

A *Posteriori* Least Squares Orthogonal Subspace Projection Approach to Desired Signature Extraction and Detection

Te-Ming Tu, Chin-Hsing Chen, and Chein-I Chang, *Senior Member, IEEE*

Abstract—One of the primary goals of imaging spectrometry in earth remote sensing applications is to determine identities and abundances of surface materials. In a recent study, an orthogonal subspace projection (OSP) was proposed for image classification. However, it was developed for an *a priori* linear spectral mixture model which did not take advantage of *a posteriori* knowledge of observations. In this paper, an *a posteriori* least squares orthogonal subspace projection (LSOSP) derived from OSP is presented on the basis of an *a posteriori* model so that the abundances of signatures can be estimated through observations rather than assumed to be known as in the *a priori* model. In order to evaluate the OSP and LSOSP approaches, a Neyman–Pearson detection theory is developed where a receiver operating characteristic (ROC) curve is used for performance analysis. In particular, a locally optimal Neyman–Pearson’s detector is also designed for the case where the global abundance is very small with energy close to zero a case to which both LSOSP and OSP cannot be applied. It is shown through computer simulations that the presented LSOSP approach significantly improves the performance of OSP.

Index Terms—*A priori* (*pr*), *A posteriori* (*ps*), detection power, false alarm probability, least-squares estimate, Neyman–Pearson (N–P) detectors, orthogonal subspace projection, ROC curve.

I. INTRODUCTION

GOVERNMENT agencies, both civilian and military, are placing a higher reliance upon remotely sensed image data as an information source crucial to decision making and planning. The fields of geology, geography, and agriculture have a strong history of exploitation of multispectral image data for analysis and classification of earth surface attributes. The use of high spatial resolution airborne and satellite sensors improves the capability of identification and discrimination of endmembers (materials), particularly for endmembers with very similar spectral signatures. Hyperspectral imaging spectrometry is a new technology for earth remote sensing applications from airborne and spaceborne platforms. One major advantage of hyperspectral imagery over multispectral imagery is that the former images a scene using as many as 200 contiguous bands as opposed to the latter

using only four–seven discrete bands. As a result, hyperspectral data permit the expansion of detection and classification activities to targets previously unresolved in multispectral images. More precisely, a hyperspectral image acquired by an imaging spectrometer such as airborne visible infrared imaging spectrometer (AVIRIS) [2]–[3] using contiguous 10 nm wide spectral bands of a spectral coverage ranging from 0.4 to 2.5 μm produces sufficient resolution for the direct identification of endmembers with diagnostic features, whereas a multispectral image such as acquired by the multispectral scanner (MSS) and thematic mapper (TM) cannot resolve these features because their spectral bandwidths are 100–200 nm and not contiguous.

Despite the vast information provided by imaging spectrometers, a major problem which arises in hyperspectral image analysis is that scene pixels are generally mixed linearly or nonlinearly by component surface endmembers or spectral reflectances of endmembers [4]–[6]. In a recent study [7], [8], an orthogonal subspace projection (OSP) approach was proposed as a classification technique for image pixels which are linearly mixed. The idea was to apply an OSP classifier to eliminate all unwanted endmembers and interferences within a pixel, then use a matched filter to extract the desired endmember present in that pixel. Therefore, OSP can be viewed as a subpixel scale version of the simultaneous diagonalization filter developed in [9]. Although there is no explicit assumption made in [7] on the knowledge about the signatures (a signature will be referred to as the spectral reflectance of an endmember hereafter), OSP was developed based on an understanding that the signatures and abundances were known and completely described in an *a priori* (abbreviated by *pr* hereafter) spectral mixture model specified by (3) below. However, in earth remote sensing problems such as crop production and damage assessment, estimating the fractions of endmembers resident in a mixed pixel can be very important and useful. Fortunately, the OSP can be still applied to this case where the true abundances are replaced by their estimates [8]. This was actually done for experiments in [7] where the true endmembers were directly extracted from the images themselves.

In this paper, we present an *a posteriori* (abbreviated by *ps* hereafter) least squares orthogonal subspace projection (LSOSP) approach from a signal processing point of view. In signal processing, a standard signal model is generally described by

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

Manuscript received March 17, 1995; revised April 16, 1996. This work was supported by the National Science Council in Taiwan for Grant 84-2213-E-006-086 (Chang) and Grant 85-2213-E-006-066 (Tu and Chen).

T.-M. Tu and C.-H. Chen are with the Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan, R.O.C.

C.-I. Chang is with the Remote Sensing Signal and Image Processing Laboratory, the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21228 USA.

Publisher Item Identifier S 0196-2892(97)00974-1.

where \mathbf{r} is an observation, \mathbf{x} is the signal of interest and \mathbf{n} is additive noise. In many practical applications, what is most useful is the observation from which a certain amount of information can be retrieved to be used for prediction. So, if we can predict the signal on the basis of the observation prior to data processing and substitute it for the true signal in (1), the difference between the predicted signal, denoted by $\hat{\mathbf{x}}$ and the observation \mathbf{r} is the one that needs to be processed and is usually referred to as the prediction error which is caused by an inaccurate prediction plus the noise. Let $\hat{\mathbf{n}}$ denote the resulting prediction error, i.e., $\hat{\mathbf{n}} = \mathbf{r} - \hat{\mathbf{x}}$, (1) can be reexpressed as

$$\mathbf{r} = \hat{\mathbf{x}} + (\mathbf{r} - \hat{\mathbf{x}}) = \hat{\mathbf{x}} + \hat{\mathbf{n}}. \quad (2)$$

If the prediction can be made perfectly in (2), $\hat{\mathbf{n}}$ will be zero, but this case does not occur in general, particularly not in real world applications in remote sensing due to a variety of unknown noise and unexpected interferences such as effects from the atmosphere. In such a case, a goodness-of-fit prediction technique is necessary. This can be accomplished by designing good estimators in some sense of optimality. If we can estimate the signal completely from the observation, the estimation error (prediction error) $\hat{\mathbf{n}}$ must be orthogonal to the estimated signal $\hat{\mathbf{x}}$. Then $\hat{\mathbf{n}}$ will be completely unpredictable and contains no information which can be retrieved from the observation. As a consequence, $\hat{\mathbf{n}}$ can be modeled by a random noise. On the other hand, if $\hat{\mathbf{n}}$ is not orthogonal to $\hat{\mathbf{x}}$, it means that $\hat{\mathbf{n}}$ is still correlated with $\hat{\mathbf{x}}$ and both $\hat{\mathbf{n}}$ and $\hat{\mathbf{x}}$ must share same information to some extent. This correlation further implies that the estimator used for $\hat{\mathbf{x}}$ is not optimal and a better estimator may exist. Accordingly, the model given by (1) will be called a *pr* signal model, while (2) will be referred to as a *ps* signal model. The $\hat{\mathbf{x}}$ can be thought of as a *a posteriori* estimate. It should be noted that a *pr* signal model is useful when the observation process is not available in the beginning of data processing. Using a statistical model as a *pr* signal model is a common practice in the signal processing community. However, the importance of a *pr* model will considerably diminish if the number of observed samples is increased. In this situation, a *ps* model begins to show its dominance and tends to replace the *pr* model. As the observation process continues on, the *ps* model may eventually take over the *pr* model. Consequently, methods using a *ps* model will subsequently improve those based on a *pr* model. This is the main focus to be addressed in our paper.

The objective of the proposed LSOSP is to take the above approach and produce an optimal estimator for signature abundances of interest based on minimizing the least squares error. Although LSOSP is derived from OSP, their underlying concepts are fundamentally different as described previously because the former is based on a *ps* linear spectral mixture model, while the latter uses a *pr* linear spectral mixture model. LSOSP is used to find the best linear least squares estimate of the abundance of a desired signature. It first decomposes the observation space into a signature space and a noise space, then projects the observation into the signature space. As a result of this projection, the noise effects are greatly reduced and the original signal-to-noise ratio (SNR) is improved. In the following stage, OSP is used to eliminate undesired signatures

in the signature space to reduce data dimensionality. Since LSOSP operates now on the noise-reduced observations, it significantly improves on the OSP approach in [7], which was developed to be applied to the observation space.

