Orthogonal Subspace Projection (OSP) Revisited: A Comprehensive Study and Analysis

Chein-I Chang, *Senior Member, IEEE*

***Abstract*—The orthogonal subspace projection (OSP) approach has received considerable interest in hyperspectral data exploitation recently. It has been shown to be a versatile technique for a wide range of applications. Unfortunately, insights into its design rationale have not been investigated and have yet to be explored. This paper conducts a comprehensive study and analysis on the OSP from several signal processing perspectives and further discusses in depth how to effectively operate the OSP using different levels of *a priori* target knowledge for target detection and classification. Additionally, it looks into various assumptions made in the OSP and analyzes filters with different forms, some of which turn out to be well-known and popular target detectors and classifiers. It also shows how the OSP is related to the well-known least-squares-based linear spectral mixture analysis and how the OSP takes advantage of Gaussian noise to arrive at the Gaussian maximum-likelihood detector/estimator and likelihood ratio test. Extensive experiments are also included in this paper to simulate various scenarios to illustrate the utility of the OSP operating under various assumptions and different degrees of target knowledge.**

***Index Terms*—Classifier, constrained energy minimization (CEM), $(\mathbf{d}, \mathbf{U})$ model, linear discriminant analysis, orthogonal subspace projection (OSP), OSP anomaly detector (OSPAD), OSP model, RX detector (RXD), signal detection, signal parameter estimation, target-constrained interference-minimized filter (TCIMF).**

I. Introduction

HYPERSPECTRAL imagery provides additional benefits over multispectral imagery in many applications, e.g., detection, discrimination, classification, quantification, identification, etc. In the early days, hyperspectral imagery has been processed and analyzed by multispectral image processing algorithms via a preprocessing such as feature extraction, dimensionality reduction, and band selection. Such multispectral-to-hyperspectral approaches have achieved some success and may have led to a brief that hyperspectral imaging is nothing more than a straightforward extension of multispectral image processing. It seems to be not the case. When the spectral resolution is low as multispectral images are, the image processing techniques are generally developed to explore spatial information such as in a geographical information system (GIS) [1] for spatial domain analysis. Therefore, as spectral resolution is increased significantly like hyperspectral imagery, such spatial domain-based multispectral imaging techniques may be found to be less effective in certain applications. In particular, if objects of interest only account for a small population, the techniques based on spatial information can easily break down. In some cases where the object size may be smaller than the pixel resolution, e.g., rare minerals in geology, special species in agriculture and ecology, small vehicles in battlefields, etc., the data analysis must rely solely on spectral information that can only be obtained and provided by a single pixel and be processed at subpixel level. In order to address this problem, linear unmixing has been developed to exploit pixel-level spectral information for image analysis. Its success in both multispectral and hyperspectral image analysis has been demonstrated in many applications.

In order to further facilitate linear unmixing applications in hyperspectral imagery, Harsanyi and Chang recently developed a hyperspectral image classification technique, referred to as orthogonal subspace projection (OSP) from a viewpoint of hyperspectral imagery [2]. Their idea was based on two aspects: 1) how to best utilize the target knowledge provided *a priori* and 2) how to effectively make use of hundreds of available contiguous spectral bands. As for case 1), the prior target knowledge is characterized in accordance with target signatures of interest, referred to as the desired target signature $\mathbf{d}$ and undesired target signature matrix $\mathbf{U}$ formed by those target signatures that are known but not wanted in image analysis. The OSP was believed to be the first approach proposed to separate the $\mathbf{d}$ from the $\mathbf{U}$ in a signal detection model, then eliminate the undesired target signatures in $\mathbf{U}$ prior to detection of the $\mathbf{d}$ so as to improve signal detectability. As for case 2), the issue of how to effectively use available spectral bands can be best explained by the well-known pigeon-hole principle in discrete mathematics [3]. Suppose that there are 13 pigeons flying into a dozen of pigeon holes (nests). The principle says that there must exist at least one pigeon hole that ought to accommodate at least two or more pigeons. Now, if we interpret the target signatures of interest and the number of spectral bands as the pigeons and the number of pigeon holes, respectively, then we can use one spectral band to accommodate a distinct target signature. In order to make sure that no more than one target signature is accommodated in a single spectral band, a spectral band must be disposed once it is used for target signature accommodation. In doing so, the orthogonal subspace projection is introduced as a mechanism to separate one spectral band from another so that target signatures accommodated in two separate spectral bands are orthogonal to each other. In this case, one band will not share its target information with another band. However, for this approach to be effective, the number of spectral bands must be no less than the number of target signatures of interest. For hyperspectral imagery, this requirement seems to be met automatically, and the pigeon-hole principle is always valid. Unfortunately, using spectral dimensionality as a means

Manuscript received December 28, 2003; revised September 22, 2004.

The author is with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250 USA (e-mail: cchang@umbc.edu).

Digital Object Identifier 10.1109/TGRS.2004.839543

for target detection, classification and identification is generally not applicable to multispectral imagery, which usually has fewer spectral bands than the number of target signatures of interest [4]. For instance, SPOT image data have three spectral bands that can be used for data analysis. If more than three target signatures need to be analyzed, the idea of using spectral bands for target detection and classification may not work effectively [4]. To circumvent this difficulty, Ren and Chang recently developed a generalized OSP that included a dimensionality expansion technique to expand the number of spectral bands nonlinearly for the OSP to have sufficient spectral dimensions to carry out orthogonal projection [5], [6]. Its utility was further extended to magnetic resonance image (MRI) classification [7].

Many OSP-based algorithms have been developed for various applications [8] since the OSP was introduced in 1994 [2]. However, insights into the OSP have not been investigated and have yet to be explored. For example, the noise assumption is not necessarily Gaussian as commonly assumed. However, if the noise is assumed to be Gaussian, it has been shown in [9]–[11] that the OSP performed essentially the Gaussian maximum-likelihood estimator [12]. Nevertheless, from a technical point of view, the design concepts of these two techniques are different. The OSP is derived from the signal-to-noise ratio (SNR) using a signal detection approach compared to the parametric estimation-based Gaussian maximum-likelihood estimator. Furthermore, the OSP can be shown in this paper to perform as a least squares estimator that is identical to the least-squares-based linear spectral mixture analysis.

When the OSP was first developed, it required the full knowledge of image endmembers present in the image data. Such complete *a priori* information may be difficult to obtain in reality, if not impossible. Two approaches have been developed to mitigate this dilemma. One is to develop unsupervised algorithms to obtain the necessary information directly from the data required for processing [8], [13]. This type of information generated by an unsupervised means is referred to as *a posteriori* information as opposed to *a priori* information provided prior to data processing. Since the accuracy of the *a posteriori* information is closely related to the unsupervised method that is used to generate the *a posteriori* information, it may not be always reliable. To avoid this problem, a second approach is to suppress unknown information with no need of *a posteriori* information. One way to do so is the constrained energy minimization (CEM) developed by Harsanyi in his dissertation [14], which only needs the knowledge of the desired signal source. Other than that, there is no knowledge required. This approach is particularly useful and attractive in the case that the image background is unknown and complicated or very difficult to characterize. The CEM was later extended to the target-constrained interference-minimized filter (TCIMF) in [15], which characterized signal sources into three separate information sources, desired, undesired, and interference. Using this three-source model, the TCIMF could detect multiple desired signal sources, annihilate undesired signal sources while suppressing interference caused by other signal sources at the same time. Comparing to the OSP that only deals with desired and undesired signal sources and the CEM that only considers the desired signal source and interference without taking into account the undesired signal sources, the TCIMF combines both the OSP and the CEM into one filter operation and includes them as its special cases. Interestingly, as will be shown, the CEM and the TCIMF can be also interpreted as various versions of the OSP operating different degrees of target knowledge. In other words, the OSP can be considered as a spectral correlation-whitened version of the TCIMF, while the TCIMF can be thought of as the OSP version of the CEM. Specifically, when the sample spectral correlation matrix in the TCIMF is whitened (i.e., decorrelated), the TCIMF performs as if it was the OSP. On the other hand, when the CEM operates in the same way that the OSP eliminates the undesired target signatures, the CEM becomes the TCIMF. In either case, both the CEM and the TCIMF can be considered to be derived from the OSP and regarded as variants of the OSP based on the knowledge to be used in their filter design. Various relationships among these approaches have been documented in [8], [16], and [17].

With all things considered above, this paper investigates two intriguing issues, "to what extent can the OSP be used?" and "how does the OSP operate the knowledge available to users?" The first issue will be addressed by deriving the OSP from three signal processing perspectives, signal detection, linear discriminant analysis and parameter estimation which provide evidence that the OSP is indeed a versatile technique for a variety of applications. In doing so, we introduce two new signal models, called $(\mathbf{d},\mathbf{U})$ model and OSP model. The former separates a desired signal source $\mathbf{d}$ from undesired signal sources in $\mathbf{U}$ based on knowledge provided *a priori* so that these two different types of signal sources can be taken care of separately. The latter annihilates the undesired signal sources in $\mathbf{U}$ from the $(\mathbf{d},\mathbf{U})$ model via an OSP operator to reduce the interference caused by the $\mathbf{U}$ so as to improve and enhance the detectability of the $\mathbf{d}$. The second issue will be investigated by looking into how the target information is used in the OSP. Of particular interest is an issue of "how does CEM perform compared to the OSP, provided that the undesired signal sources are known *a priori* and further annihilated before the CEM is implemented?" More specifically, "how does the CEM perform compared to the OSP if the OSP model is used?" In addressing this issue, many interesting results can be obtained based on such OSP model. Interestingly, under this circumstance, the commonly used least-squares-based linear spectral mixture analysis turns out to be the OSP. Additionally, we will also show how the OSP can be implemented without prior knowledge where the OSP takes advantage of the sample spectral correlation to approximate the information that is supposed to be provided by prior knowledge but is not available at the time of data processing. As a result, the OSP operates the same form of the RX algorithm developed by Reed and Yu [18]. Furthermore, the low probability detector developed in [14] can be, therefore, also interpreted as a variant of the OSP from this aspect.

This paper is organized as follows. Section II describes three different signal processing perspectives to derive the OSP, which are signal detection, linear discriminant analysis and parameter estimation. In order to accomplish this task, two signal models, the $(\mathbf{d},\mathbf{U})$ model and the OSP model are also introduced. Section III investigates the role of Gaussian noise in the OSP and also provides experiments for demonstration. Section IV shows that the CEM can be viewed as another version of the OSP im-

plemented with partial knowledge. Examples are included to illustrate their relationship. Section V further extends the OSP to the OSP with no prior knowledge which can be considered as anomaly detection. Section VI concludes some remarks.

II. Three Perspectives to Derive OSP

Suppose that L is the number of spectral bands and $\mathbf{r}$ is an L-dimensional image pixel vector. Assume that there are p targets, $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p$ present in an image scene. Let $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p$ denote their corresponding target signatures, which are generally referred to as digital numbers (DN). A linear mixture model of $\mathbf{r}$ models the spectral signature of $\mathbf{r}$ as a linear combination of $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p$ with appropriate abundance fractions specified by $\alpha_1, \alpha_2, \ldots, \alpha_p$. More precisely, $\mathbf{r}$ is an $L \times 1$ column vector, and $\mathbf{M}$ is an $L \times p$ target spectral signature matrix, denoted by $[\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p]$, where $\mathbf{m}_j$ is an $L \times 1$ column vector represented by the spectral signature of the jth target $\mathbf{t}_j$ resident in the pixel vector $\mathbf{r}$. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)^T$ be a $p \times 1$ abundance column vector associated with $\mathbf{r}$ where α_j denotes the fraction of the jth target signature $\mathbf{m}_j$ present in the pixel vector $\mathbf{r}$. A classical approach to solving a mixed pixel classification problem is linear unmixing which assumes that the spectral signature of the pixel vector $\mathbf{r}$ is linearly mixed by $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p$, the spectral signatures of the p targets, $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p$ as follows:

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

where $\mathbf{n}$ is noise or can be interpreted as a measurement or model error.

Equation (1) represents a standard signal detection model where $\mathbf{M}\boldsymbol{\alpha}$ is a desired signal vector to be detected and $\mathbf{n}$ is a corrupted noise. Since we are interested in detecting one target at a time, we can divide the set of the p targets, $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_p$ into a desired target, say $\mathbf{t}_p$ and a class of undesired targets, $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{p-1}$. In this case, a logical approach is to eliminate the effects caused by the undesired targets $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{p-1}$ that are considered as interferers to $\mathbf{t}_p$ before the detection of $\mathbf{t}_p$ takes place. With annihilation of the undesired target signatures the detectability of $\mathbf{t}_p$ can be therefore enhanced. In doing so, we first separate $\mathbf{m}_p$ from $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p$ in $\mathbf{M}$ and rewrite (1) as

$$\mathbf{r} = \mathbf{d}\alpha_p + \mathbf{U}\boldsymbol{\gamma} + \mathbf{n} \tag{2}$$

where $\mathbf{d} = \mathbf{m}_p$ is the desired spectral signature of $\mathbf{t}_p$ and $\mathbf{U} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{p-1}]$ is the undesired target spectral signature matrix made up of $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{p-1}$ which are the spectral signatures of the remaining $p-1$ undesired targets, $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{p-1}$. Here, without loss of generality we assume that the desired target is a single target $\mathbf{t}_p$ and refer (2) to the $(\mathbf{d}, \mathbf{U})$ model thereafter.

A. Signal Detection Perspective Derived From the $(\mathbf{d}, \mathbf{U})$ Model and OSP Model

Using the $(\mathbf{d}, \mathbf{U})$ model specified by (2), we can design an orthogonal subspace projector to annihilate $\mathbf{U}$ from the pixel vector $\mathbf{r}$ prior to detection of $\mathbf{t}_p$. One such desired orthogonal subspace projector was derived in [2] and given by

$$P_{\mathbf{U}}^{\perp} = \mathbf{I} - \mathbf{U}\mathbf{U}^{\#} \tag{3}$$

where $\mathbf{U}^{\#} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$ is the pseudo-inverse of $\mathbf{U}$. The notation $\perp_{\mathbf{U}}$ in $P_{\mathbf{U}}^{\perp}$ indicates that the projector $P_{\mathbf{U}}^{\perp}$ maps the observed pixel vector $\mathbf{r}$ into the orthogonal complement of $\langle \mathbf{U} \rangle$, denoted by $\langle \mathbf{U} \rangle^{\perp}$.

Applying $P_{\mathbf{U}}^{\perp}$ to the $(\mathbf{d}, \mathbf{U})$ model results in a new signal detection model

$$P_{\mathbf{U}}^{\perp}\mathbf{r} = P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p + P_{\mathbf{U}}^{\perp}\mathbf{n} \tag{4}$$

where the undesired signatures in $\mathbf{U}$ have been annihilated and the original noise $\mathbf{n}$ has been also suppressed to $\tilde{\mathbf{n}} = P_{\mathbf{U}}^{\perp}\mathbf{n}$. The model specified by (4) will be referred to as the *OSP model* afterwards in this paper.

