorname{diag}\left\{\sigma_{n_1}^2, \ldots \sigma_{n_k}^2, \ldots, \sigma_{n_K}^2\right\}$, noise can be whitened by applying $\boldsymbol{\Sigma}_n^{-1/2}\mathbf{r}$ to all the pixels.

IV. Experiment

The data used in the experiments is the Hyperspectral Digital Imagery Collection Experiment (HYDICE) data. The image scene of size 64×64 shown in Fig. 1(a) was collected in Maryland in 1995 from the flight altitude of 10 000 ft with approximately 1.5 m GSD. Removing bands with low SNR results in 169 data dimensions. There are 15 panels present in the image scene, which are arranged in a 5×3 matrix. Each element in this matrix is denoted by p_{ij} with row indexed by i and column indexed by j. The three panels in the same row were made from the same material and are of size 3×3, 2×2, and 1×1 respectively, and they are considered as a single class. The ground truth map is provided in Fig. 1(b) and shows the precise spatial locations of panel pixels where the black pixels are referred to as panel center pixels, and white pixels are considered as panel pixels mixed with background pixels. The panels in each row are in a same class and the signatures of these five panel classes are very similar. In addition to the panel signatures, two background signatures were also generated from the grass field and tree line to the left of the panels.

The classification results using OSP and CEM are shown in Figs. 2 and 3. CEM provided a better result because it correctly detected the five panel classes and successfully eliminated background noise. The result of OSP contained larger number of background pixels. It was improved after noise whitening, as shown in Fig. 4, where the improvement was obvious when classifying the panels in rows 1, 4, and 5 in terms of better background signature elimination. This is because the OSP makes white-noise assumption, and the noise-whitening

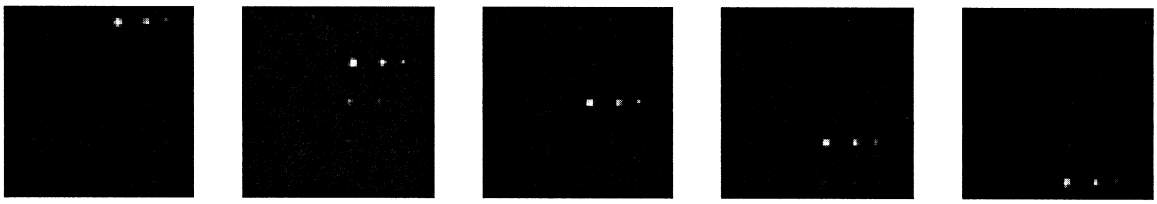


Fig. 2. Classification result of CEM. (Left to right) P1, P2, P3, P4, P5.

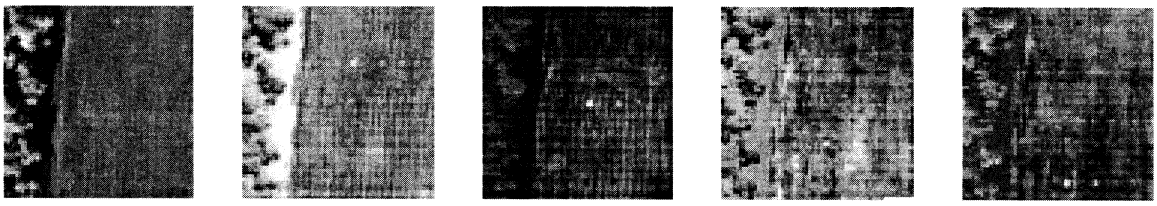


Fig. 3. Classification result of OSP. (Left to right) P1, P2, P3, P4, P5.

Fig. 4. Classification result of OSP after noise whitening (from left to right: P1, P2, P3, P4, P5).