In order to evaluate the performance of OSP and LSOSP, a signal detection model making use of the Neyman–Pearson theory is developed where the receiver operating characteristic (ROC) curve is used for performance analysis with the false alarm probability measured by the estimation error (or prediction error) and the detection power represented by the accuracy of the estimation. Two advantages can be gained by the LSOSP approach. One is that the SNR is improved by a factor of the ratio of the data dimensionality to that of the signature space. This is particularly significant for remote sensing imagery where the data dimensionality is generally much larger than that of the signature space. The second advantage is that noise effects are reduced substantially by projecting the observation into the signature space whereby the Neyman–Pearson’s detection power is increased significantly. Computer simulations show that for a desired signature with abundance less than 5%, LSOSP improves OSP performance. For the case where a pixel contains a very small fraction of a desired signature, and where OSP and LSOSP are not directly applicable, the locally optimal Neyman–Pearson detector developed in [11] and [12] is adopted for weak signature detection.

This paper is organized as follows. Section II is the problem statement and formulation where a linear spectral mixture model is described. Section III presents a least squares estimation approach to convert a *pr* linear spectral mixture model to *ps* linear spectral mixture. Section IV describes LSOSP approach and applies it to the *ps* model. In order to evaluate LSOSP and OSP, Section V derives Neyman–Pearson (N–P) theories for the *pr* and *ps* models as well as a locally optimal N–P theory for weak signature detection. Section VI presents simulation results for the performance analysis of LSOSP and OSP approaches. Section VII includes a brief conclusion.

II. PROBLEM STATEMENT AND FORMULATION

In hyperspectral image analysis, the spatial coverage of each pixel generally encompasses multiple materials. In such a case, the spectral signature of the pixel is an admixture of the spectral signatures of these endmembers and the identification of a target signature of interest within such a mixed pixel requires a detection technique at subpixel scale. In image processing, many techniques have been developed for pixel-by-pixel processing, but not particularly for subpixel analysis. As a result, most of them are not immediately applicable. For example, the matched filter/correlation filter widely used in communications/signal processing [13] cannot be directly used in mixture analysis. In this paper, we only consider the linear spectral mixture model described below [4]–[8], [14], [15].

A. Linear Mixture Model for Multispectral/Hyperspectral Images

Let \mathbf{r}_i be a $l \times l$ column vector and denote the i th pixel in a hyperspectral image where l is the number of bands. Assume that M is a $l \times p$ matrix and made up of p endmembers

denoted by $(\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p)$ where $l > p$ and \mathbf{m}_j is an $l \times 1$ column vector representing the spectral reflectance of the j th endmember. Let α_i be a $p \times 1$ nonrandom column vector given by $(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ip})^T$ and α_{ij} denote the fraction of the j th endmember present in pixel \mathbf{r}_i . A linear spectral mixture model similar to (1) for a hyperspectral image pixel \mathbf{r}_i assumes that \mathbf{r}_i is a linear combination of p endmembers with the weight coefficients designated by abundance vector $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ip})^T$ plus noise \mathbf{n}_i . More specifically,

$$\mathbf{r}_i = M\alpha_i + \mathbf{n}_i \quad (3)$$

where \mathbf{n}_i is an $l \times 1$ column vector representing additive white Gaussian noise with zero mean and covariance matrix $\sigma^2 I$, and I is the $l \times l$ identity matrix.

III. A POSTERIORI MODEL

When the model (3) is used, it generally assumes that the abundance α_i is known *a priori*. This is a Bayes approach which has important theoretical interest. However, in many real world applications, $\{\alpha_{ij}\}_{j=1}^p$ are generally not known and must be estimated from the observation vector \mathbf{r}_i . In doing so, one approach is to use a least squares error estimation given in [10] to convert the *pr* model (3) to *ps* model similar to (2).

Let $\hat{\alpha}(\mathbf{r})$ be the estimate of α based on the observation vector \mathbf{r} where the subscript i is suppressed and the ‘‘hat’’ symbol, ‘‘^’’ is used to indicate an ‘‘estimate.’’ Then the least squares error is given by

$$\begin{aligned} E &= (\mathbf{r} - M\hat{\alpha}(\mathbf{r}))^T (\mathbf{r} - M\hat{\alpha}(\mathbf{r})) \\ &= \text{trace}[(\mathbf{r} - M\hat{\alpha}(\mathbf{r}))(\mathbf{r} - M\hat{\alpha}(\mathbf{r}))^T]. \end{aligned} \quad (4)$$

Differentiating E with respect to $\hat{\alpha}(\mathbf{r})$ and setting to zero yields the optimal least squares estimate of α ,

$$\hat{\alpha}_{LS}(\mathbf{r}) = M^\# \mathbf{r} \quad (5)$$

where $M^\# = (M^T M)^{-1} M^T$ is the pseudo inverse of M . We would like to point out that (5) was also derived in [8].

As a result, the model in (3) can be equivalently stated as the following model

$$\begin{aligned} \mathbf{r} &= M\alpha + \mathbf{n} \\ &= M\hat{\alpha}_{LS}(\mathbf{r}) + \hat{\mathbf{n}}_{LS}(\mathbf{r}) \end{aligned} \quad (6)$$

where

$$\hat{\mathbf{n}}_{LS}(\mathbf{r}) = \mathbf{r} - M\hat{\alpha}_{LS}(\mathbf{r}) = M(\alpha - \hat{\alpha}_{LS}(\mathbf{r})) + \mathbf{n}. \quad (7)$$

Since all estimates are made based on the observation vector \mathbf{r} , \mathbf{r} is included in the above notations of estimates to indicate that they are functions of the observation vector \mathbf{r} . However, unless some cases need to be specified, \mathbf{r} will be omitted throughout the rest of the paper to simplify notations. According to [10], the model described by (3) is called Bayes model or *pr* model because it requires *pr* information. On the other hand, if we let $\hat{\mathbf{x}}_{LS}(\mathbf{r}) = M\hat{\alpha}_{LS}(\mathbf{r})$, the model in (6) is equivalent to that in (2), and thus the model (6) will be referred to as *ps* model because no knowledge of α is required, only the observation vector \mathbf{r} . The relationship between the *pr* model and the *ps* model is illustrated in Fig. 1 where a *pr* model and a

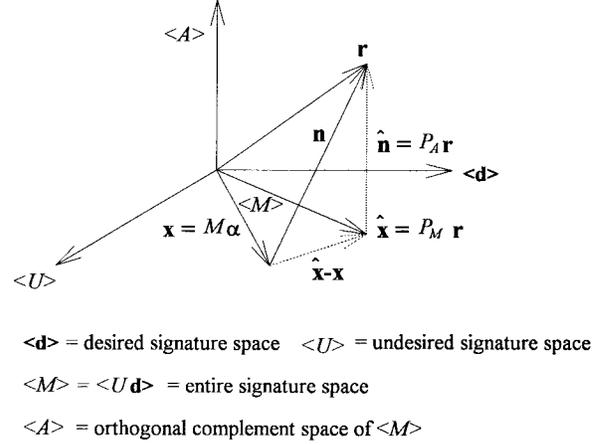


Fig. 1. Relationship between *pr* model and *ps* model.

ps model are represented by two solid vectors and two dashed vectors respectively. It is also worth noting that since the least squares estimate $\hat{\mathbf{x}}_{LS}$ is optimal, as mentioned previously in the introduction, $\hat{\mathbf{x}}_{LS}$ and $\hat{\mathbf{n}}_{LS}$ must be orthogonal.

Let

$$P_M = M M^\# \quad (8)$$

$$P_A = (I - P_M) \quad (9)$$

where I is an identity matrix.

From (5), (8) and (9) $\hat{\alpha}_{LS}$ and $\hat{\mathbf{n}}_{LS}$ can be expressed by $\hat{\alpha}_{LS} = M^\# \mathbf{r}$ and $\hat{\mathbf{n}}_{LS} = P_A \mathbf{r}$, respectively. The P_M and P_A defined by (8), and (9), are projection operators which project the observation vector \mathbf{r} into the signature space and noise space respectively. It is shown in [10] that these two operators are symmetric and idempotent. More importantly, they are orthogonal. In other words, the signature space and noise space projected by P_M and P_A are orthogonal and one is the orthogonal complement to the other. More detail of this can be found in [10].