At this point, it is noteworthy to comment on distinction among the three models specified by (1), (2), and (4). The model in (1) is a general signal detection-in-noise model which only separates a signal source from noise. The $(\mathbf{d}, \mathbf{U})$ model is a signal model derived from the general signal detection-in-noise model by breaking up the considered signal source into two types of signal sources $\mathbf{d}$ and $\mathbf{U}$ provided by prior knowledge. It is a two signal sources $(\mathbf{d}, \mathbf{U})$ model where the two signal sources can be processed separately. The OSP model is a one signal source $(\mathbf{d})$ model derived from the $(\mathbf{d}, \mathbf{U})$ model with the $\mathbf{U}$ in the $(\mathbf{d}, \mathbf{U})$ model annihilated by $P_{\mathbf{U}}^{\perp}$. Therefore, the OSP model can be considered as a custom-designed signal detection-in-nose model from (1) where the signal source in (1) has been preprocessed by $P_{\mathbf{U}}^{\perp}$ for signal enhancement.

If we operate a linear filter specified by a weight vector $\mathbf{w}$ on the OSP model, the filter output is given by $\mathbf{w}^T P_{\mathbf{U}}^{\perp}\mathbf{r} = \mathbf{w}^T P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p + \mathbf{w}^T P_{\mathbf{U}}^{\perp}\mathbf{n}$. One of most commonly used optimal criteria in communications and signal processing is to maximize the filter output SNR over the weight vector $\mathbf{w}$ defined by

$$\mathrm{SNR}(\mathbf{w}) = \frac{[\mathbf{w}^T P_{\mathbf{U}}^{\perp}\mathbf{d}]^T \alpha_p^2 [\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{w}]}{\mathbf{w}^T P_{\mathbf{U}}^{\perp} E[\mathbf{n}\mathbf{n}^T] P_{\mathbf{U}}^{\perp}\mathbf{w}}. \tag{5}$$

If we further assume that $\mathbf{n}$ is additive and zero-mean white noise with variance σ^2, (5) is reduced to $\mathrm{SNR}(\mathbf{w}) = (\alpha_p^2/\sigma^2)(([\mathbf{w}^T(P_{\mathbf{U}}^{\perp}\mathbf{d})]^2)/(\mathbf{w}^T P_{\mathbf{U}}^{\perp}\mathbf{w}))$ where the linear optimal filter can be realized by a matched filter, $M_{P_{\mathbf{U}}^{\perp}\mathbf{d}}$ defined by

$$M_{P_{\mathbf{U}}^{\perp}\mathbf{d}}(\mathbf{x}) = \kappa(P_{\mathbf{U}}^{\perp}\mathbf{d})^T\mathbf{x} = \kappa\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{x} \tag{6}$$

for some nonzero constant κ with the matched signal $P_{\mathbf{U}}^{\perp}\mathbf{d}$. Applying the matched filter $M_{P_{\mathbf{U}}^{\perp}\mathbf{d}}$ to the OSP model results in

$$M_{P_{\mathbf{U}}^{\perp}\mathbf{d}}(P_{\mathbf{U}}^{\perp}\mathbf{r}) = \kappa\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p + \kappa\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{n} \tag{7}$$

which produced the maximum SNR given by $(\alpha_p^2/\sigma^2)\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d}$.

By means of (4) and (7) we can design a linear optimal signal detector for the $(\mathbf{d}, \mathbf{U})$ model, denoted by $\delta_{\mathrm{OSPD}}(\mathbf{r})$ by an undesired target signature rejector $P_{\mathbf{U}}^{\perp}$ followed by a matched filter $M_{\mathbf{d}}$ with the matched signal $\mathbf{d}$ as follows:

$$\delta_{\mathrm{OSPD}}(\mathbf{r}) = M_{\mathbf{d}} P_{\mathbf{U}}^{\perp}\mathbf{r} = \kappa\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r} \tag{8}$$

which is exactly the one derived in [2] with $\kappa = 1$, called orthogonal subspace projection (OSP) classifier.

B. Linear Discriminant Analysis Perspective From the OSP Model

The OSP model described by (4) can be also interpreted as a two-class classification problem, signal $\tilde{\mathbf{s}} = \alpha_p P_{\mathbf{U}}^{\perp}\mathbf{d}$ and noise $\tilde{\mathbf{n}} = P_{\mathbf{U}}^{\perp}\mathbf{n}$, respectively. Let $\boldsymbol{\mu}_{\tilde{\mathbf{s}}}$ and $\Sigma_{\tilde{\mathbf{s}}}$ be the mean vector

and covariance matrix of $\tilde{\mathbf{s}} = \alpha_p P_{\mathbf{U}}^{\perp}\mathbf{d}$, and $\boldsymbol{\mu}_{\tilde{\mathbf{n}}}$ and $\Sigma_{\tilde{\mathbf{n}}}$ be the mean vector and covariance matrix of $\tilde{\mathbf{n}} = P_{\mathbf{U}}^{\perp}\mathbf{n}$. Let a linear discriminant function $\mathbf{y}(\mathbf{x})$ be denoted by a linear form specified by $\mathbf{y}(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. Then, Fisher's criterion, called Fisher's ratio, is given by [19], [20]

$$J(\mathbf{w}) = \frac{\mathbf{w}^T[(\boldsymbol{\mu}_{\tilde{\mathbf{s}}} - \boldsymbol{\mu}_{\tilde{\mathbf{n}}})(\boldsymbol{\mu}_{\tilde{\mathbf{s}}} - \boldsymbol{\mu}_{\tilde{\mathbf{n}}})^T]\mathbf{w}}{\mathbf{w}^T[\Sigma_{\tilde{\mathbf{s}}} + \Sigma_{\tilde{\mathbf{n}}}]\mathbf{w}} \tag{9}$$

where $(\boldsymbol{\mu}_{\tilde{\mathbf{s}}} - \boldsymbol{\mu}_{\tilde{\mathbf{n}}})(\boldsymbol{\mu}_{\tilde{\mathbf{s}}} - \boldsymbol{\mu}_{\tilde{\mathbf{n}}})^T$ and $\Sigma_{\tilde{\mathbf{s}}} + \Sigma_{\tilde{\mathbf{n}}}$ are called between-class and within-class scatter matrices, respectively. Since (9) can be interpreted as the SNR defined by (5), finding the Fisher linear discriminant function $\mathbf{y}(\mathbf{x}) = (\mathbf{w}^{\text{Fisher}})^T\mathbf{x}$ with the specified weight vector $\mathbf{w}^{\text{Fisher}}$ is equivalent to maximizing SNR over the $\mathbf{w}$. As a result, Fisher's discriminant function for (9) denoted by $\delta_{\text{Fisher}}(\mathbf{r})$ can be derived as [19], [20]

$$\delta_{\text{Fisher}}(\mathbf{r}) = (\mathbf{w}^{\text{Fisher}})^T\mathbf{r} = \kappa\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r}, \qquad \text{with } \mathbf{w}^{\text{Fisher}} = \kappa\mathbf{d}. \tag{10}$$

The approach to arriving at the Fisher's discriminant function in (10) was the one actually used by Harsanyi and Chang in [2] to derive the OSP classifier, $\delta_{\text{OSP}}(\mathbf{r})$ given by

$$\delta_{\text{OSP}}(\mathbf{r}) = (\mathbf{w}^{\text{OSP}})^T P_{\mathbf{U}}^{\perp}\mathbf{r} = \mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r}, \qquad \text{with } \mathbf{w}^{\text{OSP}} = \mathbf{d}. \tag{11}$$

C. Parameter Estimation Perspective From the OSP Model

In signal detection, the primary task is to detect the desired target $\mathbf{t}_p$ in noise using (1). As shown above, using the OSP model specified by (4) could eventually improve and increase signal detectability of using (1). In pattern classification, the desired target signal $\mathbf{t}_p$ was discriminated from noise using a between-class scatter matrix/within-class scatter matrix criterion specified by (9). Both of these approaches do not intend to estimate its desired signature abundance fraction α_p. In this subsection, we look into a least squares (LS) approach to estimating the abundance fraction α_p of the desired target signature $\mathbf{d}$. Using the OSP model and least squares error (LSE) as the criterion for optimality, the LS estimate of $\alpha_p, \delta_{\text{LS},\alpha_p}(\mathbf{r})$ that minimizes

$$\min_{\alpha_p}\{(P_{\mathbf{U}}^{\perp}\mathbf{r} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)^T(P_{\mathbf{U}}^{\perp}\mathbf{r} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)\} \tag{12}$$

is also the least squares solution to the linear spectral mixture analysis.

Differentiating (12) with respect to α_p and setting it to zero results in

$$[P_{\mathbf{U}}^{\perp}\mathbf{r} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p]|_{\delta_{\text{LS},\alpha_p}(\mathbf{r})} = 0 \tag{13}$$

which yields the solution to (12)

$$\begin{aligned}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d}\delta_{\text{LS},\alpha_p}(\mathbf{r}) &= \mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r}\\ \Rightarrow \delta_{\text{LS},\alpha_p}(\mathbf{r}) &= (\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r}.\end{aligned} \tag{14}$$

Comparing $\delta_{\text{LS},\alpha_p}(\mathbf{r})$ to $\delta_{\text{OSP}}(\mathbf{r})$, there is a constant $(\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}$ appearing in $\delta_{\text{LS},\alpha_p}(\mathbf{r})$, but absent in $\delta_{\text{OSP}}(\mathbf{r})$. In other words, (11) and (14) are related by

$$\delta_{\text{LS},\alpha_p}(\mathbf{r}) = (\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}\delta_{\text{OSP}}(\mathbf{r}). \tag{15}$$

The constant $(\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}$ in (14) is the consequence of the LSE resulting from the estimation problem using the OSP model in (12). It is included to account for estimation accuracy, not a normalization constant as commonly assumed.

It should be noted that the approach presented above to derive $\delta_{\text{OSP}}(\mathbf{r})$ is different from that developed in [8]–[11], which used the oblique subspace projection [21].

D. Relationship Between $\delta_{\text{LS},\alpha_p}(\mathbf{r})$ and Least Squares Linear Spectral Mixture Analysis

In order to see how $\delta_{\text{LS},\alpha_p}(\mathbf{r})$ is related to the commonly used least squares linear spectral mixture analysis (LSMA), we minimize the least squares error resulting from (1) as follows:

$$\min_{\boldsymbol{\alpha}}\{(\mathbf{r} - \mathbf{M}\boldsymbol{\alpha})^T(\mathbf{r} - \mathbf{M}\boldsymbol{\alpha})\}. \tag{16}$$

The least squares solution to (16), denoted by $\hat{\alpha}_{\text{LS}}(\mathbf{r})$, is given by [22]

$$\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r}) = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{r}. \tag{17}$$

The major difference between $\delta_{\text{LS},\alpha_p}(\mathbf{r})$ and $\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r})$ is that the former is the scalar parameter estimate of α_p, whereas the latter is a vector parameter estimate of the abundance vector $\boldsymbol{\alpha}$. It has been shown in [9] that $\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r})$ can be decomposed as $\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r}) = (\hat{\boldsymbol{\gamma}}_{\text{LS}}(\mathbf{r}), \hat{\boldsymbol{\alpha}}_{\text{LS},p}(\mathbf{r}))^T$ with

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r}) &= \begin{pmatrix}\hat{\boldsymbol{\gamma}}_{\text{LS}}(\mathbf{r})\\ \hat{\alpha}_{\text{LS},p}(\mathbf{r})\end{pmatrix} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{r}\\ &= \begin{bmatrix}(\mathbf{U}^T\mathbf{U})^{-1} + \beta\mathbf{U}^{\#}\mathbf{d}\mathbf{d}^T(\mathbf{U}^{\#})^T & -\beta\mathbf{U}^{\#}\mathbf{d}\\ -\beta\mathbf{d}^T(\mathbf{U}^{\#})^T & \beta\end{bmatrix}\begin{pmatrix}\mathbf{U}^T\\ \mathbf{d}^T\end{pmatrix}\mathbf{r}\\ &= \begin{pmatrix}\mathbf{U}^{\#} + \beta\mathbf{U}^{\#}\mathbf{d}\mathbf{d}^T\mathbf{U}\mathbf{U}^{\#} - \beta\mathbf{U}^{\#}\mathbf{d}\mathbf{d}^T\\ \beta\mathbf{d}^T - \beta\mathbf{d}^T\mathbf{U}\mathbf{U}^{\#}\end{pmatrix}\mathbf{r}\\ &= \begin{pmatrix}\mathbf{U}^{\#} - \beta\mathbf{U}^{\#}\mathbf{d}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\\ \beta\mathbf{d}^T P_{\mathbf{U}}^{\perp}\end{pmatrix}\mathbf{r}\end{aligned} \tag{18}$$

where $\hat{\boldsymbol{\gamma}}_{\text{LS}}(\mathbf{r})$ is the least squares estimated abundance vector of $(\alpha_1, \alpha_2, \ldots, \alpha_{p-1})^T$ and $\beta = (\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}$. Combining (15) and (18) results in

$$\hat{\alpha}_{\text{LS},p}(\mathbf{r}) = (\mathbf{d}^T P_{U}^{\perp}\mathbf{d})^{-1}\mathbf{d}^T P_{U}^{\perp}\mathbf{r} = \delta_{\text{LS},\alpha_p}(\mathbf{r}) \tag{19}$$

where $\hat{\alpha}_{\text{LS},p}(\mathbf{r})$ is the pth component of $\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r})$ in (18) and also the least squares estimate of α_p in (15). The same argument can be carried out for all other abundance fractions, $\alpha_1, \alpha_2, \ldots, \alpha_{p-1}$.

Let $\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r}) = (\hat{\alpha}_{\text{LS},1}(\mathbf{r}), \hat{\alpha}_{\text{LS},2}(\mathbf{r}), \ldots, \hat{\alpha}_{\text{LS},p}(\mathbf{r}))^T$ and $\delta_{\text{LS},\boldsymbol{\alpha}}(\mathbf{r}) = (\delta_{\text{LS},\alpha_1}(\mathbf{r}), \delta_{\text{LS},\alpha_1}(\mathbf{r}), \ldots, \delta_{\text{LS},\alpha_p}(\mathbf{r}))^T$, then

$$\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r}) = \delta_{\text{LS},\boldsymbol{\alpha}}(\mathbf{r}) \tag{20}$$

where $\delta_{\text{LS},\alpha_j}(\mathbf{r}) = \hat{\alpha}_{\text{LS},j}(\mathbf{r})$ for $j = 1, 2, \ldots, p$.

If we further introduce a jth component projection function $\mathbf{1}_j$ defined by

$$\mathbf{1}_j(\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r})) = \hat{\alpha}_{\text{LS},j}(\mathbf{r}) \tag{21}$$

then we can rewrite (21) as

$$\mathbf{1}_j(\hat{\boldsymbol{\alpha}}_{\text{LS}}(\mathbf{r})) = \mathbf{1}_j(\delta_{\text{LS},\boldsymbol{\alpha}}(\mathbf{r})) \Rightarrow \hat{\alpha}_{\text{LS},j}(\mathbf{r}) = \delta_{\text{LS},\alpha_j}(\mathbf{r}), \quad \text{for } j = 1, 2, \ldots, p \tag{22}$$

with a particular case given by (19), $j = p$.