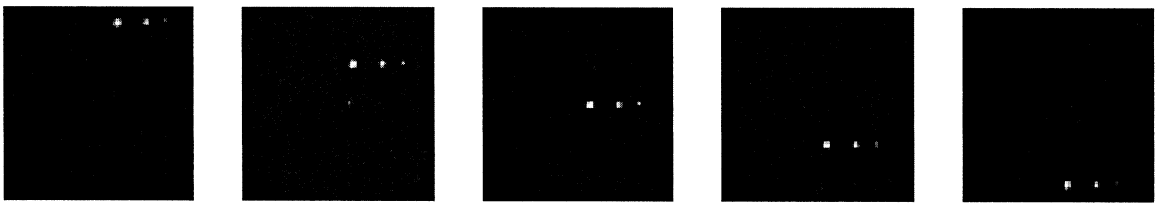


Fig. 5. The classification result of OSP after data whitening. (Left to right) P1, P2, P3, P4, P5.

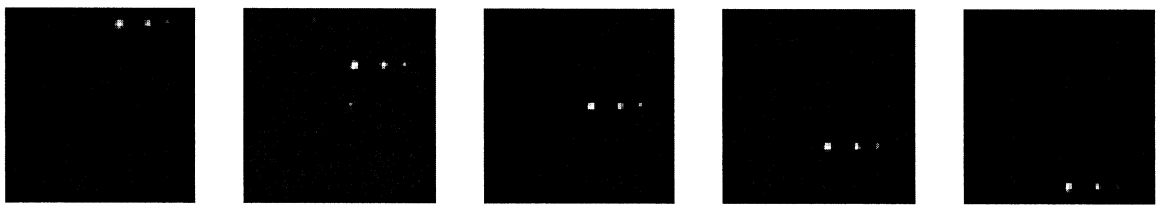


Fig. 6. Classification result of OSP after data whitening and noise whitening. (Left to right) P1, P2, P3, P4, P5.

process should be able to improve its performance. Fig. 5 shows the result of OSP after data whitening. The improvement was significant and the result was comparable to that of CEM in Fig. 2. Fig. 6 presents the result of OSP after data whitening followed by a noise-whitening process. The difference between Figs. 5 and 6 is inappreciable. This may be because the noise correlation was greatly reduced by the data-whitening process. Using the noise estimate method in Section III-B, we found that the noise variance in 161 out of 169 bands were close to unity. So in this experiment, the noise-whitening process based on the noise estimation technique in Section III-B could not further improve the performance after the data was whitened.

The images shown in Figs. 2–6 are grayscale images with the pixel gray level corresponding to the abundance fractions of a specific endmember. In order to make a quantitative comparison, we converted them to binary images by using the 50% of the maximal abundance fraction as cut-off threshold. Table I lists the number of correctly detected pixels N_D and false-alarm pixels N_F using the CEM in Fig. 2, the original OSP in Fig. 3, the improved OSP in Fig. 4 with the noise-whitening process only (OSP-M1), the improved OSP in Fig. 5 with the data-whitening process only (OSP-M2) and the improved OSP with the data-whitening process followed by the noise-whitening process (OSP-M3). The CEM could detect the

TABLE I
TALLY OF THE NUMBER OF PIXELS DETECTED AND FALSE-ALARMED USING DIFFERENT METHODS

		CEM		OSP		OSP-M1		OSP-M2		OSP-M3	
	N	N_D	N_F	N_D	N_F	N_D	N_F	N_D	N_F	N_D	N_F
P1	3	2	0	1	232	2	0	2	0	2	0
P2	4	3	0	4	2693	4	315	3	0	3	0
P3	4	3	0	4	57	4	5	3	0	3	0
P4	4	3	0	4	2274	4	28	3	0	3	0
P5	4	3	0	4	1191	3	4	3	0	3	0
Total	19	14		17		17		14		14	

14 out of 19 panel pixels without false alarm. The result of OSP contained large false-alarm rate, which means the panels could not be classified correctly as shown in Fig. 3. The OSP-M1 greatly reduced the false-alarm rates while detecting the 17 out of 19 panel pixels. The OSP-M2 significantly improved the performance of the OSP and provided the same results as CEM. No further improvement was provided by the OSP-M3 in this experiment.