Now applying P_M to (3) results in

$$\begin{aligned} P_M \mathbf{r} &= M\alpha + P_M \mathbf{n} \\ &= M\alpha + M(\hat{\alpha}_{LS} - \alpha) + P_M \hat{\mathbf{n}}_{LS} \quad (\text{due to (7)}) \\ &= M\alpha + M\epsilon \end{aligned} \quad (10)$$

$$= \hat{\mathbf{x}}_{LS} \quad (\text{since } \hat{\mathbf{x}}_{LS} = M\hat{\alpha}_{LS} \text{ and } P_M \hat{\mathbf{n}}_{LS} = 0) \quad (11)$$

where $\epsilon = \hat{\alpha}_{LS} - \alpha$ is the estimation error term and P_M annihilates $\hat{\mathbf{n}}_{LS}$ because $P_M P_A = 0$ and $P_M \hat{\mathbf{n}}_{LS} = P_M P_A \mathbf{n} = 0$. Furthermore, from [10] $\epsilon = \hat{\alpha}_{LS} - \alpha$ is unbiased and its covariance matrix is given by

$$\begin{aligned} E[\epsilon \epsilon^T] &= E[(\hat{\alpha}_{LS}(\mathbf{r}) - \alpha)(\hat{\alpha}_{LS}(\mathbf{r}) - \alpha)^T] \\ &= \sigma^2 (M^T M)^{-1}. \end{aligned} \quad (12)$$

Since \mathbf{n} is Gaussian, so is $\hat{\mathbf{x}}_{LS}$ from (11) with the probability distribution given by $N(P_M M \alpha, \sigma^2 P_M) = N(M \alpha, \sigma^2 P_M)$ where $P_M M = M$. Similarly, from (9) and (11) the estimate of the noise, $\hat{\mathbf{n}}_{LS}$ has the Gaussian probability distribution

$$N(P_A M \alpha, \sigma^2 P_A) = N(0, \sigma^2 P_A), \quad (13)$$

It is important to notice that the estimator $\hat{\mathbf{x}}_{LS}$ is unbiased, i.e., $E[\hat{\mathbf{x}}_{LS}(\mathbf{r})] = M\boldsymbol{\alpha}$. In particular, the variance of $\hat{\mathbf{x}}_{LS}$ is given by

$$\begin{aligned} \text{var}(\hat{\mathbf{x}}_{LS}) &= E[(\hat{\mathbf{x}}_{LS}(\mathbf{r}) - M\boldsymbol{\alpha})^T(\hat{\mathbf{x}}_{LS}(\mathbf{r}) - M\boldsymbol{\alpha})] \\ &= \text{trace} E[\hat{\mathbf{x}}_{LS}(\mathbf{r}) - M\boldsymbol{\alpha})(\hat{\mathbf{x}}_{LS}(\mathbf{r}) - M\boldsymbol{\alpha})^T] \\ &= \sigma^2 p \end{aligned} \quad (14)$$

where p is the dimension of the signature space.

If we define the SNR for a random vector V as

$$\text{SNR} = \frac{[\text{mean}(\mathbf{V})]^T[\text{mean}(\mathbf{V})]}{\text{var}(\mathbf{V})} \quad (15)$$

then the SNR for \mathbf{r} in the pr model of (3) is given by

$$\text{SNR}_{pr} = \frac{[M\boldsymbol{\alpha}]^T[M\boldsymbol{\alpha}]}{\text{trace}[\sigma^2 I]} = \frac{[M\boldsymbol{\alpha}]^T[M\boldsymbol{\alpha}]}{\sigma^2 \cdot l} \quad (16)$$

where l is the dimension of the observation space [10; pp. 379–380]. Similarly, the SNR for the estimator, $\hat{\mathbf{x}}_{LS}$ in the ps model is given by

$$\text{SNR}_{ps} = \frac{[M\boldsymbol{\alpha}]^T[M\boldsymbol{\alpha}]}{\text{trace}[\sigma^2 P_M]} = \frac{[M\boldsymbol{\alpha}]^T[M\boldsymbol{\alpha}]}{\sigma^2 \cdot p} \quad (17)$$

where the last equality follows from (14).

As a result, the ratio of the SNR_{ps} to SNR_{pr} can be obtained by

$$\frac{\text{SNR}_{ps}}{\text{SNR}_{pr}} = \frac{[M\boldsymbol{\alpha}]^T[M\boldsymbol{\alpha}]/\sigma^2 \cdot p}{[M\boldsymbol{\alpha}]^T[M\boldsymbol{\alpha}]/\sigma^2 \cdot l} = \frac{l}{p}. \quad (18)$$

It is worth noting that the result of (18) is significant. It simply says that the ps model can produce a greater improvement upon SNR than that of pr model, provided that $l > p$. This is certainly true for remote sensing imagery in which the data dimensionality is usually much larger than that of the signature space.

IV. LEAST SQUARES SUBSPACE PROJECTION

Despite the SNR improvement by an amount of l/p by (18) via converting from a pr model to a ps model, the data dimensionality was not reduced by the ps projection operator P_M . However, this can be achieved by the OSP method proposed by Harsanyi and Chang in [7] to simultaneously reduce data dimensionality and suppress unwanted or interfering signatures while enhancing the desired signature.

In order to illustrate the above idea, we rewrite $M\boldsymbol{\alpha}$ in the following form

$$M\boldsymbol{\alpha} = \mathbf{d}\alpha_p + U\boldsymbol{\gamma} \quad (19)$$

where \mathbf{d} is the desired signature assumed without loss of generality to be the column vector \mathbf{m}_p , U is an $l \times (p-1)$ matrix given by $U = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{p-1})$, and $\boldsymbol{\gamma}$ is a vector which contain the first $(p-1)$ components of $\boldsymbol{\alpha}$, $\boldsymbol{\gamma} = (\alpha_1, \alpha_2, \dots, \alpha_{p-1})^T$. Therefore, the ps model (4) can be rewritten as follows

$$\begin{aligned} \mathbf{r} &= M\hat{\boldsymbol{\alpha}}_{LS} + \hat{\mathbf{n}}_{LS} \\ &= \mathbf{d}\hat{\alpha}_p + U\hat{\boldsymbol{\gamma}}_{LS} + \hat{\mathbf{n}}_{LS} \end{aligned} \quad (20)$$

where the notation $\hat{\alpha}_p$ is used to simplify the notation $\hat{\alpha}_{LS,p}(\mathbf{r})$ and indicates the estimate of the abundance of the p -th signature. It is worth noting that the estimate of the noise, $\hat{\mathbf{n}}_{LS}$ is a combination of the signature estimation error $\boldsymbol{\varepsilon}$ and the noise projection of \mathbf{n} into $\langle M \rangle$.

So, the first stage is to eliminate the uninteresting signatures which are contained in U , then followed by a second stage using a matched filter to pick up the desired signature \mathbf{d} . The suppression of undesired signals U is accomplished by an OSP operator P_{OSP} given by

$$P_{\text{OSP}} = (I - UU^\#) \quad \text{where } U^\# \text{ is the pseudo inverse of } U \quad (21)$$

which maps the desired signature \mathbf{d} into a space orthogonal to the space spanned by the uninteresting signatures in U . In other words, P_{OSP} annihilates all undesired signatures.

Applying P_{OSP} to (10) renders

$$\begin{aligned} P_{\text{OSP}}P_M\mathbf{r} &= P_{\text{OSP}}P_M(M\boldsymbol{\alpha} + M\boldsymbol{\varepsilon}) \\ &= P_{\text{OSP}}P_M(\mathbf{d}\alpha_p + U\boldsymbol{\gamma}) + P_{\text{OSP}}P_MM\boldsymbol{\varepsilon} \\ &= P_MP_{\text{OSP}}(\mathbf{d}\alpha_p + U\boldsymbol{\gamma}) + P_MP_{\text{OSP}}M\boldsymbol{\varepsilon} \end{aligned} \quad (22)$$

It can be shown that $\text{rank}(P_{\text{OSP}}P_M) = \text{rank}(P_MP_{\text{OSP}})$ and $P_{\text{OSP}}P_M = P_MP_{\text{OSP}}$. In addition, $P_{\text{OSP}}P_M$ is symmetric and idempotent.