In light of (18)–(22), if $\delta_{\mathrm{LS},\alpha_j}(\mathbf{r})$ operates on every individual signature with α_j being one of signatures in $\mathbf{M}$, it becomes the commonly used linear unmixing solution, $\hat{\alpha}_{\mathrm{LS},j}(\mathbf{r})$. Compared to $\hat{\boldsymbol{\alpha}}_{\mathrm{LS}}(\mathbf{r})$ which solves for all p abundance fractions as a vector, the advantage of using $\delta_{\mathrm{LS},\alpha_j}(\mathbf{r})$ over $\hat{\boldsymbol{\alpha}}_{\mathrm{LS}}(\mathbf{r})$ is conceptually easy to understand and mathematically simple to implement. In other words, if we are interested in detection or estimation of a particular target signature, all we have to do is, (1) to designate the desired target signature as $\mathbf{d}$, (2) to annihilate all uninteresting signatures in $\mathbf{U}$ by $P_{\mathbf{U}}^{\perp}$, (3) to extract the $\mathbf{d}$ using a matched filter with the matched signature specified by $\mathbf{d}$. This is equivalent to using the OSP model to estimate the abundance fraction of the desired target after the undesired target signatures have been annihilated by $P_{\mathbf{U}}^{\perp}$ rather than using $\hat{\boldsymbol{\alpha}}_{\mathrm{LS}}(\mathbf{r})$ to directly estimate the entire abundance fractions $\alpha_1, \alpha_2, \ldots, \alpha_p$ via (17). More specifically, if the least squares estimation is performed for (17) using the OSP model, then (17) is reduced to

$$(\hat{\alpha}_j)_{\mathrm{LS}}(P_{\mathbf{U}}^{\perp}\mathbf{r}) = \mathbf{1}_j(\hat{\boldsymbol{\alpha}}_{\mathrm{LS}}(\mathbf{r})) = \hat{\alpha}_{\mathrm{LS},j}(\mathbf{r}), \quad \text{for } j = 1, 2, \ldots, p \tag{23}$$

where $(\hat{\alpha}_j)_{\mathrm{LS}}$ is the least squares estimate of α_j based on the $(\mathbf{d}, \mathbf{U})$ model in (2) with $\mathbf{d}$ replaced with $\mathbf{m}_j$ and $\mathbf{U} = [\mathbf{m}_1, \ldots, \mathbf{m}_{j-1}, \mathbf{m}_{j+1}, \ldots, \mathbf{m}_p]$. As a consequence, (23) is exactly identical to (22). Both (22) and (23) suggest two different ways to estimate the abundance fraction α_j for $j = 1, 2, \ldots, p$. Equation (23) first projects the data to the space that is orthogonal to the space linearly spanned by the undesired target signatures in $\mathbf{U}$ using $P_{\mathbf{U}}^{\perp}$, then estimates the abundance fraction of the desired target signature in the least squares sense. This is actually the approach taken by the OSP in (11). By contrast, (17) is the commonly used least squares LSMA which performs a vector parameter estimation, then uses a projection function defined by (21) to yield the abundance fraction of a particular desired target signature. The relationship between these two equations is delivered by (21) and (22), which have been overlooked in the past. It is very important because many subspace-based vector parameter estimation methods can be interpreted by the OSP via (18)–(23). A diagram to illustrate the least squares OSP and the least squares LSMA is depicted in Fig. 1.

As a concluding remark, it is worth noting that the idea of using the OSP model to rederive the OSP provides new insights into the OSP, particularly, for the approaches to linear discriminant analysis and parameter estimation, and the relationship between the OSP and the least squares LSMA via the OSP model.

III. Gaussian Noise in OSP

The noise assumed in (1) is nothing more than additive, zero-mean and white. More precisely, the noise is assumed to be *uncorrelated* with target signatures in $\mathbf{M}$ and also a zero-mean *decorrelated* (i.e., the noise covariance matrix is an identity matrix) random process. These two assumptions are not crucial and can be relaxed by data preprocessing. The assumption of additivity can be achieved by an estimation technique such as least squares methods in [8], [11], and [22] to remove correlation between target signal subspace and noise subspace. The assumption of zero-mean white noise can be accomplished by a prewhitening process, a widely used technique in communications and signal processing community in [23]. Since the signal-to-noise ratio is generally very high in hyperspectral imagery, the correlation of the noise subspace with the target signature subspace is significantly reduced compared to that in multispectral imagery. This may be one of major reasons that the OSP has been successful even though it violates the additivity assumption and white noise, and the consequence does not cause much performance deterioration. Nevertheless, by taking advantage of the Gaussian assumption many research efforts have produced satisfactory results [8]–[11], [24].

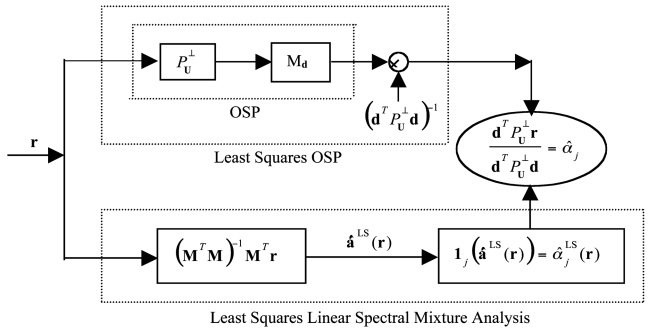

Fig. 1. Diagram to illustrate the relationship between least squares OSP and least squares linear spectral mixture analysis.

In this section, we investigate the role of Gaussian noise assumption in the OSP. Specifically, when the OSP model is cast as a two-hypotheses (signal and noise) problem, the OSP becomes a Gaussian maximum-likelihood detector. Moreover, if the OSP is used as a signal estimator, it can be further shown to be equivalent to the Gaussian maximum-likelihood estimator, which includes the two-class Gaussian discriminant function as a special case.

A. Signal Detector in Gaussian Noise Using the OSP Model

In what follows, we assume that the noise in (1) is zero-mean Gaussian with the covariance matrix given by $\Sigma_{\mathbf{n}}$. In this case, the probability distribution of $\mathbf{r}$ in (1) is a Gaussian distribution $p(\mathbf{r} \mid \mathbf{M}\boldsymbol{\alpha}) = N(\mathbf{M}\boldsymbol{\alpha}, \Sigma_{\mathbf{n}})$ with the mean vector and covariance matrix given by $\mathbf{M}\boldsymbol{\alpha}$ and $\Sigma_{\mathbf{n}}$, respectively. Similarly, we can obtain the probability distribution for $P_{\mathbf{U}}^{\perp}\mathbf{r}$ in the OSP model specified by (4), which is $p(P_{\mathbf{U}}^{\perp}\mathbf{r} \mid P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p) = N(P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p, \Sigma_{\tilde{\mathbf{n}}})$ with $\Sigma_{\tilde{\mathbf{n}}} = P_{\mathbf{U}}^{\perp}\Sigma_{\mathbf{n}}(P_{\mathbf{U}}^{\perp})^T$. Using the OSP model as a signal detection model, a standard signal detection problem can be formed by the following binary hypothesis test

$$H_0 : \tilde{\mathbf{n}} \approx p_0(\mathbf{z}) = N(\mathbf{0}, \Sigma_{\tilde{\mathbf{n}}})$$

versus

$$H_1 : P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p + \tilde{\mathbf{n}} \approx p_1(\mathbf{z}) = N(P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p, \Sigma_{\tilde{\mathbf{n}}}) \tag{24}$$

where $\mathbf{z} = P_{\mathbf{U}}^{\perp}\mathbf{r}$ and $\tilde{\mathbf{n}} = P_{\mathbf{U}}^{\perp}\mathbf{n}$. Following a standard derivation in [23], a likelihood ratio test (LRT), $\Lambda(\mathbf{z})$ resulting from (24) can be obtained by

$$\Lambda(\mathbf{z}) = \log\left(\frac{p_1(\mathbf{z})}{p_0(\mathbf{z})}\right) = (P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{z} = \alpha_p\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{r}. \tag{25}$$

However, any color Gaussian noise can be further simplified by a whitening process [23, pp. 58–60] and reduced to a white

Gaussian noise with $\Sigma_{\mathbf{n}} = \sigma^2\mathbf{I}$. In this case, the LRT $\Lambda(\mathbf{z})$ in (24) becomes $\Lambda(\mathbf{z}) = \alpha_p\sigma^{-2}\mathbf{d}^T\mathbf{z} = \alpha_p\sigma^{-2}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r}$ which is essentially $\delta_{\mathrm{OSPD}}(\mathbf{r})$ specified by (8) and $\delta_{\mathrm{OSP}}(\mathbf{r})$ specified by (11) subject to a scale constant.

B. Gaussian Maximum-Likelihood Classifier Using the OSP Model

Once again, the OSP model is used and $\mathbf{z} = P_{\mathbf{U}}^{\perp}\mathbf{r}$. Let C_0 and C_1 represent two classes corresponding to noise and signal, respectively. The discriminant functions associated with C_0 and C_1 are specified by their corresponding *a posteriori* probability distributions given by $\mathbf{y}_i(\mathbf{z}) = P(C_i\,|\,\mathbf{z})$ for $i = 0, 1$. In other words, $\mathbf{z}$ is assigned to class C_1, i.e., $\mathbf{z} \in C_1$ if $P(C_1\,|\,\mathbf{z}) > P(C_0\,|\,\mathbf{z})$, and $\mathbf{z} \in C_0$, otherwise, i.e.,

$$\begin{aligned} & P(C_1\,|\,\mathbf{z}) > P(C_0\,|\,\mathbf{z}) \\ & \Rightarrow P(\mathbf{z}\,|\,C_1)P(C_1) > P(\mathbf{z}\,|\,C_0)P(C_0) \\ & \Rightarrow \frac{P(\mathbf{z}\,|\,C_1)}{P(\mathbf{z}\,|\,C_0)} > \frac{P(C_0)}{P(C_1)} \\ & \Rightarrow \Lambda(\mathbf{z}) \equiv \log\left(\frac{p_1(\mathbf{z})}{p_0(\mathbf{z})}\right) > \log\left(\frac{P(C_0)}{P(C_1)}\right) \end{aligned} \tag{26}$$

where $p_1(\mathbf{z}) = P(\mathbf{z}\,|\,C_1)$ and $p_0(\mathbf{z}) = P(\mathbf{z}\,|\,C_0)$, and $P(C_0)$ and $P(C_1)$ are prior probabilities of C_0 and C_1, respectively. Equation (26) becomes the LRT $\Lambda(\mathbf{z})$ in (25) with the threshold given by $\log(P(C_0)/P(C_1))$. If we further assume that the prior probabilities $P(C_0)$ and $P(C_1)$ are equally likely, (26) is reduced to the maximum-likelihood detector given by

$$P(\mathbf{z}\,|\,C_1) > P(\mathbf{z}\,|\,C_0) \Leftrightarrow p_1(\mathbf{z}) > p_0(\mathbf{z}) \Leftrightarrow \mathbf{z} \in C_1. \tag{27}$$

With the Gaussian noise assumption, (27) can be calculated and expressed as follows:

$$-\frac{(\mathbf{z} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)^T\Sigma_{\tilde{\mathbf{n}}}^{-1}(\mathbf{z} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)}{2\sigma^2} > -\frac{\mathbf{z}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{z}}{2\sigma^2} \Leftrightarrow \mathbf{z} \in C_1 \tag{28}$$

$$(\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{z} > \alpha_p/2 \Leftrightarrow \mathbf{z} \in C_1. \tag{29}$$

Equation (29) makes sense, since we assume that the noise is zero-mean and the prior probabilities of the noise class C_0 and the signal class C_1 are equally likely. Substitute $P_{\mathbf{U}}^{\perp}\mathbf{r}$ for $\mathbf{z}$ in (29) yields

$$(\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r} > \alpha_p/2 \Leftrightarrow P_{\mathbf{U}}^{\perp}\mathbf{r} \in C_1 \tag{30}$$

where the left-hand side of (30) is exactly $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ given by (15). In this case, (30) can be considered as Gaussian discriminant function for the $(\mathbf{d}, \mathbf{U})$ model.

C. Gaussian Maximum-Likelihood Estimator

With the Gaussian noise assumed in the OSP model, we can obtain the maximum-likelihood estimate of the abundance fraction $\alpha_p, \delta_{\mathrm{GML},\alpha_p}(\mathbf{z})$ by

$$\delta_{\mathrm{GML},\alpha_p}(\mathbf{z}) = \max_{\alpha_p}\{p(\mathbf{z}\,|\,P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)\} \tag{31}$$

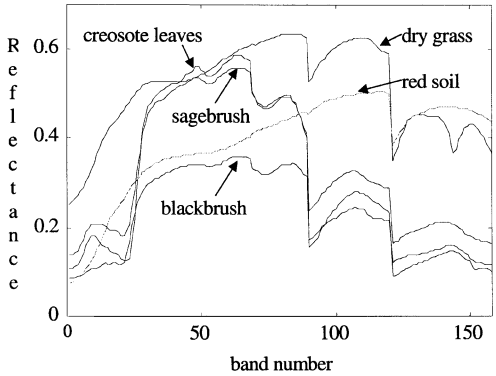

Fig. 2. Spectra of five AVIRIS signatures.

where $\mathbf{z} = P_{\mathbf{U}}^{\perp}\mathbf{r}, \Sigma_{\tilde{\mathbf{n}}} = P_{\mathbf{U}}^{\perp}\Sigma_{\mathbf{n}}(P_{\mathbf{U}}^{\perp})^T$, and $p(\mathbf{z}\,|\,P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p) = N(P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p, \Sigma_{\tilde{\mathbf{n}}})$ as defined in (24). Solving (31) is equivalent to minimizing the following Mahalanobis distance [25].

$$\delta_{\mathrm{GML},\alpha_p}(\mathbf{z}) = \min_{\alpha_p}\{(\mathbf{z} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)^T\Sigma_{\tilde{\mathbf{n}}}^{-1}(\mathbf{z} - P_{\mathbf{U}}^{\perp}\mathbf{d}\alpha_p)\} \tag{32}$$

with the solution given by

$$\delta_{\mathrm{GML},\alpha_p}(\mathbf{z}) = \left(\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{z}. \tag{33}$$

Substituting $\mathbf{z} = P_{\mathbf{U}}^{\perp}\mathbf{r}$ and using $(P_{\mathbf{U}}^{\perp})^2 = P_{\mathbf{U}}^{\perp}$ we obtain

$$\begin{aligned} \delta_{\mathrm{GML},\alpha_p}(P_{\mathbf{U}}^{\perp}\mathbf{r}) &= \left(\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{r} \\ &= \left(\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{d}\right)^{-1}\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{r}. \end{aligned} \tag{34}$$

If the Gaussian noise is further whitened, i.e., $\Sigma_{\tilde{\mathbf{n}}} = \sigma^2\mathbf{I}$, (34) is reduced to

$$\delta_{\mathrm{GML},\alpha_p}(P_{\mathbf{U}}^{\perp}\mathbf{r}) = (\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r} = \delta_{\mathrm{GML},\alpha_p}(\mathbf{r}) \tag{35}$$

which is exactly the one derived in [8]–[11]. The abundance fraction of the desired target signature $\mathbf{d}$ estimated by $\delta_{\mathrm{GML},\alpha_p}(\mathbf{r})$ was directly obtained from the Gaussian maximum-likelihood estimator $\delta_{\mathrm{GML}}(\mathbf{r}) = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{r}$ which is identical to (17). The preprocessing of using $P_{\mathbf{U}}^{\perp}$ to annihilate the undesired target signatures is not necessary for $\delta_{\mathrm{GML}}(\mathbf{r})$ because it has been taken care of in the least squares estimator shown in (18) (see [8]–[11]). Once again, (35) includes a constant $(\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}$ that represents the least square estimation error and was absent in $\delta_{\mathrm{OSP}}(\mathbf{r})$ given by (11).