This experiment demonstrates that either a noise-whitening process and a data-whitening process can bring about improvement to the performance of OSP. But the improvement from the noise-whitening process is limited by the accuracy of the noise estimate.

V. CONCLUSION

The relationship between OSP and CEM is investigated. It has been shown that when the noise is white with large SNR, the OSP and the CEM perform very closely. In this case, they can be considered essentially the same filter. Based on this relationship, the performance of OSP can be improved through data-whitening and noise-whitening processes. Future research will focus on a more effective technique to estimate the noise covariance matrix to be used in the noise-whitening process.

ACKNOWLEDGMENT

This work was conducted when H. Ren was a National Research Council Associateship fellow and worked in the U.S. Army ECBC.

REFERENCES

[1] J. B. Adams and M. O. Smith, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 suite," *J. Geophys. Res.*, vol. 91, no. 8, pp. 8098–8112, 1986.

[2] J. W. Boardman, "Inversion of imaging spectrometry data using singular value decomposition," in *Proc. IGARSS*, 1989, pp. 2069–2072.

[3] A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fischer, III, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images," in *Proc. 2nd AVIRIS Workshop*, Owens Valley, CA, 1990, pp. 243–270.

[4] A. F. H. Goetz and J. W. Boardman, "Quantitative determination of imaging spectrometer specifications based on spectral mixing models," in *Proc. IGARSS*, 1989, pp. 1036–1039.

[5] D. E. Sabol, J. B. Adams, and M. O. Smith, "Quantitative sub-pixel spectral detection of targets in multispectral images," *J. Geophys. Res.*, pp. 2659–2672, 1992.

[6] J. B. Adams, M. O. Smith, and A. R. Gillespie, "Image spectroscopy: Interpretation based on spectral mixture analysis," in *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, C. M. Pieters and P. A. Englert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1993, pp. 145–166.

[7] M. O. Smith, J. B. Adams, and D. E. Sabol, "Spectral mixture analysis-new strategies for the analysis of multispectral data," in *Image Spectroscopy—A Tool for Environmental Observations*, J. Hill and J. Mergier, Eds. Brussels, Belgium: ECSE, EEC, EAEC, , 1994, pp. 125–143.

[8] J. J. Settle and N. A. Drake, "Linear mixing and estimation of ground cover proportions," *Int. J. Remote Sens.*, vol. 14, no. 6, pp. 1159–1177, 1993.

[9] 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, pp. 779–785, July 1994.

[10] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 1045–1046, July 1996.

[11] C.-I Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1030–1032, May 1998.

[12] C.-I Chang, X. Zhao, M. L. G. Althouse, and J.-J. Pan, "Least squares subspace projection approach to mixed pixel classification in hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 898–912, Mar. 1998.

[13] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Dept. Elect. Eng., Univ. Maryland Baltimore County, Baltimore, MD, 1993.

[14] J. J. Settle, "On constrained energy minimization and the partial unmixing of multispectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 718–721, Mar. 2002.

[15] C.-I Chang, "Relationship among orthogonal subspace projection, constrained energy minimization and RX-algorithm," in *SPIE Conf. Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery VIII*, vol. 4725, Orlando, FL, Apr. 1–5, 2002.

[16] C.-I Chang and D. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1144–1159, May 2000.

[17] S. Haykin, *Advances in Spectrum Analysis and Array Processing*. Upper Saddle River, NJ: Prentice-Hall, 1991.

[18] ——, *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[19] R. E. Roger, "Principal components transform with simple, automatic noise adjustment," *Int. J. Remote Sens.*, vol. 17, no. 14, pp. 2719–2727, 1996.

[20] R. E. Roger and J. F. Arnold, "Reliability estimating the noise in AVIRIS hyperspectral images," *Int. J. Remote Sens.*, vol. 17, no. 10, pp. 1951–1962, 1996.

[21] A. 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, pp. 65–74, Jan. 1988.

[22] T. Yamamoto, H. Hanaizumi, and S. Chino, "A change detection method for remotely sensed multispectral and multitemporal images using 3-D segmentation," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 976–985, May 2001.