In order to find a vector which maximizes the SNR defined by (15), we apply \mathbf{x} to (22) and obtain

$$\mathbf{x}^T P_{\text{OSP}}P_M\mathbf{r} = \mathbf{x}^T P_{\text{OSP}}P_M\mathbf{d}\alpha_p + \mathbf{x}^T P_{\text{OSP}}M\boldsymbol{\varepsilon} \quad (23)$$

where $\mathbf{x}^T P_{\text{OSP}}P_MU\boldsymbol{\gamma} = \mathbf{0}$ due to $P_{\text{OSP}}P_MU\boldsymbol{\gamma} = P_MP_{\text{OSP}}U\boldsymbol{\gamma} = \mathbf{0}$.

The $\text{SNR}_{\hat{\alpha}_p}$ is given by

$$\begin{aligned} \text{SNR}_{\hat{\alpha}_p} &= \frac{(\mathbf{x}^T P_{\text{OSP}}P_M\mathbf{d})\alpha_p^2(\mathbf{d}^T P_M^T P_{\text{OSP}}^T \mathbf{x})}{(\mathbf{x}^T P_{\text{OSP}}P_MM)E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T](M^T P_M^T P_{\text{OSP}}^T \mathbf{x})} \\ &= \frac{\alpha_p^2 \mathbf{x}^T P_{\text{OSP}}P_M\mathbf{d}\mathbf{d}^T P_M^T P_{\text{OSP}}^T \mathbf{x}}{\sigma^2 \mathbf{x}^T P_{\text{OSP}}P_MM(M^T M)^{-1}M^T P_M^T P_{\text{OSP}}^T \mathbf{x}} \\ &= \frac{\alpha_p^2 \mathbf{x}^T P_{\text{OSP}}P_M\mathbf{d}\mathbf{d}^T P_M^T P_{\text{OSP}}^T \mathbf{x}}{\sigma^2 \mathbf{x}^T P_{\text{OSP}}P_MP_M^T P_{\text{OSP}}^T \mathbf{x}} \end{aligned} \quad (24)$$

where the second equality of (24) is true from (12).

Maximizing (24) turns out to be a generalized eigenvalue problem [16]–[17]

$$P_{\text{OSP}}P_M\mathbf{d}\mathbf{d}^T P_M^T P_{\text{OSP}}^T \mathbf{x} = \lambda P_{\text{OSP}}P_MP_M^T P_{\text{OSP}}^T \mathbf{x}. \quad (25)$$

The solution to (25) is given by

$$\mathbf{x} = \kappa\mathbf{d}. \quad (26)$$

Without loss of generality, let $\kappa = 1$. The maximum eigenvalue $\lambda_{\hat{\alpha}_p, \text{max}}$ is obtained by

$$\begin{aligned} \lambda_{\hat{\alpha}_p, \text{max}} &= \frac{\alpha_p^2 \mathbf{d}^T P_{\text{OSP}}P_M\mathbf{d}\mathbf{d}^T P_M^T P_{\text{OSP}}^T \mathbf{d}}{\mathbf{d}^T P_{\text{OSP}}P_MP_M^T P_{\text{OSP}}^T \mathbf{d}} \\ &= \frac{\alpha_p^2}{\sigma^2} \mathbf{d}^T P_{\text{OSP}}P_M\mathbf{d} \end{aligned} \quad (27)$$

where the second equality follows from $P_{\text{OSP}}P_MP_M^T P_{\text{OSP}}^T = P_{\text{OSP}}P_M$.

Comparing (27) to the SNR given in [7],

$$\lambda_{\alpha_p, \max} = \frac{\alpha_p^2 \mathbf{d}^T P_{\text{OSP}} \mathbf{d} \mathbf{d}^T P_{\text{OSP}}^T \mathbf{d}}{\sigma^2 \mathbf{d}^T P_{\text{OSP}} P_{\text{OSP}}^T \mathbf{d}} = \frac{\alpha_p^2}{\sigma^2} \mathbf{d}^T P_{\text{OSP}} \mathbf{d}. \quad (28)$$

Equation (27) is obtained by replacing P_{OSP} in (28) by $P_{\text{OSP}} P_M$ where the ps projector P_M is introduced in (27) to account for the estimation of α_p . It is easy to see that $P_{\text{OSP}} P_M \mathbf{d} = P_M \mathbf{d}$ and $\lambda_{\hat{\alpha}_p, \max} = \lambda_{\alpha_p, \max}$ which implies that the ps least squares projector P_M does not alter the maximum eigenvalue of (25).

The operation solving equation (25) is called the matched filter, denoted by $\mathcal{M}_{\mathbf{d}}$, a linear filter widely used in communication systems which is an inner product of a vector \mathbf{d} with an another vector \mathbf{v} , i.e.,

$$\mathcal{M}_{\mathbf{d}}(\mathbf{v}) = \langle \mathbf{d}, \mathbf{v} \rangle = \mathbf{d}^T \mathbf{v}. \quad (29)$$

From (27) and (28) the matched filter designed by (29) is exactly the same as one obtained in [7].

By coupling the ps projector P_M given by (8) and the OSP operator P_{OSP} by (21) with the matched filter $\mathcal{M}_{\mathbf{d}}$ by (29), the final desired linear operator, $P_{\mathbf{q}}$ can be derived by

$$\begin{aligned} P_{\mathbf{q}} &= \mathcal{M}_{\mathbf{d}} P_{\text{OSP}} P_M, \mathbf{q} = \mathbf{d}^T (I - UU^{\#}) MM^{\#} \quad \text{and} \\ P_{\mathbf{q}} \mathbf{r} &= \mathbf{q}^T \mathbf{r} = [\mathbf{d}^T (I - UU^{\#}) MM^{\#}] \mathbf{r}. \end{aligned} \quad (30)$$

The operator $P_{\mathbf{q}}$ given by (30) achieves a, b and c SNR improvement by making use of P_M b) undesired signal removal and desired signal enhancement by OSP operator P_{OSP} and c) data dimensionality reduction by the matched filter $\mathcal{M}_{\mathbf{d}}$.

As a matter of fact, the estimate of $\alpha_p, \hat{\alpha}_p$ can be obtained from (20) and (30) as follows

$$\begin{aligned} P_{\mathbf{q}} \mathbf{r} &= \mathbf{d}^T P_{\text{OSP}} P_M \mathbf{d} \hat{\alpha}_p = \mathbf{d}^T P_{\text{OSP}} \mathbf{d} \hat{\alpha}_p \quad (\text{due to (20)}) \\ \Rightarrow \hat{\alpha}_p &= \frac{P_{\mathbf{q}} \mathbf{r}}{\mathbf{d}^T P_{\text{OSP}} \mathbf{d}} = \frac{\mathbf{d}^T P_{\text{OSP}} P_M \mathbf{r}}{\mathbf{d}^T P_{\text{OSP}} \mathbf{d}} \\ \Rightarrow \hat{\alpha}_p &= \alpha_p + \frac{\mathbf{d}^T P_{\text{OSP}} P_M \mathbf{n}}{\mathbf{d}^T P_{\text{OSP}} \mathbf{d}} \quad (\text{due to (19)}). \end{aligned} \quad (31)$$

From (27), (28), and (31), OSP and LSOSP are equivalent to the spectral linear unmixing approach specified by (3)–(5). It is worth noting that a similar result was also derived in [8].

V. NEYMAN-PEARSON DETECTION THEORY

In the OSP method [7], the fractions of abundance of spectral signatures $\{\alpha_1, \dots, \alpha_p\}$ are assumed to be known and constant. However, in most of real cases, $\{\alpha_1, \dots, \alpha_p\}$ are unknown. Thus, these $\{\alpha_1, \dots, \alpha_p\}$ need to be estimated prior to application of the OSP method. Section III suggested the LSOSP method to use a ps signature projector to produce $\hat{\alpha}$, the best least squares estimate of α based on observations, to replace the true α in the OSP method. In this section, we develop a N–P detector to evaluate the performance of OSP and LSOSP where the former is based on the pr observation model given by (3) and the latter is based on the ps model given by (6). The computer simulations in Section VI show that LSOSP outperforms OSP. In the following, we first briefly

review the N–P detection theory in Section V-A, then develop the N–P detector and locally optimal N–P detector for LSOSP in Section V-B. Finally, the same theory is applied to OSP in Section V-C.