D. Examples

In what follows, we conduct experiments to examine the noise assumption used in the OSP. Two scenarios will be simulated, white Gaussian noise versus white uniform noise and color Gaussian noise versus white Gaussian noise.

Example 1 (White Gaussian Noise Versus White Uniform Noise): This example demonstrates that the Gaussian noise is an unnecessary assumption for the OSP. The set of reflectance spectra considered in [2] was used for illustration and contained five reflectance spectra, dry grass, red soil, creosote leaves, blackbrush, and sagebrush. These spectra have 158 bands after water bands were removed, as shown in Fig. 2. A signature matrix $\mathbf{M}$ was formed by the dry grass, red soil and creosote leaves signatures, $\mathbf{M} = [\mathbf{m}_1\ \mathbf{m}_2\ \mathbf{m}_3]$ with their associated

TABLE I
DETECTED ABUNDANCE FRACTIONS OF "CREOSOTE LEAVES" AT 198–202 PIXELS BY $\delta_{\mathrm{OSP}}(\mathbf{r})$ AND $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$

	SNR	198	199	200	201	202	LSE
$\delta_{\mathrm{OSP}}(\mathbf{r})$ (WGN)	10:1	0.2616	0.1850	0.2531	0.2518	0.2167	0.093424
	20:1	0.2436	0.2053	0.2393	0.2386	0.2211	0.084988
	30:1	0.2375	0.2120	0.2347	0.2343	0.2226	0.082672
$\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ (WGN)	10:1	0.1160	0.0820	0.1122	0.1116	0.0961	0.00087834
	20:1	0.1080	0.0910	0.1061	0.1058	0.0980	0.00021959
	30:1	0.1053	0.0940	0.1041	0.1039	0.0987	0.000097594
$\delta_{\mathrm{OSP}}(\mathbf{r})$ (WUN)	10:1	0.1851	0.2088	0.1842	0.1828	0.2769	0.064323
	20:1	0.2053	0.2172	0.2049	0.2042	0.2512	0.069532
	30:1	0.2121	0.2200	0.2117	0.2113	0.2427	0.072167
$\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ (WUN)	10:1	0.0821	0.0926	0.0817	0.0811	0.1228	0.0015905
	20:1	0.0910	0.0963	0.0908	0.0905	0.1114	0.00039763
	30:1	0.0940	0.0975	0.0939	0.0937	0.1076	0.00017672

TABLE II
DETECTED ABUNDANCE FRACTIONS OF "CREOSOTE LEAVES" AT 198–202 PIXELS BY $\delta_{\mathrm{OSP}}(\mathbf{r})$ AND $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$

	SNR	CC	198	199	200	201	202	LSE
$\delta_{\mathrm{OSP}}(\mathbf{r})$ (GMN)	10:1	0.5	0.2419	0.1819	0.2617	0.1642	0.2054	0.068228
		0.6	0.2439	0.1790	0.2655	0.1550	0.2012	0.067641
		0.7	0.2462	0.1749	0.2697	0.1405	0.1924	0.065953
		0.8	0.2477	0.1683	0.2737	0.1141	0.1759	0.062597
	201	0.5	0.2337	0.2037	0.2436	0.1949	0.2155	0.071598
		0.6	0.2347	0.2023	0.2455	0.1903	0.2134	0.070798
		0.7	0.2359	0.2002	0.2476	0.1830	0.2090	0.069052
		0.8	0.2366	0.1968	0.2496	0.1698	0.2007	0.06545
	30:1	0.5	0.2310	0.2110	0.2376	0.2051	0.2188	0.07357
		0.6	0.2317	0.2100	0.2389	0.2020	0.2174	0.072923
		0.7	0.2324	0.2087	0.2402	0.1972	0.2145	0.071559
		0.8	0.2329	0.2064	0.2416	0.1884	0.2090	0.06873
$\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ (GMN)	10:1	0.5	0.1072	0.0807	0.1160	0.0728	0.0911	0.0015011
		0.6	0.1082	0.0794	0.1177	0.0687	0.0892	0.0018998
		0.7	0.1092	0.0776	0.1196	0.0623	0.0853	0.0026091
		0.8	0.1098	0.0746	0.1214	0.0506	0.0780	0.0041221
	20:1	0.5	0.1036	0.0903	0.1080	0.0864	0.0955	0.0003752
		0.6	0.1041	0.0897	0.1089	0.0844	0.0946	0.0004749
		0.7	0.1046	0.0888	0.1098	0.0811	0.0926	0.0006522
		0.8	0.1049	0.0873	0.1107	0.0753	0.0890	0.0010305
	30:1	0.5	0.1024	0.0936	0.1053	0.0909	0.0970	0.0001667
		0.6	0.1027	0.0931	0.1059	0.0896	0.0964	0.0002110
		0.7	0.1031	0.0925	0.1065	0.0874	0.0951	0.0002899
		0.8	0.1033	0.0915	0.1071	0.0835	0.0927	0.0004580

abundance fractions denoted by $\boldsymbol{\alpha} = (\alpha_1 \ \alpha_2 \ \alpha_3)^T$. The simulation consisted of 401 mixed pixel vectors. We started the first pixel vector with 100% red soil and 0% dry grass, then began to increase 0.25% dry grass and decrease 0.25% red soil every pixel vector until the 401st pixel vector which contained 100% dry grass. We then added creosote leaves to pixel vector numbers 198–202 at abundance fractions 10% while reducing the abundance of red soil and dry grass by multiplying their abundance fractions by 90%. For example, after addition of creosote leaves, the resulting pixel vector 200 contained 10% creosote leaves, 45% red soil and 45% dry grass. Two types of noise were simulated, white zero-mean Gaussian noise with variance σ^2 and white zero-mean uniform noise with its probability density function defined on $[-a, a]$ and variance $\sigma^2 = a^3/6$. They were added to each band to achieve the SNR defined in [2] as 50% reflectance divided by the standard deviation of the noise. Table I also tabulates the abundance fractions produced by $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ in detecting creosote leaves where only small least squares errors (LSEs) between white Gaussian and white uniform noise were observed. Nevertheless, the abundance fractions detected by $\delta_{\mathrm{OSP}}(\mathbf{r})$ and estimated by $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ were quite different. Clearly, $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ produced much more accurate estimates of abundance fractions than did $\delta_{\mathrm{OSP}}(\mathbf{r})$. This was because the scale constant $\kappa = (\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}$ in $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ was included to effectively account for estimation error.

Example 2 (Gaussian–Markov Noise): According to the model specified by (1) and the $(\mathbf{d}, \mathbf{U})$ model, the noise is assumed to be zero-mean and white. Interestingly, to the author's best knowledge, most of OSP-based techniques do not include a whitening process, but still successfully achieve their goals including Harsanyi and Chang's OSP in [2]. The reason for this is that due to high spectral resolution the SNR is generally very high in hyperspectral imagery. In this case, the noise has little impact on the OSP performance. Whether or not the noise is white becomes immaterial. This evidence is shown in the following experiments where the whitening process does improve the performance, but the gain is very small.

According to the OSP model, the whitening is only performed on interband spectral correlation at pixel level. In this

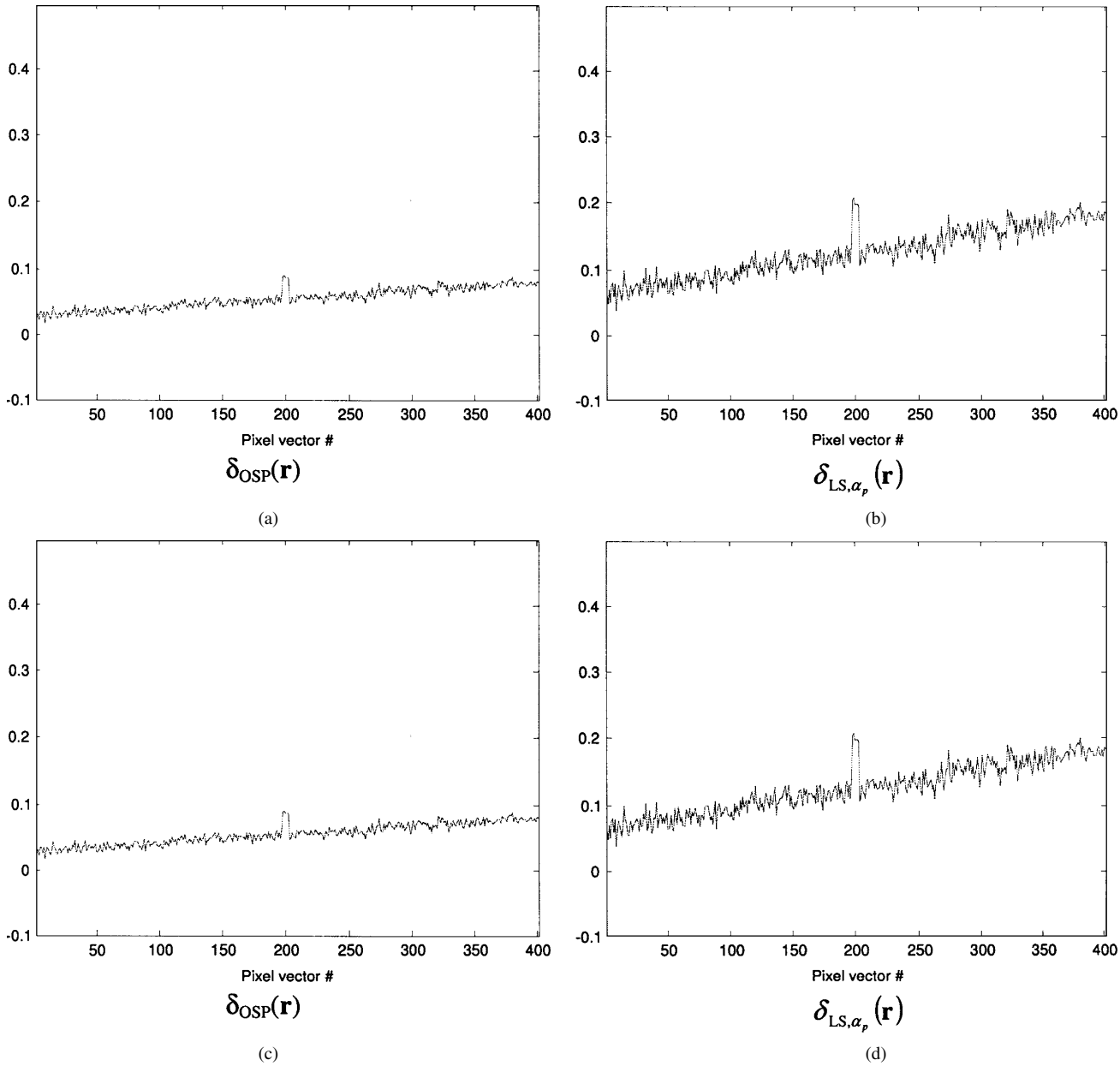

Fig. 3. Results of $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ where the GMN with CC $= 0.8$ was/was not whitened. (a) GMN was whitened. (b) GMN was not whitened.

case, a first-order zero-mean Gaussian Markov noise with the between-band correlation coefficient (CC) specified by ρ was added to each pixel vector simulated in Example 1 to achieve various levels of SNRs. The covariance matrix of such Gaussian–Markov noise has the form given by $\Sigma_{\mathbf{n}} = [\rho^{|i-j|}]$, i.e.,

$$\Sigma_{\mathbf{n}} = \begin{bmatrix} 1 & \rho & \cdots & \rho^{L-1} \\ \rho & 1 & \rho & \rho^{L-2} \\ \vdots & \rho & \ddots & \rho \\ \rho^{L-1} & \cdots & \rho & 1 \end{bmatrix}. \tag{36}$$

Table II tabulates abundance fractions produced by $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ along with their respective LSEs in detecting the creosote leaves. Comparing Table II to Table I, $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ performed slightly better in white nose than they did in color noise, but the improvements were very limited. Table II also demonstrated that the performance of $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ was deteriorated as the CC was increased. Furthermore, in order to see the effect of noise whitening, Fig. 3 shows the results of $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ where the Gaussian–Markov noise with CC $= 0.8$ was not whiten and also whitened by using the square root matrix of $\Sigma_{\tilde{\mathbf{n}}}^{-1}, \Sigma_{\tilde{\mathbf{n}}}^{-1/2}$ analytically [23, p. 60]. As shown, the whitening has slight impact on the performance of $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ in the sense that the abundance fractions of creosote leaves and background signatures were detected more accurately. This is particularly visible for $\delta_{\mathrm{OSP}}(\mathbf{r})$. These simple experiments also demonstrated that the OSP performance could be improved by a whitening process, but might not be significant. So, the payoff may not be great given that a reliable estimation of noise covariance may be difficult to obtain.

IV. OSP Implemented With Partial Knowledge

In Sections II and III, the OSP assumed the complete knowledge of target signatures, $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p$. In many practical applications, obtaining such full knowledge is generally very difficult, if not impossible. It is particularly true when the image background is not known. In this section, we investigate the issue of how to implement the OSP when there is no full knowledge available, specifically, for the case that we are only interested in specific targets, not image background or other natural sources.

In order to circumvent this dilemma, Harsanyi relaxed the requirement of complete knowledge for the OSP by developing an approach, called constrained energy minimization (CEM) in his dissertation [14]. The idea is to constrain the desired target signature while minimizing interfering effects caused by unknown as well as undesired signal sources. Since the undesired target signatures in $\mathbf{U}$ are assumed to be unknown in the CEM, these signatures cannot be annihilated by a specific operator such as $P_{\mathbf{U}}^{\perp}$. Instead, they are suppressed along with noise in terms of their energies. Despite the fact that the relationship between the OSP and CEM was reported in [8], [16], and [17], intriguing results can be obtained by showing that if the CEM is given by the same target knowledge that is provided for the OSP, the CEM actually performs exactly as does the $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$.

A. CEM

Let $\{\mathbf{r}^1, \mathbf{r}^2, \ldots, \mathbf{r}^N\}$ be a collection of image pixel vectors in an remotely sensed image where $\mathbf{r}^i = (r_1^i, r_2^i, \ldots, r_L^i)^T$ for $1 \le i \le N$ is an L-dimensional pixel vector, N is the total number of pixels in the image, and L is the total number of spectral channels. The goal is to design a finite-impulse response (FIR) linear filter with L filter coefficients $\{w_1, w_2, \ldots, w_L\}$, denoted by an L-dimensional vector $\mathbf{w} = (w_1, w_2, \ldots, w_L)^T$ that minimizes the filter output energy subject to the constraint $\mathbf{d}^T\mathbf{w} = \mathbf{w}^T\mathbf{d} = 1$.