A. N–P’s Detection Theory

A detection problem is generally described as a binary hypothesis testing

$$\begin{aligned} H_0: \quad Z &= n \sim p_0(z) \\ \text{versus} \\ H_1: \quad Z &= s + n \sim p_1(z) \end{aligned} \quad (32)$$

where Z is an observable random variable, s is a target signal and n is noise.

The hypothesis H_0 is called the null hypothesis, representing noise only and H_1 is the alternative hypothesis indicating the presence of a target signal s plus the noise. The decision of detecting a target signal is made based on an observation z generated by the random variable Z governed by one of two probability density functions $p_0(z)$ and $p_1(z)$ depending upon which hypothesis is true. An optimal detector is generally determined by criteria such as cost functions and prior probabilities associated with the hypotheses. If both are not known *a priori*, an alternative criterion is needed in this case. The N–P approach is one which adopts the false alarm probability P_F as a criterion measure defined as the probability of declaring the presence of a target when there is actually no target, i.e., the detector declares H_1 when H_0 is true. A counterpart of the false alarm probability is the probability of detection or detection power, P_D which is defined as the probability of detecting a target when the target is actually present, i.e., the detector declares H_1 when H_1 is true. The performance of a N–P’s detector is generally evaluated based on an ROC curve, a graph of the false alarm probability P_F plotted versus the detection power P_D .

It is known that the optimal detector yielded by the N–P approach turns out to be a likelihood ratio test with the following form

$$\delta_{NP}(z) = \begin{cases} 1 \text{ (declaring } H_1) & \text{if } L(z) = \frac{p_1(z)}{p_0(z)} > \tau \\ \eta \text{ (declaring } H_1) & \text{if } L(z) = \frac{p_1(z)}{p_0(z)} = \tau \\ 0 \text{ (declaring } H_0) & \text{if } L(z) = \frac{p_1(z)}{p_0(z)} < \tau \end{cases} \quad (33)$$

where $L(z)$ defined in (33) is called a likelihood ratio test or test statistic and τ is a threshold determined by the false alarm probability P_F . The η in (33) is the probability of the detector $\delta_{NP}(z)$ declaring H_1 when the likelihood ratio test $L(z)$ equals the threshold τ . As a result, the $\delta_{NP}(z)$ is called a randomized detector.

B. A Posteriori Signal Detection Model for LSOSP

Applying $P_{\mathbf{q}}$ in (30) to \mathbf{r} given by (6) or (20) yields

$$z = \mathbf{q}^T \mathbf{r} = \mathbf{q}^T \mathbf{d} \hat{\alpha}_p + \mathbf{q}^T \hat{\mathbf{n}}. \quad (34)$$

Equation (34) represents a standard signal model with a signal $\mathbf{q}^T \mathbf{d} \hat{\alpha}_p$ corrupted by the noise given by

$$\tilde{n} = \mathbf{q}^T \hat{\mathbf{n}}, \quad (34)$$

where $\hat{\mathbf{n}} = P_A \mathbf{n}$, z and the noise \tilde{n} are scalars.

Let Z be the random variable representing the observation z in (34). The hypothesis testing problem (32) becomes

$$\begin{aligned} H_0: & Z = \tilde{n} \sim p_0(z) \\ \text{versus} \\ H_1: & Z = \mathbf{q}^T \mathbf{d} \hat{\alpha}_p + \tilde{n} \sim p_1(z) \end{aligned} \quad (36)$$

where the null hypothesis H_0 and the alternative hypothesis H_1 represent noise alone and the presence of the estimated desired signature in the signature space respectively. $p_0(z)$ and $p_1(z)$ are Gaussian distributions given by

$$\begin{aligned} p_0(z) &= N(0, \sigma^2 \mathbf{q}^T P_A \mathbf{q}) \\ p_1(z) &= N(\mathbf{q}^T \mathbf{d} \hat{\alpha}_p, \sigma^2 \mathbf{q}^T P_A \mathbf{q}), \end{aligned} \quad (37)$$

and the estimated noise variance is given by

$$\sigma_{\tilde{n}}^2 = \sigma^2 \mathbf{q}^T P_A \mathbf{q}. \quad (38)$$

It should be noted that theoretically, the noise variance $\sigma_{\tilde{n}}^2$ in (38) is zero because $\hat{\mathbf{n}}$ is *ps* noise and when $P_A \mathbf{q}$ operates on $\hat{\mathbf{n}}$, the resultant noise \tilde{n} will be removed. This is due to the fact that (5) is the solution to an unconstrained least squares problem described by (3) and (4). In many applications, however, there are constraints such as all abundances must lie between 0 and 1, and add up to one [14], the lack of exact knowledge about signatures of endmembers, atmospheric effects, etc. As a result, $\sigma_{\tilde{n}}^2$ is generally very small but not zero.

Substituting (37) into (33) results in the following N-P detector

$$\tilde{\delta}_{NP}(z) = \begin{cases} 1 & \text{(declaring } H_1) \text{ if } z = \mathbf{q}^T \mathbf{r} \geq \tau' \\ 0 & \text{(declaring } H_0) \text{ if } z = \mathbf{q}^T \mathbf{r} < \tau' \end{cases} \quad (39)$$

where

$$\tau' = \sigma_{\tilde{n}} \Phi^{-1}(1 - a) \quad (40)$$

is the threshold obtained by specifying a particular false alarm probability a given by $a = \int_{\tau'}^{\infty} p_0(z) dz$ and $p_0(z)$ given by (37). The notation Φ is the standard Gaussian distribution with zero mean and unit variance given by $\Phi(x) = \int_{-\infty}^x e^{-(z^2/2)} dz$ and $\Phi^{-1}(y)$ is the inverse function of $\Phi(x)$ with $\Phi(x) = y$. It should be noted that since the observable random variable Z in (36) is governed by Gaussian distributions given by (37), a N-P's detector can be designed by a nonrandomized detector as given by (39) without using a randomized detector as in (33).

For the a given in (40) as a false alarm probability, P_D , can be calculated as follows

$$\begin{aligned} a &= \int_{\tau'}^{\infty} p_0(z) dz \\ P_D(\tilde{\delta}_{NP}) &= \int_{\tau'}^{\infty} p_1(z) dz. \end{aligned} \quad (41)$$

Substituting (40) into (41) yields

$$\begin{aligned} P_D(\tilde{\delta}_{NP}) &= 1 - \Phi\left(\frac{\tau' - \hat{\alpha}_p \mathbf{q}^T \mathbf{d}}{\sigma_{\tilde{n}}}\right) \\ &= 1 - \Phi\left(\Phi^{-1}(1 - a) - \frac{\hat{\alpha}_p \mathbf{q}^T \mathbf{d}}{\sigma_{\tilde{n}}}\right). \end{aligned} \quad (42)$$

In (42), we can see the detection power is determined by $\sigma_{\tilde{n}}$ for any fixed false alarm probability a . However, the $\sigma_{\tilde{n}}$ is in turn determined by $\sigma \sqrt{\mathbf{q}^T P_A \mathbf{q}}$. Since σ is a fixed constant, the P_D is inversely proportional to the quantity $\mathbf{q}^T P_A \mathbf{q}$.

B. Locally Optimal N-P's Detectors

In the hypothesis testing problem (36), the probability distribution $p_1(z)$ under hypothesis H_1 depends on the strength of the desired signature α_p . If the signal energy is very small and close to zero, i.e., $\alpha_p \rightarrow 0$, the solution given by (39) is not applicable. In order to account for weak signals, a composite hypothesis testing problem described below is used to replace (35) (see [11]–[12])

$$\begin{aligned} H_0: & Z = \tilde{n} \sim p_0(z) \\ \text{versus} \\ H_1: & Z = \theta s + \tilde{n} \sim p_{\theta}(z) \end{aligned} \quad (43)$$

where the noise \tilde{n} is given in (35), θ is a parameter controlling the energy of the signal s or SNR and supposed to be very small close to zero.

In this case, we expand the detection power P_D in terms of a Taylor series about 0. Namely

$$P_d(\delta; \theta) = P_D(\delta; 0) + \theta \left. \frac{\partial P_D(\delta; \theta)}{\partial \theta} \right|_{\theta=0} + O(\theta^2). \quad (44)$$

Since $P_D(\delta; 0) = P_F(\delta)$ as $\theta \rightarrow 0$, $P_d(\delta; \theta)$ can be approximated by

$$P_D(\delta; \theta) \approx P_F(\delta) + \theta \left[\left. \frac{\partial P_D(\delta; \theta)}{\partial \theta} \right|_{\theta=0} \right]. \quad (45)$$

From (45), the maximum power $P_D(\delta; \theta)$ can be approximately achieved by choosing a decision rule δ which maximizes $\partial P_D(\delta; \theta) / \partial \theta |_{\theta=0}$. The optimal detector for (43) obtained by maximizing $\partial P_D(\delta; \theta) / \partial \theta |_{\theta=0}$ is called a locally optimal N-P's detector which can be expressed by

$$\tilde{\delta}_{LNP}(z) = \begin{cases} 1 & \text{if } \left. \frac{\partial L_{\theta}(z)}{\partial \theta} \right|_{\theta=0} = \frac{\left. \frac{\partial p_{\theta}(z)}{\partial \theta} \right|_{\theta=0}}{p_0(z)} \geq \tau \\ 0 & \text{if } \left. \frac{\partial L_{\theta}(z)}{\partial \theta} \right|_{\theta=0} = \frac{\left. \frac{\partial p_{\theta}(z)}{\partial \theta} \right|_{\theta=0}}{p_0(z)} < \tau \end{cases} \quad (46)$$

where $\partial L_{\theta}(z) / \partial \theta$ is the partial derivative of the likelihood ratio test $L_{\theta}(z)$ with respect to θ . Substituting $\hat{\alpha}_p$ for θ and $\mathbf{q}^T \mathbf{d}$ for s in (43) yields the following hypothesis testing problem for weak signature $\hat{\alpha}_p$:

$$\begin{aligned} H_o: & Z = \tilde{n} \sim p_0(z) \\ \text{versus} \\ H_1: & Z = \hat{\alpha}_p [\mathbf{q}^T \mathbf{d}] + \tilde{n} \sim p_1(z) \end{aligned} \quad (47)$$

where \tilde{n} is given by (34).

From (47), the locally optimal N-P's detector for a very small signature α_p (i.e., close to zero) is found by

$$\tilde{\delta}_{\text{LNP}}(z) = \begin{cases} 1 & \text{if } \left. \frac{\partial L_{\hat{\alpha}_p}(z)}{\partial \hat{\alpha}_p} \right|_{\hat{\alpha}_p=0} = \frac{\left. \frac{\partial p_{\hat{\alpha}_p}(z)}{\partial \hat{\alpha}_p} \right|_{\hat{\alpha}_p=0}}{p_0(z)} = \mathbf{z}^T \frac{\mathbf{q}^T \mathbf{d}}{\sigma_{\tilde{n}}^2} \geq \tau \\ 0 & \text{if } \left. \frac{\partial L_{\hat{\alpha}_p}(z)}{\partial \hat{\alpha}_p} \right|_{\hat{\alpha}_p=0} = \frac{\left. \frac{\partial p_{\hat{\alpha}_p}(z)}{\partial \hat{\alpha}_p} \right|_{\hat{\alpha}_p=0}}{p_0(z)} = \mathbf{z}^T \frac{\mathbf{q}^T \mathbf{d}}{\sigma_{\tilde{n}}^2} < \tau \end{cases} \quad (48)$$

From (48) the false alarm probability is given by

$$a = \int_{\tau'}^{\infty} p_0(z) dz \quad (49)$$

where

$$\tau' = \tau \left(\frac{\mathbf{q}^T \mathbf{d}}{\sigma_{\tilde{n}}^2} \right)^{-1} = \tau \left(\frac{\sigma_{\tilde{n}}^2}{\mathbf{q}^T \mathbf{d}} \right). \quad (50)$$

This implies that

$$\tau' = \sigma_{\tilde{n}} \Phi^{-1}(1-a). \quad (51)$$

In order to find the locally optimal detection power specified by (45) with $\theta = \hat{\alpha}_p$, we need calculate $\partial P_D(\tilde{\delta}_{\text{LNP}}; \hat{\alpha}_p) / \partial \hat{\alpha}_p |_{\hat{\alpha}_p=0}$ as follows:

$$\begin{aligned} & \left. \frac{\partial P_D(\tilde{\delta}_{\text{LNP}}; \hat{\alpha}_p)}{\partial \hat{\alpha}_p} \right|_{\hat{\alpha}_p=0} \\ &= \int_{\tau'}^{\infty} \left[\frac{\mathbf{q}^T \mathbf{d}}{\sqrt{2\pi\sigma_{\tilde{n}}^2}} e^{-[(z-\hat{\alpha}_p\mathbf{q}^T\mathbf{d})^2/2\sigma_{\tilde{n}}^2]} \cdot \left(\frac{z-\hat{\alpha}_p\mathbf{q}^T\mathbf{d}}{\sigma_{\tilde{n}}^2} \right) \mathbf{q}^T \mathbf{d} \right] \Bigg|_{\hat{\alpha}_p=0} dz \\ &= \frac{\mathbf{q}^T \mathbf{d}}{\sqrt{2\pi\sigma_{\tilde{n}}}} e^{-[(\Phi^{-1}(1-a))^2/2]}. \end{aligned} \quad (52)$$

The last equality of (52) follows from substituting (51) for τ' .

Placing (49) and (52) into (45) yields the locally optimal detection power

$$P_D(\tilde{\delta}_{\text{LNP}}; \hat{\alpha}_p) = a + \hat{\alpha}_p \left[\frac{(\mathbf{q}^T \mathbf{d})}{\sqrt{2\pi\sigma_{\tilde{n}}}} e^{-(1/2)(\Phi^{-1}(1-a))^2} \right]. \quad (53)$$

From (53), we note that the locally optimal detection power is also determined by the standard deviation of the noise \tilde{n} , i.e., $\sigma_{\tilde{n}} = \sigma \sqrt{\mathbf{q}^T P_A \mathbf{q}}$ and the signal strength $\mathbf{q}^T \mathbf{d}$, the projection of \mathbf{d} by the LSOSP P_q .

C. A Priori Signal Detection Model for OSP

Following the same approach given in Sections V-A and B we obtain similar results for the pr signal model for OSP without further derivations as follows

$$\begin{aligned} H_0: & Z = \bar{n} \sim p_0(z) \\ \text{versus} & \\ H_1: & Z = \mathbf{q}_{\text{OSP}}^T \mathbf{d} \alpha_p + \bar{n} \sim p_1(z) \end{aligned} \quad (54)$$

TABLE I
PARAMETERS USED FOR FSS

Number of Bands	60
Spectral Cover	0.4-2.4 μm
Altitude	60m
IFOV (ground)	25m

where $\mathbf{q}_{\text{OSP}}^T = \mathbf{d}^T P_{\text{OSP}}$ is the operator developed in [7], $\bar{n} = \mathbf{q}_{\text{OSP}}^T \mathbf{n}$ with variance $\sigma_{\tilde{n}}^2 = \sigma^2 \mathbf{q}_{\text{OSP}}^T \mathbf{q}_{\text{OSP}}$ and $p_0(z), p_1(z)$ are given by

$$\begin{aligned} p_0(z) &= N(0, \sigma^2 \mathbf{q}_{\text{OSP}}^T \mathbf{q}_{\text{OSP}}) \\ p_1(z) &= N(\mathbf{q}_{\text{OSP}}^T \mathbf{d} \alpha_p, \sigma^2 \mathbf{q}_{\text{OSP}}^T \mathbf{q}_{\text{OSP}}). \end{aligned} \quad (55)$$