More specifically, let y^i denote the output of the designed FIR filter resulting from the input $\mathbf{r}^i$. Then, y^i can be expressed by

$$y^i = \sum_{l=1}^{L} w_l r_l^i = (\mathbf{w})^T\mathbf{r}^i = (\mathbf{r}^i)^T\mathbf{w}. \tag{37}$$

The average energy of the filter output is then given by

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}(y^i)^T y^i &= \frac{1}{N}\left[\sum_{i=1}^{N}((\mathbf{r}^i)^T\mathbf{w})^T(\mathbf{r}^i)^T\mathbf{w}\right] \\ &= \mathbf{w}^T\left[\frac{1}{N}\sum_{i=1}^{N}\mathbf{r}^i(\mathbf{r}^i)^T\right]\mathbf{w} = \mathbf{w}^T\mathbf{R}\mathbf{w}\end{aligned} \tag{38}$$

where $\mathbf{R} = (1/N)[\sum_{i=1}^{N}\mathbf{r}^i(\mathbf{r}^i)^T]$ is the $L \times L$ autocorrelation sample matrix of the image. The CEM was developed to solve the following linearly constrained optimization problem [26]

$$\min_{\mathbf{w}}\{\mathbf{w}^T\mathbf{R}\mathbf{w}\} \quad \text{subject to } \mathbf{d}^T\mathbf{w} = \mathbf{w}^T\mathbf{d} = 1. \tag{39}$$

The optimal solution to (39) can be derived in [14] and [26] by

$$\mathbf{w}^{\mathrm{CEM}} = \frac{\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d}}. \tag{40}$$

With the optimal weight vector $\mathbf{w}^{\mathrm{CEM}}$ specified by (40) the CEM filter, $\delta_{\mathrm{CEM}}(\mathbf{r})$ derived in [14] can be obtained by

$$\delta_{\mathrm{CEM}}(\mathbf{r}) = (\mathbf{w}^{\mathrm{CEM}})^T\mathbf{r} = \left(\frac{\mathbf{R}^{-1}\mathbf{d}}{\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d}}\right)^T\mathbf{r} = \frac{\mathbf{d}^T\mathbf{R}^{-1}\mathbf{r}}{\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d}}. \tag{41}$$

Four special cases are of interest and described as follows.

1) $\mathbf{d}\perp\mathbf{U}$ (i.e., $\mathbf{d}$ Is Orthogonal to $\mathbf{U}$) and $\mathbf{R} = \mathbf{I}$ (i.e., Noise Whitened): In this case, the noise in the image data to be processed is whitened and assumed to be zero-mean and uncorrelated. So, the sample spectral correlation matrix $\mathbf{R}$ is reduced to the identity matrix $\mathbf{I}$ and $\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d} = \mathbf{d}^T\mathbf{d}$. As a result, the CEM becomes a normalized spectral matched filter, that is, $\delta_{\mathrm{CEM}}(\mathbf{r}) = (\mathbf{d}^T\mathbf{d})^{-1}\mathbf{d}^T\mathbf{r}$. On the other hand, if the desired target signature $\mathbf{d}$ is further assumed to be orthogonal to $\mathbf{U}$, i.e., $P_{\mathbf{U}}^{\perp}\mathbf{d} = \mathbf{d}$, then $\delta_{\mathrm{OSP}}(\mathbf{r}) = \mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{r} = (P_{\mathbf{U}}^{\perp}\mathbf{d})^T\mathbf{r} = \mathbf{d}^T\mathbf{r}$. This implies that CEM is identical to OSP subject to a constant $\kappa = (\mathbf{d}^T\mathbf{d})^{-1}$. Thus, both OSP and CEM can be considered as the same detector and reduced to a commonly used matched filter with the designated matched signature specified by $\mathbf{d}$.

2) Alternative Approach to Implementing CEM: Comparing $\delta_{\mathrm{CEM}}(\mathbf{r})$ in (41) to $\delta_{\mathrm{OSPD}}(\mathbf{r})$ in (8), $\delta_{\mathrm{OSP}}(\mathbf{r})$ in (11), and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ in (15), we will discover that there is a very close relationship between $P_{\mathbf{U}}^{\perp}$ and $\mathbf{R}^{-1}$. Since the knowledge of the undesired target signature matrix $\mathbf{U}$ that is assumed to be known in $P_{\mathbf{U}}^{\perp}$ in $\delta_{\mathrm{OSPD}}(\mathbf{r}), \delta_{\mathrm{OSP}}(\mathbf{r})$, and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ is not available in $\delta_{\mathrm{CEM}}(\mathbf{r}), \delta_{\mathrm{CEM}}(\mathbf{r})$ must estimate $P_{\mathbf{U}}^{\perp}$ directly from the image data. One way of doing so is to approximate the "$P_{\mathbf{U}}^{\perp}$" in the sense of minimum LSE by $\mathbf{R}^{-1}$ that can be obtained directly from the image data. More specifically, $\delta_{\mathrm{CEM}}(\mathbf{r})$ makes use of the *a posteriori* information, $\mathbf{R}^{-1}$ to approximate the *a priori* information $P_{\mathbf{U}}^{\perp}$ to accomplish what $\delta_{\mathrm{OSPD}}(\mathbf{r}), \delta_{\mathrm{OSP}}(\mathbf{r})$, and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ are able to do. The only difference is the constant κ. Since both $\delta_{\mathrm{OSPD}}(\mathbf{r})$ and $\delta_{\mathrm{OSP}}(\mathbf{r})$ are used only for abundance detection, κ has been set to 1. To the contrary, $\hat{\alpha}_{\mathrm{LS},p}(\mathbf{r})$ is an abundance estimator and the constant $\kappa = (\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d})^{-1}$ is included to account for estimation error [8]–[11]. In this case, if we replace $P_{\mathbf{U}}^{\perp}$ in $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ specified by (15) with $\mathbf{R}^{-1}$, $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ becomes $\delta_{\mathrm{CEM}}(\mathbf{r})$. This suggests that $\delta_{\mathrm{CEM}}(\mathbf{r})$ can be considered as partial knowledge version of $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ with the only knowledge provided by the desired signature $\mathbf{d}$. In this case, the sample spectral information provided by $\mathbf{R}^{-1}$ is used to replace the unknown signature matrix $P_{\mathbf{U}}^{\perp}$.

As noted above, the *a posteriori* information $\mathbf{R}^{-1}$ used in $\delta_{\mathrm{CEM}}(\mathbf{r})$ is intended to approximate the *a priori* information $P_{\mathbf{U}}^{\perp}$ used in $\delta_{\mathrm{OSP}}(\mathbf{r})$. However, the $P_{\mathbf{U}}^{\perp}$ excludes the information provided by the desired target signature $\mathbf{d}$ which is included in $\mathbf{R}^{-1}$. This observation suggests that a more accurate data sample correlation matrix used by $\delta_{\mathrm{CEM}}(\mathbf{r})$ should be the one that removes all the pixel vectors specified by $\mathbf{d}$ from $\mathbf{R}$. If we let $\tilde{\mathbf{R}}$ be such a matrix which excludes all desired target pixel vectors specified by $\mathbf{d}$ and be defined by

$$\tilde{\mathbf{R}} = \mathbf{R} - \frac{1}{N}\sum_{i}\alpha_p^i\mathbf{d}\mathbf{d}^T \tag{42}$$

where the superscript "i" runs through all target pixel vectors whose signatures are specified by $\mathbf{d}$ and α_p^i indicates their re-

spective abundance fractions contained in pixel vector $\mathbf{r}^i$. Equation (42) allows us to rewrite $\mathbf{w}^T\mathbf{R}\mathbf{w}$ in (38) as

$$\mathbf{w}^T\mathbf{R}\mathbf{w} = \mathbf{w}^T\tilde{\mathbf{R}}\mathbf{w} + \frac{1}{N}\mathbf{w}^T\left(\sum_i \alpha_p^i \mathbf{d}\mathbf{d}^T\right)\mathbf{w}. \tag{43}$$

In addition, the second term at the right-hand side of (43), $(1/N)\sum_i \alpha_p^i \mathbf{w}^T\mathbf{d}\mathbf{d}^T\mathbf{w} = (1/N)\sum_i \alpha_p^i$ can be shown to be independent of $\mathbf{w}$ because $(\mathbf{w}^T\mathbf{d})(\mathbf{w}^T\mathbf{d})^T = 1$. Solving (39) is equivalent to solving the following optimization problem

$$\min_{\mathbf{w}}\{\mathbf{w}^T\tilde{\mathbf{R}}\mathbf{w}\} \quad \text{subject to } \mathbf{w}^T\mathbf{d} = 1 \tag{44}$$

with the optimal solution given by

$$\mathbf{w}^* = (\mathbf{d}^T\tilde{\mathbf{R}}^{-1}\mathbf{d})^{-1}\tilde{\mathbf{R}}^{-1}\mathbf{d} \tag{45}$$

where the $\mathbf{w}^*$ can be obtained by simply replacing $\mathbf{R}^{-1}$ in $\mathbf{w}^{\mathrm{CEM}}$ specified by (40) with $\tilde{\mathbf{R}}^{-1}$. Therefore, technically speaking, (45) should be a more appropriate form for the CEM which is also demonstrated in [8, pp. 63–67]. Nevertheless, the CEM solution outlined by (39)–(41) is still desirable for two reasons. One is that if the number of target pixel vectors specified by the desired signature $\mathbf{d}$ is small, the impact of without removing these pixel vectors will not be significant. Another is that in many practical applications, finding all pixel vectors that are specified by the desired signature $\mathbf{d}$ may not be realistic, particularly, if the $\mathbf{d}$ is a mixed pixel signature.

3) CEM Implemented in Conjunction With $P_{\mathbf{U}}^{\perp}$: More interestingly, when the $\mathbf{U}$ is actually known, the CEM should be able to take advantage of such knowledge to annihilate the undesired signatures via $P_{\mathbf{U}}^{\perp}$ instead of suppressing these signatures. In this case, the $\mathbf{r}$ in (37) will be replaced be preprocessed image pixel, $P_{\mathbf{U}}^{\perp}\mathbf{r}$ and the desired target signature $\mathbf{d}$ is also projected to $P_{\mathbf{U}}^{\perp}\mathbf{d}$. Consequently, the constraint $\mathbf{w}^T\mathbf{d} = 1$ and the object function $\mathbf{w}^T\mathbf{R}\mathbf{w}$ in (39) must be replaced with $\mathbf{w}^T(P_{\mathbf{U}}^{\perp}\mathbf{d}) = 1$ and $\mathbf{w}^T[(P_{\mathbf{U}}^{\perp}\mathbf{r})(P_{\mathbf{U}}^{\perp}\mathbf{r})^T]\mathbf{w}$, respectively, which result in

$$\begin{aligned}\mathbf{w}^T[(P_{\mathbf{U}}^{\perp}\mathbf{r})(P_{\mathbf{U}}^{\perp}\mathbf{r})^T]\mathbf{w} &= \alpha_p^2\mathbf{w}^T[(P_{\mathbf{U}}^{\perp}\mathbf{d})(P_{\mathbf{U}}^{\perp}\mathbf{d})^T]\mathbf{w} \\ &\quad + \mathbf{w}^T[E((P_{\mathbf{U}}^{\perp}\mathbf{n})(P_{\mathbf{U}}^{\perp}\mathbf{n})^T)]\mathbf{w} \\ &= \alpha_p^2\mathbf{w}^T[(P_{\mathbf{U}}^{\perp}\mathbf{d})(P_{\mathbf{U}}^{\perp}\mathbf{d})^T]\mathbf{w} \\ &\quad + \mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w} \\ &= \alpha_p^2 + \mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w}\end{aligned} \tag{46}$$

where $\Sigma_{\tilde{\mathbf{n}}} = E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T] = E[(P_{\mathbf{U}}^{\perp}\mathbf{n})(P_{\mathbf{U}}^{\perp}\mathbf{n})^T]$ and the cross-term $\mathbf{d}^T E[\tilde{\mathbf{n}}]$ vanishes if $\mathbf{n}$ is zero-mean. As a consequence, (39) becomes

$$\min_{\mathbf{w}}\{\mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w}\} \quad \text{subject to } \mathbf{w}^T(P_{\mathbf{U}}^{\perp}\mathbf{d}) = 1. \tag{47}$$

In order to see the relationship between the OSP and the CEM, we use (5) to obtain the filter output SNR as follows:

$$\begin{aligned}\mathrm{SNR}(\mathbf{w}) &= \frac{\mathbf{w}^T\left[\left(P_{\mathbf{U}}^{\perp}\mathbf{r}\right)\left(P_{\mathbf{U}}^{\perp}\mathbf{r}\right)^T\right]\mathbf{w}}{\mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w}} \\ &= \frac{\alpha_p^2\mathbf{w}^T\left[\left(P_{\mathbf{U}}^{\perp}\mathbf{d}\right)\left(P_{\mathbf{U}}^{\perp}\mathbf{d}\right)^T\right]\mathbf{w} + \mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w}}{\mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w}} \\ &= \frac{\alpha_p^2\mathbf{w}^T[(P_{\mathbf{U}}^{\perp}\mathbf{d})(P_{\mathbf{U}}^{\perp}\mathbf{d})^T]\mathbf{w}}{\mathbf{w}^T\Sigma_{\tilde{\mathbf{n}}}\mathbf{w}} + 1.\end{aligned} \tag{48}$$

So, solving (47) is equivalent to finding the solution to the following constrained optimization problem

$$\max_{\mathbf{w}}\{\mathrm{SNR}(\mathbf{w})\} \quad \text{subject to } \mathbf{w}^T(P_{\mathbf{U}}^{\perp}\mathbf{d}) = 1. \tag{49}$$

The solution to (49) can be easily obtained by

$$\mathbf{w}^* = \Sigma_{\tilde{\mathbf{n}}}^{-1}\left(\mathbf{d}^T\Sigma_{\tilde{\mathbf{n}}}^{-1}\mathbf{d}\right)^{-1}P_{\mathbf{U}}^{\perp}\mathbf{d}. \tag{50}$$

The CEM specified by (50) is denoted by $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ and called CEM implemented in conjunction with $P_{\mathbf{U}}^{\perp}$.

4) CEM Implemented in Conjunction With $P_{\mathbf{U}}^{\perp}$ in White Noise: If we further assume that the noise in (49) is white and given by $\Sigma_{\tilde{\mathbf{n}}} = \sigma^2\mathbf{I}_{L\times L}$, (49) is reduced to

$$\min_{\mathbf{w}}\{\mathbf{w}^T\mathbf{w}\} \quad \text{subject to } \mathbf{w}^T\left(P_{\mathbf{U}}^{\perp}\mathbf{d}\right) = 1 \tag{51}$$

and (50) becomes

$$\mathbf{w}^* = (\mathbf{d}^T P_{\mathbf{U}}^{\perp}\mathbf{d})^{-1}P_{\mathbf{U}}^{\perp}\mathbf{d}. \tag{52}$$

Let the filter specified by (52) be denoted by $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp},\mathrm{white}}(\mathbf{r})$ where the CEM assumes that noise is white and the knowledge of the $\mathbf{U}$ is provided *a priori*. Coupled with the undesired signature projector $P_{\mathbf{U}}^{\perp}$, the $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp},\mathrm{white}}(\mathbf{r})$ becomes the least squares estimator $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ given by (15) and the Gaussian maximum-likelihood estimator $\delta_{\mathrm{GML},\alpha_p}(\mathbf{r})$ given by (35), respectively. This implies that if the noise is zero-mean and white and the undesired target signatures are annihilated by $P_{\mathbf{U}}^{\perp}$, then $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp},\mathrm{white}}(\mathbf{r})$ performs as if it is an abundance fraction estimator.