The resulting N-P detector is given by

$$\tilde{\delta}_{\text{OSP,NP}}(z) = \begin{cases} 1 & \text{(declaring } H_1) \text{ if } z = \mathbf{q}_{\text{OSP}}^T \mathbf{r} \geq \tau'' \\ 0 & \text{(declaring } H_0) \text{ if } z = \mathbf{q}_{\text{OSP}}^T \mathbf{r} < \tau'' \end{cases} \quad (56)$$

and

$$\tau'' = \sigma_{\tilde{n}} \Phi^{-1}(1-a) \quad (57)$$

with the threshold τ'' obtained by specifying a particular false alarm probability a given by $a = \int_{\tau''}^{\infty} p_0(z) dz$ where $p_0(z)$ is given by (55). The corresponding detection power is

$$\begin{aligned} P_{\text{OSP},D}(\tilde{\delta}_{\text{OSP,NP}}) &= 1 - \Phi \left(\frac{\tau'' - \alpha_p \mathbf{q}_{\text{OSP}}^T \mathbf{d}}{\sigma_{\tilde{n}}} \right) \\ &= 1 - \Phi \left(\Phi^{-1}(1-a) - \frac{\alpha_p \mathbf{q}_{\text{OSP}}^T \mathbf{d}}{\sigma_{\tilde{n}}} \right). \end{aligned} \quad (58)$$

Analogous to (53), the locally optimal detection power for (54) can be obtained similarly by

$$\begin{aligned} P_{\text{OSP},D}(\tilde{\delta}_{\text{OSP,LNP}}; \alpha_p) \\ &= a + \alpha_p \left[\frac{\mathbf{q}_{\text{OSP}}^T \mathbf{d}}{\sqrt{2\pi\sigma_{\tilde{n}}}} e^{-(1/2)(\Phi^{-1}(1-a))^2} \right]. \end{aligned} \quad (59)$$

Comparing (58) with (42) and (59) with (53), LSOSP introduces an extra noise annihilator operator P_A in (42) and (53) which takes advantage of the ps knowledge about $\hat{\alpha}_p$ based on (6) or (20) to suppress noise while (58) and (59) only relying on α_p and the pr information provided by the model (3). This effect makes a significant difference in detecting weak signatures which is shown in the computer simulations in Section VI.

VI. COMPUTER SIMULATION RESULTS

In this section, experiments are conducted by computer simulations based on the field spectrometer system (FSS) data which has 60 spectral bands [18]. No other effects in actual remote sensing from sensors and atmosphere but white Gaussian noise is simulated. Namely, the raw radiance data collected by sensors are converted to reflectance data

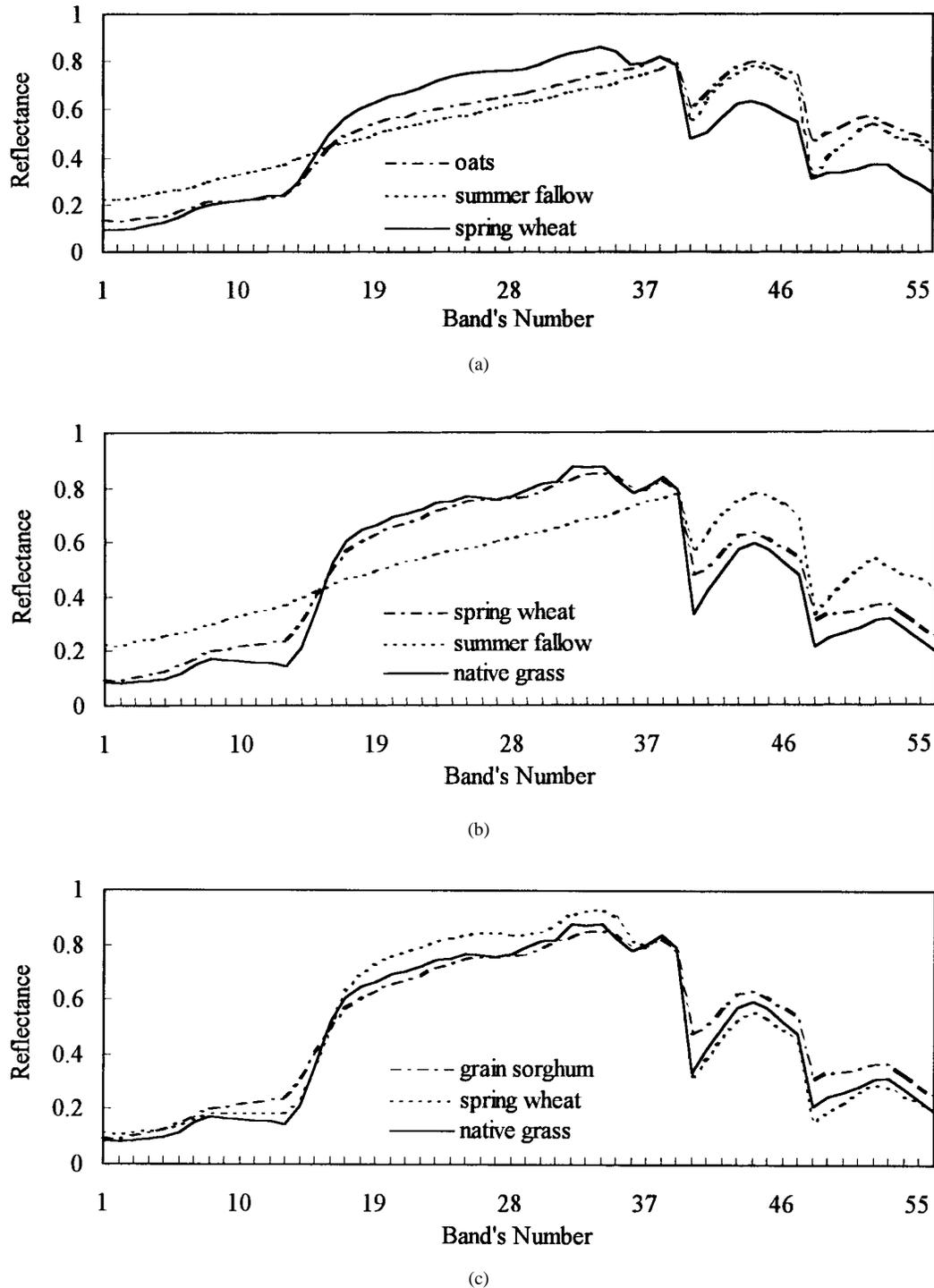


Fig. 2. (a) Reflectance spectra of data set 1. (b) Reflectance spectra of data set 2. (c) Reflectance spectra of data set 3.

by removing the effects of atmospheric transmission and scattering. The major parameters of the FSS data are listed in Table I. Since bands corresponding to the water absorption regions have no useful energy, they are removed prior to processing which leaves 56 bands in this study. In [7] the OSP method failed to classify a signature with abundance less than 5%. In Experiment 1, we consider the situation where the desired signature contains abundance less than 5%. It is found that in the N-P's detection approach described in Section IV,

projecting observation vectors into the signature space where only a very small part of the noise is contributed to the signature space significantly improves the detection power. As a result, the desired signature with a small fraction of abundance can be picked up by the LSOSP method but not by the OSP method. If the abundance is very small and close to zero, the N-P detector cannot be applied. Instead, a locally optimal N-P's detector needs to be implemented. Experiment 2 examines this case.