B. TCIMF

The CEM was originally designed to detect a single target signature. If there are multiple target signatures, it must be carried out one target signature at a time. In order to extend the CEM to a multiple-target detection technique, a target-constrained interference-minimized filter was recently developed by Ren and Chang [15], which can be viewed as a generalization of the OSP and CEM.

Let $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{n_{\mathbf{D}}}]$ denote the desired target signature matrix and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_{\mathbf{U}}}]$ be the undesired target signature matrix where $n_{\mathbf{D}}$ and $n_{\mathbf{U}}$ are the number of the desired target signatures and the number of the undesired target signatures, respectively. Now, we can develop an FIR filter that passes the desired target signatures in $\mathbf{D}$ using an $n_{\mathbf{D}} \times 1$ unity constraint vector $\mathbf{1}_{n_{\mathbf{D}}\times 1} = \underbrace{(1, 1, \ldots, 1)}_{n_{\mathbf{D}}}{}^T$ while annihilating the undesired target signatures in $\mathbf{U}$ using an $n_{\mathbf{U}} \times 1$ zero constraint vector $\mathbf{0}_{n_{\mathbf{U}}\times 1} = \underbrace{(0, 0, \ldots, 0)}_{n_{\mathbf{U}}}{}^T$. In doing so, the constraint in (39) is replaced by

$$[\mathbf{D}\mathbf{U}]^T\mathbf{w} = \begin{bmatrix}\mathbf{1}_{n_{\mathbf{D}}\times 1} \\ \mathbf{0}_{n_{\mathbf{U}}\times 1}\end{bmatrix} \tag{53}$$

and the optimization problem in (39) becomes the following linearly constrained optimization problem:

$$\min_{\mathbf{w}}\{\mathbf{w}^T\mathbf{R}\mathbf{w}\} \quad \text{subject to } [\mathbf{D}\mathbf{U}]^T\mathbf{w} = \begin{bmatrix}\mathbf{1}_{n_{\mathbf{D}}\times 1} \\ \mathbf{0}_{n_{\mathbf{U}}\times 1}\end{bmatrix}. \tag{54}$$

TABLE III
COMPARISON OF DETECTED ABUNDANCE FRACTIONS BETWEEN $\delta_{\mathrm{CEM}}(\mathbf{r})$ AND $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ WITH $P_{\mathbf{U}}^{\perp}$ ANNIHILATION

	SNR	198	199	200	201	202	LSE
$\delta_{\mathrm{CEM}}(\mathbf{r})$	30:1	0.0614	0.0597	0.0668	0.0771	0.0633	0.0060883
$\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$	30:1	0.0630	0.0606	0.0671	0.0782	0.0651	0.0056946

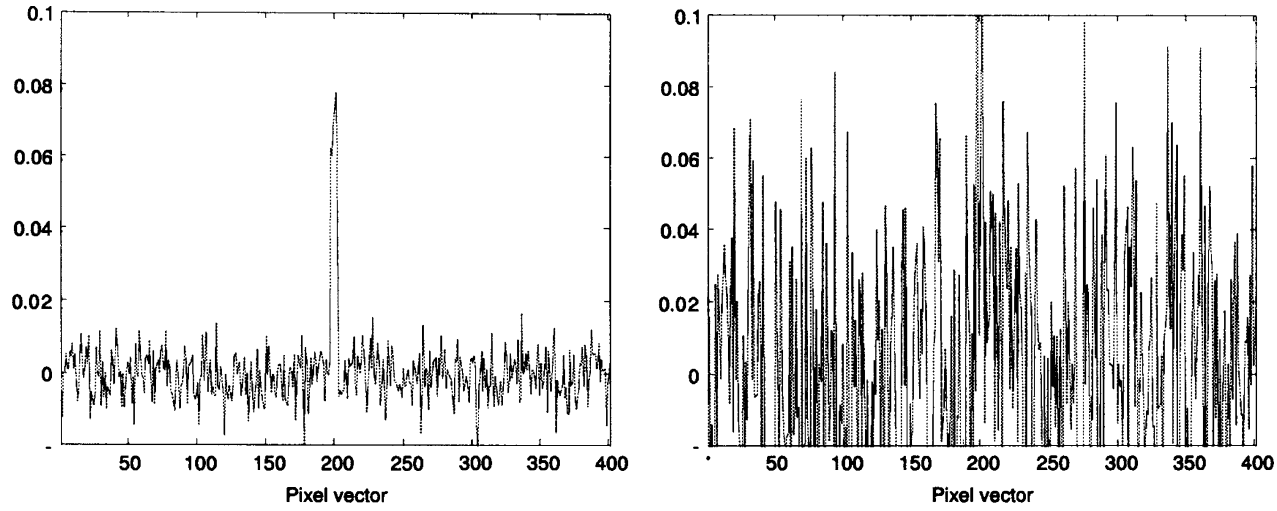

Fig. 4. Detection results of $\delta_{\mathrm{CEM}}(\mathbf{r})$ and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ with dry grass, red soil, and sagebrush annihilated by $P_{\mathbf{U}}^{\perp}$. (a) $\delta_{\mathrm{CEM}}(\mathbf{r})$. (b) $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$.

The filter solving (54) is called target-constrained interference-minimized filter in [15] and given by

$$\delta_{\mathrm{TCIMF}}(\mathbf{r}) = (\mathbf{w}^{\mathrm{TCIMF}})^T\mathbf{r} \tag{55}$$

with the optimal weight vector $\mathbf{w}^{\mathrm{TCIMF}}$ given by

$$\mathbf{w}^{\mathrm{TCIMF}} = \mathbf{R}^{-1}[\mathbf{DU}]\left([\mathbf{DU}]^T\mathbf{R}_{L\times L}^{-1}[\mathbf{DU}]\right)^{-1}\begin{bmatrix}\mathbf{1}_{n_{\mathbf{D}}\times 1}\\ \mathbf{0}_{n_{\mathbf{U}}\times 1}\end{bmatrix}. \tag{56}$$

A discussion on the relationship between the OSP and the CEM via the TCIMF can be found in [8] and [16]. In what follows, we describe more details about this relationship by considering three special cases of $\delta_{\mathrm{TCIMF}}(\mathbf{r})$.

1) $\mathbf{D} = \mathbf{m}_p = \mathbf{d}$ *With* $n_{\mathbf{D}} = 1$ *and* $\mathbf{U} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{p-1}]$ *With* $n_{\mathrm{U}} = p-1$*:* In this case, $\delta_{\mathrm{TCIMF}}(\mathbf{r})$ performs like $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$. However, there is a difference in the sense of algorithm implementation. The former performs extraction of the desired signature $\mathbf{m}_p$ and annihilation of the undesired signatures $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{p-1}$ simultaneously, whereas the latter performs the undesired signature annihilation using $P_{\mathbf{U}}^{\perp}$ followed by the CEM, $\delta_{\mathrm{CEM}}(\mathbf{r})$ in sequence. Therefore, despite the fact that both filters may produce the same results, they should be considered as separate filters. In particular, $\delta_{\mathrm{TCIMF}}(\mathbf{r})$ can be carried out in real time as noted in [8] and [27].

2) $\mathbf{R} = \mathbf{I}, \mathbf{D} = \mathbf{m}_p = \mathbf{d}$ *With* $n_{\mathbf{D}} = 1$ *and* $\mathbf{U} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{p-1}]$ *With* $n_{\mathrm{U}} = p-1$*:* In this case, $\delta_{\mathrm{TCIMF}}(\mathbf{r})$ performs like $\delta_{\mathrm{OSP}}(\mathbf{r})$, but in the mean time, it also suppresses all signatures other than desired and undesired target signatures, an operation that the OSP does not do. Let the weight vector in this case be denoted by $\mathbf{w}_{\mathbf{R}=\mathbf{I}}^{\mathrm{TCIMF}}$ and its corresponding TCIMF be denoted by $\delta_{\mathrm{TCIMF},\mathbf{R}=\mathbf{I}}(\mathbf{r})$. As derived in [8] and [16], the $\delta_{\mathrm{TCIMF},\mathbf{R}=\mathbf{I}}(\mathbf{r})$ can be shown to be equivalent to $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ or $\delta_{\mathrm{OSP}}(\mathbf{r})$ as follows:

$$\mathbf{w}_{\mathbf{R}=\mathbf{I}}^{\mathrm{TCIMF}} = \kappa P_{\mathbf{U}}^{\perp}\mathbf{d} \tag{57}$$

where $\kappa = (\mathbf{d}^T\mathbf{R}^{-1}\mathbf{d})^{-1}$, and

$$\delta_{\mathrm{TCIMF},\mathbf{R}=\mathbf{I}}(\mathbf{r}) = \delta_{\mathrm{LS},\alpha_p}(\mathbf{r}) = \kappa\delta_{\mathrm{OSP}}(\mathbf{r}). \tag{58}$$

It should be noted that the extra constant κ in (58) was a result of interference/noise suppression from the TCIMF that the OSP does not perform.

3) $\mathbf{D} = \mathbf{d}$ *and* $\mathbf{U} = \emptyset$ *(i.e., Only the Desired Signature* $\mathbf{d}$ *Is Available):* In this case, the TCIMF is further reduced to the CEM given by

$$\delta_{\mathrm{TCIMF},\mathbf{D}=\mathbf{d},\mathbf{U}=\emptyset}(\mathbf{r}) = \left(\mathbf{w}_{\mathbf{D}=\mathbf{d},\mathbf{U}=\emptyset}^{\mathrm{TCIMF}}\right)^T\mathbf{r} = \delta_{\mathrm{CEM}}(\mathbf{r}). \tag{59}$$

So, according to the three cases discussed above, on one hand, the OSP and the CEM can be considered as special cases of the TCIMF by virtue of (58) and (59). On the other hand, the OSP can be also interpreted as a data correlation-whitened version of the TCIMF with interference/noise suppression (58) and the CEM is an undesired target signature-suppressed version of the TCIMF (59). Nevertheless, there is a subtle and substantial distinction between the TCIMF and other filters such as OSP and CEM. The TCIMF can be implemented to detect and classify multiple targets, annihilate undesired targets and suppress unknown signal sources in one-shot operation in real time [27], whereas $\delta_{\mathrm{CEM}}(\mathbf{r})$ and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ can be only used for detection of a single target at a time.

C. Examples

In this subsection, we conduct a comparative analysis between the OSP and the CEM with the undesired signatures annihilated as they are in the OSP.

Example 3 (Comparative Study Between the $(\mathbf{d},\mathbf{U})$ Model and the OSP Model): This example assumes that the complete knowledge of target signatures is available. We study how the undesired signatures in $\mathbf{U}$ affect the performance of $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$. The same 401 simulated pixels used in Example 1 were also used in this example with added SNR 30 : 1 white Gaussian noise so that the results of $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ derived in Example 1 can be used for comparison. Two scenarios were studied. One is the $\mathbf{U}$ made up of dry grass and red soil. Table III tabulates the detection results of $\delta_{\mathrm{CEM}}(\mathbf{r})$ and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ with dry grass and red soil annihilated by $P_{\mathbf{U}}^{\perp}$ along with their respective averaged LSEs. As we can see from Table III, $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ with the dry grass and red soil annihilated performed slightly better than $\delta_{\mathrm{CEM}}(\mathbf{r})$ which suppressed only the dry grass and red soil. However, if we further included the sagebrush in the $\mathbf{U}$ as another undesired signature even if it was absent in the 401 simulated pixel vectors to repeat the same experiment. Fig. 4 shows the results of $\delta_{\mathrm{CEM}}(\mathbf{r})$ and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ with dry grass and red soil along with sagebrush annihilated. As we can see clearly from Fig. 4, $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ with the annihilation of $\mathbf{U}$ performed poorly compared to its counterpart $\delta_{\mathrm{CEM}}(\mathbf{r})$. This is because the signature of the sagebrush was so close to that of creosote leaves as shown in [8] and [28] that the annihilation of sagebrush also eliminated the most part of the signature of creosote leaves, which resulted in significant deterioration of signal detectability of $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$. This example demonstrated a significant difference between annihilation of undesired signatures and suppression of undesired signatures.

Example 4 (Partial Knowledge): The same 401 simulated pixel vectors used in Example 1 were once again used for the following experiments except that the blackbrush and sagebrush were added to pixel vector numbers 98–102 and pixel numbers 298–302, respectively, at abundance fractions 10%, while reducing the abundance of red soil and dry grass by multiplying their abundance fractions by 90%. In this case, there were three target signatures of interest, blackbrush, creosote leaves and sagebrush with two background signatures, red soil, and dry grass. In this example, the complete knowledge of targets of interest, blackbrush, creosote leaves, and sagebrush was also assumed to be available and the background signatures, soil and dry grass were unknown and considered to be interferers as interference. We also let $\mathbf{d}$ be the desired target signature, the $\mathbf{U}$ consist of the other two known target signatures. Like Example 3, the CEM was implemented in conjunction with/without $P_{\mathbf{U}}^{\perp}$, $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ and $\delta_{\mathrm{CEM}}(\mathbf{r})$, respectively. Fig. 5 shows the results of $\delta_{\mathrm{OSP}}(\mathbf{r}),\delta_{\mathrm{LS},\alpha_p}(\mathbf{r}),\delta_{\mathrm{CEM}}(\mathbf{r})$, and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ where figures labeled by (a), (b), and (c) were detection results of blackbrush, creosote leaves, and sagebrush, respectively, with the $\mathbf{U}$ formed by the other two signatures which served as undesired signatures. As shown in [8] and [28], the three target signatures, blackbrush, creosote leaves, and sagebrush were very similar. As a consequence, annihilating any two of these three signatures would certainly have tremendous impacts on the detection performance of the third signature. The results of Fig. 5 confirmed what we expected. That is, detecting one signature would also detect the other two signatures. Additionally, it also significantly deteriorated the ability of $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ in estimating abundance fractions as shown in Fig. 5 where the estimated abundance fractions of each of the three signatures were far more being accurate as tabulated in Table IV. Surprisingly, the detection of creosote leaves was quite different from that of blackbrush and sagebrush, as $\delta_{\mathrm{OSP}}(\mathbf{r})$ and $\delta_{\mathrm{LS},\alpha_p}(\mathbf{r})$ were implemented. The detection of creosote leaves also detected significant amounts of blackbrush and sagebrush signatures by suppressing the background signatures even if it was not supposed to do so. The detection of blackbrush and sagebrush showed very similar results and also detected visible amounts of the three signatures except that different amounts of abundance fractions of the background signatures, dry grass and red soil were detected. Similar phenomena were also observed from the detection results of $\delta_{\mathrm{CEM}}(\mathbf{r})$ and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ where detection of one signature also picked up the other signatures. Because the spectra of these three signatures were very similar, the $P_{\mathbf{U}}^{\perp}$ used in $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ also annihilated part of the desired signature before the detection of the desired signature. Consequently, $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ performed poorly compared to $\delta_{\mathrm{CEM}}(\mathbf{r})$.

Since the soil and dry grass were used as interference, the TCIMF was implemented in two scenarios. One was with $\mathbf{d}=$ a single desired target signature and $\mathbf{U}=[\text{dry grass, red soil}]$, and the other was with $\mathbf{D}=[\text{blackbrush, creosote leaves, sagebrush}]$ and $\mathbf{U}=[\text{dry grass, red soil}]$. Fig. 6 shows their detection results of $\delta_{\mathrm{TCIMF}}(\mathbf{r})$, where Fig. 6(a)–(c) was detection results of blackbrush, creosote leaves, and sagebrush, respectively, and Fig. 6(d) was the simultaneous detection result of the three signatures blackbrush, creosote leaves, and sagebrush with $\mathbf{D}=[\text{blackbrush, creosote leaves, sagebrush}]$. As indicated previously, when the TCIMF was implemented with a single target signature designated as the desired signature $\mathbf{D}=\mathbf{d}$, it would perform like $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$. This was verified by comparing the results of $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ in Figs. 5 to 6(a)–(c). Interestingly, it was not true as shown in Fig. 6(d) when $\delta_{\mathrm{TCIMF}}(\mathbf{r})$ was implemented with $\mathbf{D}=[\text{blackbrush, creosote leaves, sagebrush}]$ and $\mathbf{U}=[\text{dry grass redsoil}]$ where it performed significantly better than $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$.

V. OSP Implemented Without Knowledge

As the OSP was originally developed in [2], it required complete knowledge about the image endmembers present in the image data. Unfortunately, such requirement is seldom satisfied in reality. In order to resolve this issue, two approaches were developed previously. One is to generate desired complete knowledge directly from the image data in an unsupervised manner so that the obtained unsupervised knowledge can be used as if it was provided *a priori* [8], [13] to make the $(\mathbf{d},\mathbf{U})$ model applicable where the undesired target signature projector $P_{\mathbf{U}}^{\perp}$ can be constructed from the generated $\mathbf{U}$. Due to the fact that such

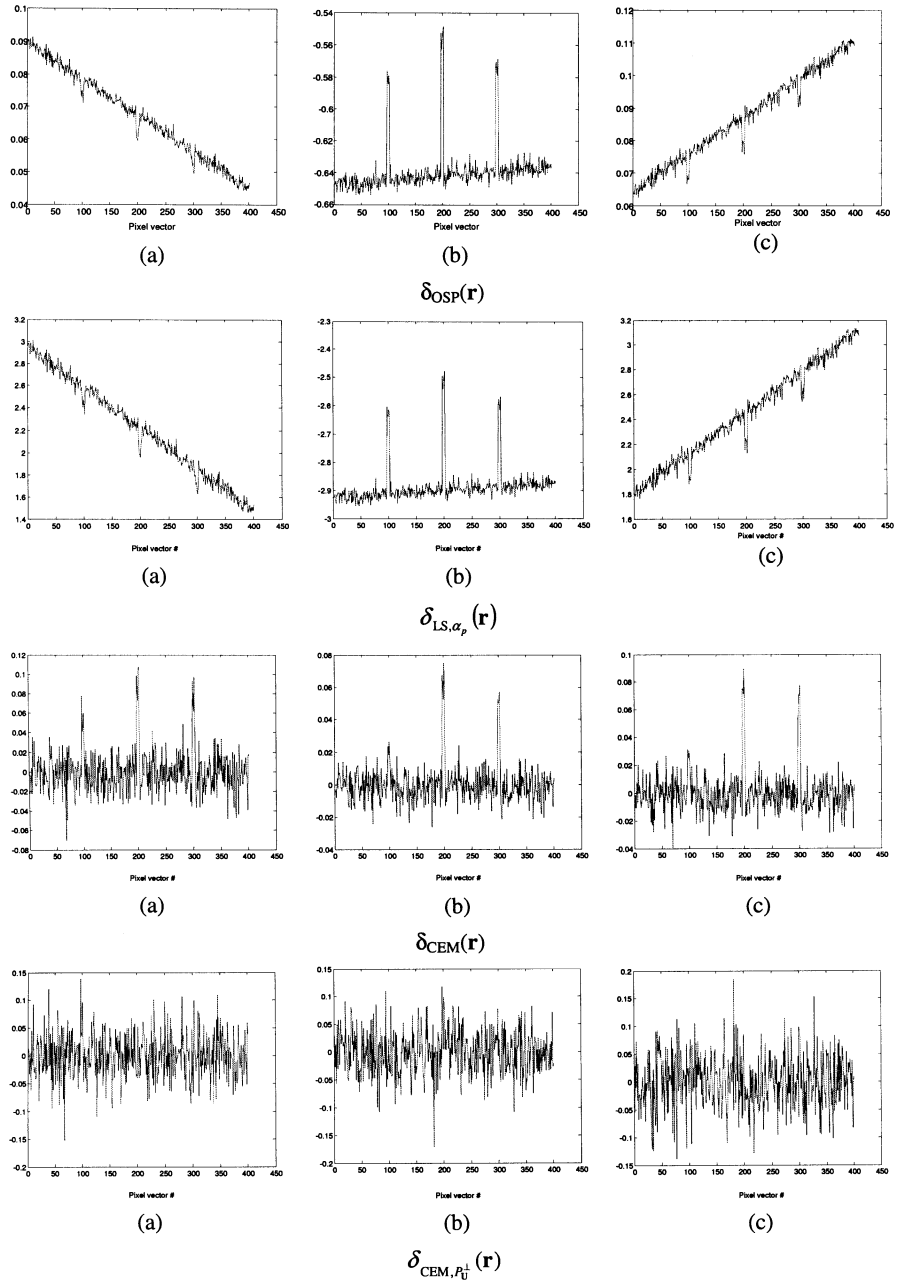

Fig. 5. Detection results of $\delta_{\mathrm{OSP}}(\mathbf{r}), \delta_{\mathrm{LS},\alpha_p}(\mathbf{r}), \delta_{\mathrm{CEM}}(\mathbf{r})$, and $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$.

TABLE IV
COMPARISON OF DETECTED ABUNDANCE FRACTIONS BY $\delta_{\mathrm{OSP}}(\mathbf{r}), \delta_{\mathrm{LS},\alpha_p}(\mathbf{r}), \delta_{\mathrm{CEM}}(\mathbf{r})$, AND $\delta_{\mathrm{CEM},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ WITH $P_{\mathbf{U}}^{\perp}$ ANNIHILATION

	198	199	200	201	202	LSE
$\delta_{OSP}(\mathbf{r})$	-0.5534	-0.5613	-0.5560	-0.5529	-0.5496	2.143
$\delta_{LS,\alpha_p}(\mathbf{r})$	-2.5008	-2.5366	-2.5124	-2.4986	-2.4837	33.9685
$\delta_{CEM}(\mathbf{r})$	0.0676	0.0618	0.0551	0.0757	0.0626	0.0065128
$\delta_{CEM,P_U^\perp}(\mathbf{r})$	0.1171	0.0678	0.0177	0.0912	0.0903	0.0082764

generated unsupervised knowledge may not be accurate, an alternative approach is to implement the OSP without appealing for unsupervised knowledge. One such approach is the CEM described in Section IV-A where only the desired target knowledge, $\mathbf{d}$ was required. Instead of trying to find unknown signatures in $\mathbf{U}$ for annihilation, the CEM suppresses all signatures other than the signature of interest. In order to accomplish that, the CEM makes use of the inverse of the sample correlation matrix, $\mathbf{R}^{-1}$ to approximate the complete knowledge provided by $P_{\mathbf{U}}^{\perp}$ in the OSP. As a result, the OSP and the CEM can be related by (58) and (59) using the TCIMF as a bridge. Another approach is to implement the OSP without prior knowledge. As

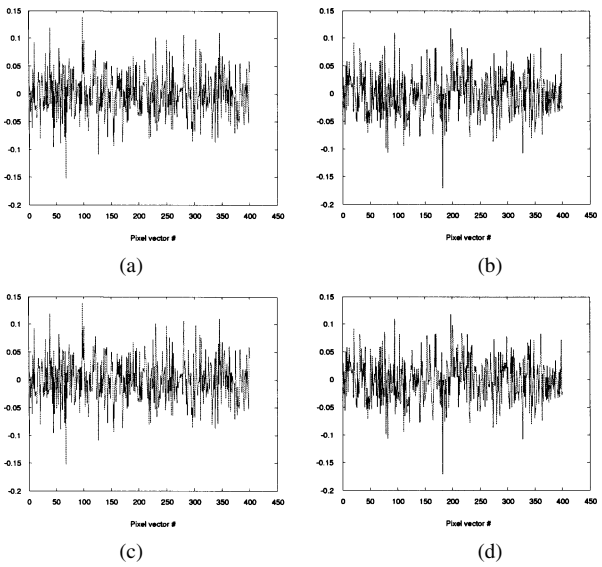

Fig. 6. Detection results of $\delta_{\mathrm{TCIMF}}(\mathbf{r})$. (a) $\mathbf{d}$ = blackbrush, (b) $\mathbf{d}$ = creosote leaves, (c) $\mathbf{d}$ = sagebrush, (d) D = [blackbrush, creosote, leaves, sagebrush].

noted, the $(\mathbf{d}, \mathbf{U})$ model requires the knowledge of the desired target signature $\mathbf{d}$ and the undesired target signature matrix, $\mathbf{U}$. When both the $\mathbf{d}$ and the $\mathbf{U}$ are not available, the OSP must be implemented whatever it can obtain directly from the data. According to (58) and (59), when the knowledge about the $\mathbf{U}$ is not available, the inverse of the sample spectral correlation matrix, $\mathbf{R}^{-1}$ can be used to approximate $P_{\mathbf{U}}^{\perp}$. Moreover, if the knowledge of the $\mathbf{d}$ is further not provided, the only available information that can be used for the OSP is the image pixel $\mathbf{r}$. In this case, the matched signatures used in the OSP must be replaced by the $\mathbf{d}$. So, substituting $\mathbf{r}$ and $\mathbf{R}^{-1}$ for $\mathbf{d}$ and $P_{\mathbf{U}}^{\perp}$ in (11), respectively, results in a new filter, which can be used for anomaly detection. Such a filter is referred to as OSP anomaly detector (OSPAD), denoted by $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and given by

$$\delta_{\mathrm{OSPAD}}(\mathbf{r}) = \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r}. \tag{60}$$

It should be noted that if we replace the $\mathbf{d}$ in the CEM in (41) with the image pixel $\mathbf{r}$, the resulting form would be the constant 1 for all the image pixel $\mathbf{r}$, i.e., $(\mathbf{r}^T\mathbf{R}^{-1}\mathbf{r})^{-1}\mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} = 1$. This is because the CEM performs as an estimator rather than a detector as the OSP does. Since no desired signature $\mathbf{d}$ needs to be estimated, the quantity of $(\mathbf{r}^T\mathbf{R}^{-1}\mathbf{r})^{-1}$ that is included in the CEM is to account for the estimation accuracy varies with the image pixel $\mathbf{r}$. Therefore, the CEM cannot be used to derive for anomaly detection as we did for the OSP in (60). That also explains why the OSP has better generalized properties than the CEM, and the CEM can be considered as partial-knowledge version of the OSP.

Interestingly, if we replace $\mathbf{r}$ and $\mathbf{R}^{-1}$ in the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ with $\mathbf{r} - \boldsymbol{\mu}$ and $\mathbf{K}^{-1}$ where $\boldsymbol{\mu}$ and $\mathbf{K}$ are the sample mean and the sample covariance matrix, the resulting filter $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ turns out to be the well-known anomaly detector, referred to as RX detector $\delta_{\mathrm{RXD}}(\mathbf{r})$ and also known as Mahalanobis distance [25]

$$\delta_{\mathrm{RXD}}(\mathbf{r}) = (\mathbf{r} - \boldsymbol{\mu})^T\mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) \tag{61}$$

that was developed by Reed and Yu [18]. If we once again replace the matched signature $\mathbf{r}^T$ in (60) and (61) with the L-dimensional unity vector $\mathbf{1}^T = (\underbrace{1, 1, \ldots 1}_{L})$, $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ becomes so-called low probability detection (LPD), $\delta_{\mathrm{LPD}}(\mathbf{r})$ in [14] and [29] given by

$$\delta_{\mathrm{LPD}}(\mathbf{r}) = \mathbf{1}^T\mathbf{R}^{-1}\mathbf{r} \tag{62}$$

which was developed in Harsanyi's dissertation [14] and uniform target detector $\delta_{\mathrm{UTD}}(\mathbf{r})$

$$\delta_{\mathrm{UTD}}(\mathbf{r}) = (\mathbf{1} - \boldsymbol{\mu})^T\mathbf{K}^{-1}(\mathbf{r} - \boldsymbol{\mu}) \tag{63}$$

which was derived in [8]. More details about $\delta_{\mathrm{LPD}}(\mathbf{r})$ and $\delta_{\mathrm{UTD}}(\mathbf{r})$ can be found in [8] and [30].

As discussed in Section IV, we may sometimes have partial knowledge about target signatures that are not wanted, such as background. In this case, we may think that removing this knowledge prior to anomaly detection could improve anomaly detectability. As will be explained later, this is not necessarily true.

Following a similar treatment in Section IV, suppose that the knowledge about $\mathbf{U}$ is provided. We can implement the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ in conjunction with undesired signature annihilator $P_{\mathbf{U}}^{\perp}$ in the same way that it is implemented in the $\delta_{\mathrm{OSP}}(\mathbf{r})$ to remove the undesired target signatures before anomaly detection. The resulting detector is called the $P_{\mathbf{U}}^{\perp}$-OSP anomaly detector ($P_{\mathbf{U}}^{\perp}$-OSPAD), $\delta_{\mathrm{OSPAD},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ defined by

$$\delta_{\mathrm{OSPAD},P_{\mathbf{U}}^{\perp}}(\mathbf{r}) = (P_{\mathbf{U}}^{\perp}\mathbf{r})^T\mathbf{R}^{-1}(P_{\mathbf{U}}^{\perp}\mathbf{r}). \tag{64}$$

Similarly, the RXD can be also implemented in conjunction with $P_{\mathbf{U}}^{\perp}$, called $P_{\mathbf{U}}^{\perp}$-RXD and denoted by $\delta_{\mathrm{RXD},P_{\mathbf{U}}^{\perp}}(\mathbf{r})$ as follows:

$$\delta_{\mathrm{RXD},P_{\mathbf{U}}^{\perp}}(\mathbf{r}) = (P_{\mathbf{U}}^{\perp}\mathbf{r} - P_{\mathbf{U}}^{\perp}\boldsymbol{\mu})^T\mathbf{K}^{-1}(P_{\mathbf{U}}^{\perp}\mathbf{r} - P_{\mathbf{U}}^{\perp}\boldsymbol{\mu}). \tag{65}$$

Surprisingly, according to the conducted experiments, the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$, and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ will be shown to perform very closely regardless of whether or not $P_{\mathbf{U}}^{\perp}$ is included in detection. This is because anomaly detectors are generally designed to extract pixels whose signatures spectrally distinct its surroundings rather than suppress signatures. Another reason is that $\mathbf{R}^{-1}$ can be viewed as an approximation of $P_{\mathbf{U}}^{\perp}$. Therefore, an additional inclusion of $P_{\mathbf{U}}^{\perp}$ does not improve the performance of the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$, both of which already perform a similar task to $P_{\mathbf{U}}^{\perp}$ that is carried out by $\mathbf{R}^{-1}$ in (64)–(65). So, in this paper, only experiments for the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ will be presented.

Example 5 (Anomaly Detection): In this example, several experiments were conducted to evaluate the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ specified by (60) and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ specified by (61). The same 401 simulated pixel vectors with added SNR 30 : 1 white Gaussian noise in Example 1 were used to detect the creosote leaves as an anomalous target. Table V tabulates the abundance fractions produced by the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$, and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ in detection of creosote leaves when the pixels of the creosote leaves with abundance 10% were expanded from one pixel (pixel number 200), three pixels (pixel numbers 199, 200, 201), 5 pixels (pixel numbers, 198–202) to 11 pixels (pixel numbers 195–205). As we can see from Table V, the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ performed slightly better

TABLE V
ABUNDANCE FRACTIONS DETECTED BY OSPAD AND RXD IN FIG. 12

	Pixel #	197	198	199	200	201	202	203	LSE
$\delta_{\mathrm{OSPAD}}(\mathbf{r})$	200				0.7139				0.50961
	199-201			0.4941	0.5484	0.5940			0.24411
	198-202		0.4964	0.4215	0.4726	0.5442	0.4700		0.24637
	195-205	0.4252	0.4562	0.3739	0.3957	0.4593	0.4057	0.4602	0.17875
$\delta_{\mathrm{RXD}}(\mathbf{r})$	200				0.7114				0.50616
	199-201			0.4916	0.5459	0.5916			0.24169
	198-202		0.4939	0.4190	0.4701	0.5417	0.4675		0.24391
	195-205	0.4228	0.4537	0.3714	0.3993	0.4568	0.4032	0.4577	0.17666

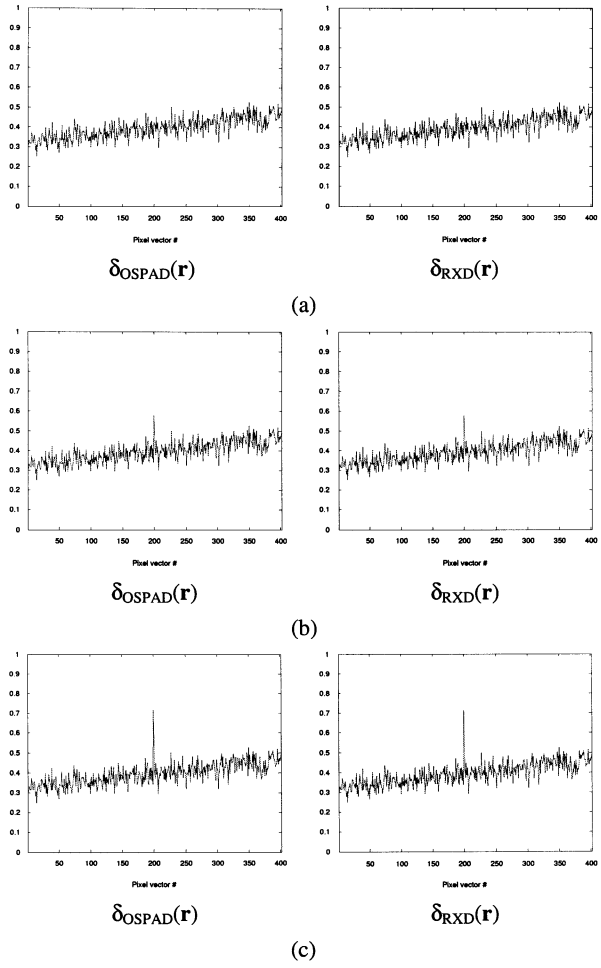

Fig. 7. Detection results of creosote leaves at pixel number 200 by $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and $\delta_{\mathrm{RXD}}(\mathbf{r})$ with various SNRs. (a) SNR 10 : 1. (b) SNR 20 : 1. (c) SNR 30 : 1.

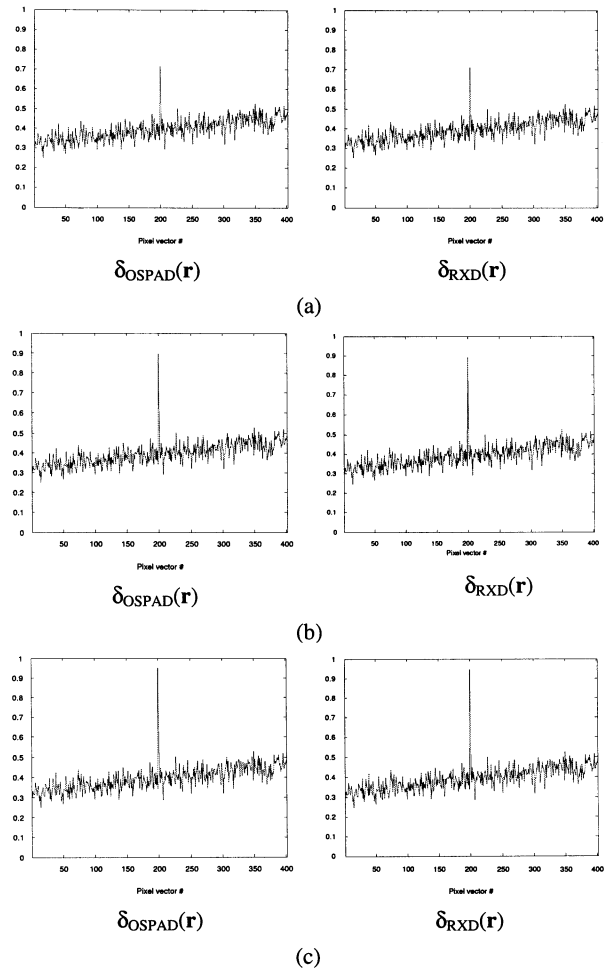

Fig. 8. Detection results of creosote leaves at pixel number 200 by the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ with various abundance fractions. (a) Abundance fraction 10%. (b) Abundance fraction 20%. (c) Abundance fraction 30%.

than the $\delta_{\mathrm{RXD}}(\mathbf{r})$. As a matter of fact, according to our experiments, in order for the creosote leaves to qualify for an anomalous target, the number of pixels should not exceed 3. Figs. 7 and 8, also show how SNR (10 : 1, 20 : 1, 30 : 1) and abundance fractions (10%, 20%, 30%) affected the performance of anomaly detection for the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$, respectively, where the higher the SNR, the better the anomaly detection, and the more the abundance fractions of anomaly, the better the anomaly detection. From Table V and Figs. 7 and 8, both the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ performed comparably in terms of detected abundance fractions.

The next experiment was designed to see how many anomalies could be detected as distinct targets by the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ if the same 401 simulated pixels with added SNR 30 : 1 white Gaussian noise in Example 1 were also used. Table VI tabulates the detection results of the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ where two of three target signatures, blackbrush, creosote leaves, and sagebrush were selected with same abundance 10% at pixel number 100 and pixel number 300. As we can see from Table VI, the results were not good, but the pixel number 300 was always detected. Interestingly, if the three target signatures blackbrush, creosote leaves, and sagebrush were present at

TABLE VI
ABUNDANCE FRACTIONS OF TWO SIGNATURES DETECTED BY $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ AND $\delta_{\mathrm{RXD}}(\mathbf{r})$

	blackbrush	Creosote leaves	sagebrush
$\delta_{\mathrm{OSPAD}}(\mathbf{r})$	0.3319	0.7020	
	0.3523		0.6220
		0.5800	0.5452
$\delta_{\mathrm{RXD}}(\mathbf{r})$	0.3296	0.6995	
	0.3500		0.6195
		0.5775	0.5428

TABLE VII
ABUNDANCE FRACTIONS OF THREE SIGNATURES DETECTED BY $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ AND $\delta_{\mathrm{RXD}}(\mathbf{r})$

	blackbrush	Creosote leaves	sagebrush
$\delta_{\mathrm{OSPAD}}(\mathbf{r})$	0.3202	0.5956	0.5204
$\delta_{\mathrm{RXD}}(\mathbf{r})$	0.3179	0.5931	0.5179

pixel numbers 100, 200, and 300 with the same abundance 10%, Table VII, tabulates the detection results of the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ where both detectors failed to detect these three targets. This implied that among the 401 simulated pixels, the $\delta_{\mathrm{OSPAD}}(\mathbf{r})$ and the $\delta_{\mathrm{RXD}}(\mathbf{r})$ could only detect two types of distinct anomalies, but no more than two. The investigation of the issue of anomaly detection on the size and signatures of anomalies is beyond the scope of this paper. This experiment only provides evidence that anomaly detection cannot be blindly implemented, and some extra care must be taken. However, a detailed study and performance analysis on this issue can be found in [31] and [32].

VI. CONCLUSION

The OSP has become a standard hyperspectral imaging technique [33] that can be used in many versatile applications. Despite the fact that various relationships among the OSP, the CEM, and the RXD have been studied in [8], [16], and [17], this paper is accomplished as a consequence of many interesting results derived from the OSP that were not explored in [8], [16], and [17]. It revisits the OSP from several signal processing perspectives and offer many insights into its design rationale that have not been investigated previously. In particular, it shows that the OSP can be derived from various view points of signal detection, linear discriminant analysis, and parameter estimation where the least squares OSP is essentially equivalent to the least squares linear spectral mixture analysis via the proposed OSP model. It further studies effects of the Gaussian noise and white noise assumptions on the performance of the OSP. Finally, it derives various forms of the OSP when the OSP is provided by different levels of information, where some well-known and popular filters such as the CEM, the TCIMF, and the RX anomaly detector can be considered as members of the OSP family. As a concluding remark, in order to investigate various assumptions, only computer simulations have been used to substantiate the results. Since many experiments conducted based on real hyperspectral images using various forms of the OSP have been reported in the literature and can be also found in [8], image experiments are not included in this paper.

ACKNOWLEDGMENT

The author would like to thank the support received from an NRC senior associateship from the U.S. Army Soldier and Biological Command, Edgewood Chemical and Biological Center (ECBC), and the TRW Foundation. The author also thanks M. Hsueh for generating all figures and tables in this paper and J. C. Harsanyi for providing the AVIRIS data.

REFERENCES

[1] J. R. Jensen, *Introductory Digital Image Processing, A Remote Sensing Perspective*. Upper Saddle River, New Jersey: Prentice-Hall, 1996.
[2] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
[3] S. S. Epp, *Discrete Mathematics with Applications*, 2nd ed. Pacific Grove, CA: Brooks/Cole, 1995.
[4] C.-I Chang and C. Brumbley, "A Kalman filtering approach to multispectral image classification and detection of changes in signature abundance," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 257–268, Jan. 1999.
[5] H. Ren and C.-I Chang, "A generalized orthogonal subspace projection approach to unsupervised multispectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2515–2528, Nov. 2000.
[6] C.-M. Wang, S.-C. Yang, P.-C. Chung, C. S. Lo, C.-I Chang, C. C. Chen, C.-W. Yang, and C. H. Wen, "Orthogonal subspace projection-based approaches to classification of MR image sequences," *Comput. Med. Imag. Graph.*, vol. 25, no. 6, pp. 465–476, Oct. 2001.
[7] C. M. Wang, C. C. Chen, S.-C. Yang, Y.-N. Chung, P. C. Chung, C. W. Yang, and C.-I Chang, "An unsupervised orthogonal subspace projection approach to MR image classification MR images for classification," *Opt. Eng.*, vol. 41, no. 7, pp. 1546–1557, Jul. 2002.
[8] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Orlando, FL: Kluwer Academic, 2003.
[9] J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1045–1046, Jul. 1996.
[10] C.-I Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 3, pp. 1030–1032, May 1998.
[11] 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 Sens.*, vol. 36, no. 3, pp. 898–912, May 1998.
[12] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 2nd ed. New York: Springer-Verlag, 1999.
[13] C.-I Chang, Q. Du, S.-S. Chiang, D. Heinz, and I. W. Ginsberg, "Unsupervised target subpixel detection in hyperspectral imagery," in *Proc. SPIE Conf. Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, Orlando, FL, Apr., 20–24 2001, pp. 370–379.
[14] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Dept. Elect. Eng., Univ. Maryland Baltimore County, Baltimore, MD, Aug. 1993.
[15] H. Ren and C.-I Chang, "Target-constrained interference-minimized approach to subpixel target detection for hyperspectral imagery," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, Dec. 2000.
[16] C.-I Chang, "Relationship among orthogonal subspace projection, constrained energy minimization and RX-algorithm," in *Proc. SPIE Conf. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII, SPIE*, vol. 4725, Orlando, FL, Apr., 1–5 2002.
[17] Q. Du, H. Ren, and C.-I Chang, "A comparative study for orthogonal subspace projection and constrained energy minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1525–1529, Jun. 2003.
[18] S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
[19] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[20] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York: Oxford Univ. Press, 1995.
[21] L. L. Scharf, *Statistical Signal Processing*. Reading, MA: Addison-Wesley, 1991.

[22] T. M. Tu, C.-H. Chen, and C.-I Chang, "*A posteriori* least-squares orthogonal subspace projection approach to weak signature extraction and detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 127–139, Jan. 1997.
[23] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
[24] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing mdoel," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.
[25] K. Fukunaga, *Statistical Pattern Recognition*, 2nd ed. Orlando, FL: Academic, 1990.
[26] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, 1972.
[27] C.-I Chang, H. Ren, and S. S. Chiang, "Real-time processing algorithms for target detection and classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 760–768, Apr. 2001.
[28] C.-I Chang, "An information theoretic-based approach to spectral variability, similarity, and discriminability for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.
[29] C.-I Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.
[30] J. Wang and C.-I Chang, "A uniform projection-based unsupervised detection and classification for hyperspectral imagery," in *Proc. IGARSS*, Anchorage, AK, Sep. 20–24, 2004.
[31] M. Huseh and C.-I Chang, "Adaptive causal anomaly detection for hyperspectral imagery," in *Proc. IGARSS*, Anchorage, AK, Sep. 20–24, 2004.
[32] M. Hsueh, "Adaptive causal anomaly detection," M.S. thesis, Dept. Comput. Sci. Electrical Engineering, Univ. Maryland Baltimore County, Baltimore, MD, Aug. 2004.
[33] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 2nd ed. San Diego, CA: Academic, 1997, pp. 470–471.

Chein-I Chang (S'81–M'87–SM'92) received the B.S. degree from Soochow University, Taipei, Taiwan, R.O.C., in 1973, the M.S. degree from the Institute of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 1975, and the M.A. degree from the State University of New York, Stony Brook, in 1977, all in mathematics. He received the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana-Champaign in 1982 and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, in 1987.

He has been with the University of Maryland Baltimore County (UMBC), Baltimore, since 1987, as a Visiting Assistant Professor from January 1987 to August 1987, Assistant Professor from 1987 to 1993, Associate Professor from 1993 to 2001, and Professor in the Department of Computer Science and Electrical Engineering since 2001. He was a Visiting Research Specialist in the Institute of Information Engineering at the National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995. He has three patents on automatic pattern recognition and several pending patents on image processing techniques for hyperspectral imaging and detection of microcalcifications. His research interests include automatic target recognition, multispectral/hyperspectral image processing, medical imaging, information theory and coding, signal detection and estimation, and neural networks. He is the author of a book *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Norwell, MA: Kluwer). He is on the editorial board and was the Guest Editor of a special issue on telemedicine and applications of the *Journal of High Speed Networks*.

Dr. Chang received a National Research Council Senior Research Associateship Award from 2002 to 2003 at the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, MD. He is an Associate Editor in the area of hyperspectral signal processing for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is a Fellow of SPIE and a member of Phi Kappa Phi and Eta Kappa Nu.