TABLE II
SIMULATIONS OF DESIRED AND UNDESIRED SIGNATURES FOR THREE DATA SETS

Data set	Undersired Signatures	Desired Signature
Data set-1	oats and summer fallow	spring wheat
Data set-2	spring wheat and summer fallow	native grass
Data set-3	spring wheat and grain sorghum	native grass

In the following experiments we consider $p = 3$ and $\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$ is a spectral abundance vector corresponding to a signature matrix $M = (\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)$, with \mathbf{m}_3 as the desired signature and α_3 as its associated spectral abundance. $U = (\mathbf{m}_1, \mathbf{m}_2)$ is the undesired signature vector. Three simulated data sets are used to test the proposed the LSOSP and OSP methods. Each data set contains three signatures from five endmembers listed in Table II. Data set 1 contains three distinct signatures as shown in Fig. 2(a) where spring wheat is designated as the desired signature, and oats and summer fallow are undesired signatures. Data set 2 shown in Fig. 2(b) contains summer fallow and two signatures with similar spectral reflectances, spring wheat and native grass, with native grass selected as the desired signature. Data set 3 contains spring wheat, grain sorghum and native grass whose spectral reflectances are nearly indistinguishable as shown in Fig. 2(c), with native grass chosen to be the desired signature. All the three data sets are simulated based on ground truth. 50 pixels to be used in the experiments are simulated as follows. Each pixel contains three different signatures with various spectral reflectance abundances. White Gaussian noise is also simulated and added to each pixel to generate three SNR's, 50:1, 30:1, and 10:1, respectively, with the SNR is defined as 50% reflectance divided by the standard deviation of the noise [7]. It should be noted that this defined SNR is an accepted convention used in the remote sensing community and is different from the SNR defined by (15).

A. Experiment 1 (N - P 's Detection)

As shown in [7], when a desired signature has abundance less than 5%, the OSP method failed to pick it up. In this experiment, we will show that this case can be improved by the LSOSP method. To do so we simulate 50 pixels for each of the three data sets as given in Table III where the 50 pixels are divided equally into five classes, each of which contains 10 pixels. The 10 pixels in each class contain the same amount of signature abundance. For example, The pixels in the first class contain 1% abundance of the desired signature and 49.5% for each of two undesired signatures.

Three experiments (false alarm probabilities $P_F = 0.1, 0.01$, and 0.001) are conducted, respectively, for each data set under three different SNR's. The results are shown in Figs. 3–5 for $P_F = 0.1$ and Table IV for $P_F = 0.01$ and 0.001 . Fig. 3 is obtained for data set 1 for three different SNR's, Fig. 3(a) for SNR = 50:1, Fig. 3(b) for SNR = 30:1 and Fig. 3(c) for SNR = 10:1. Similarly, Figs. 4 and 5 are the results of data sets 2 and 3, respectively. As we can see from these figures

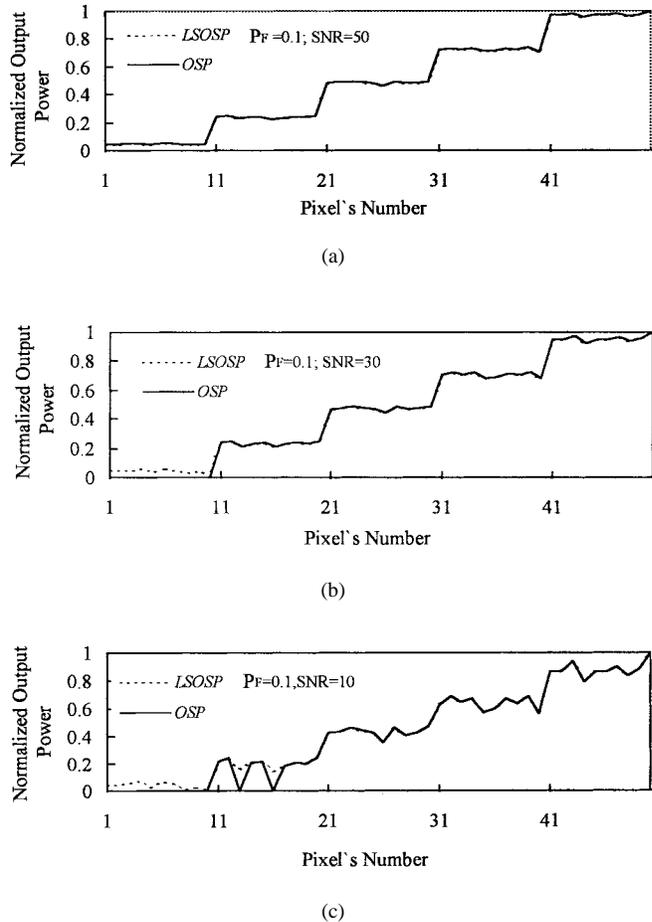


Fig. 3. Simulation results for false alarm probability = 0.1 for data set 1. (a) Normalized output power versus pixel's number for SNR = 50:1. (b) Normalized output power versus pixel's number for SNR = 30:1. (c) Normalized output power versus pixel's number for SNR = 10:1.

(a), (b), and (c), the detection performance degrades as the SNR decreases. In addition, Figs. 3–5 show that the detection capability also depends upon the spectral similarity between signatures. The more similar the spectra of signatures, the more difficult the signatures are to detect.

Since the LSOSP works more impressively than OSP for low false alarm probabilities, we tabulate in detail the results of $P_F = 0.01$ and 0.001 in Table IV. The columns of the table show the three SNR's, 50:1, 30:1 and 10:1 using the LSOSP and OSP methods under two different false alarm probabilities $P_F = 0.01$ and 0.001 . The numbers in each row represent the number of pixels detected by the LSOSP and OSP methods. For example, the row labeled by 1% under the data set 1 shows that the OSP method missed all pixels which contain 1% desired signature abundance even in the case that the SNR is very strong (50:1). This situation significantly improves using the LSOSP method which detected all 10 pixels for SNR = 50:1 and 30:1. In the case of SNR = 10:1, it detected eight pixels and only missed two pixels.

A more important factor that dominates the detection performance is the similarity of signature spectra. This is shown in Table IV for data set 3 where all three signatures are similar. Neither of the OSP and LSOSP methods can detect the case

TABLE III
ABUNDANCE FOR 50 SIMULATED PIXELS FOR THREE DATA SETS

%	pixel 1-10	pixel 11-20	pixel 21-30	pixel 31-40	pixel 41-50
Desired Signature	1%	5%	10%	15%	20%
Undesired Signature 1	49.5%	47.5%	45%	42.5%	40%
Undesired Signature 2	49.5%	47.5%	45%	42.5%	40%

TABLE IV
SIMULATIONS FOR FALSE ALARM PROBABILITY = 0.01 AND 0.001 FOR THREE DATA SETS

%	$P_F=0.01$						$P_F=0.001$					
	SNR=50		SNR=30		SNR=10		SNR=50		SNR=30		SNR=10	
	LSOSP	OSP	LSOSP	OSP	LSOSP	OSP	LSOSP	OSP	LSOSP	OSP	LSOSP	OSP
Data set 1												
1%	10	0	10	0	8	0	10	0	10	0	8	0
5%	10	10	10	10	10	0	10	10	10	10	10	0
10%	10	10	10	10	10	10	10	10	10	10	10	9
15%	10	10	10	10	10	10	10	10	10	10	10	10
20%	10	10	10	10	10	10	10	10	10	10	10	10
Data set 2												
1%	8	0	3	0	3	0	8	0	3	0	3	0
5%	10	9	10	0	10	0	10	0	10	0	10	0
10%	10	10	10	7	10	0	10	10	10	0	10	0
15%	10	10	10	10	10	0	10	10	10	8	10	0
20%	10	10	10	10	10	0	10	10	10	10	10	0
Data set 3												
1%	0	0	0	0	0	0	0	0	0	0	0	0
5%	10	0	3	0	0	0	10	0	3	0	0	0
10%	10	3	10	0	2	0	10	0	10	0	2	0
15%	10	10	10	1	3	0	10	9	10	0	3	0
20%	10	10	10	9	9	0	10	10	10	2	9	0

where the desired signature has 1% abundance even when the SNR is as high as 50:1. Nevertheless, LSOSP improves its performance rapidly as SNR increases. In other words, three factors affect the detection performance, SNR, false alarm probability P_F , and the spectral similarity between desired and undesired signatures. The higher the SNR, the better the detection; the smaller the false alarm probability, the worse the detection; and the more similar the spectra, the more difficult the detection.

This example clearly demonstrates that the LSOSP method outperforms the OSP method in all experiments. In particular, in an unknown environment, using *pr* knowledge may not be reliable. Instead, *ps* knowledge based on observations will be more realistic and provide better estimates for the environment.

B. Experiment 2 (Locally Optimal N-P Detection)

In Experiment 1, we found that both OSP and LSOSP could not detect the existence of a signature with abundance

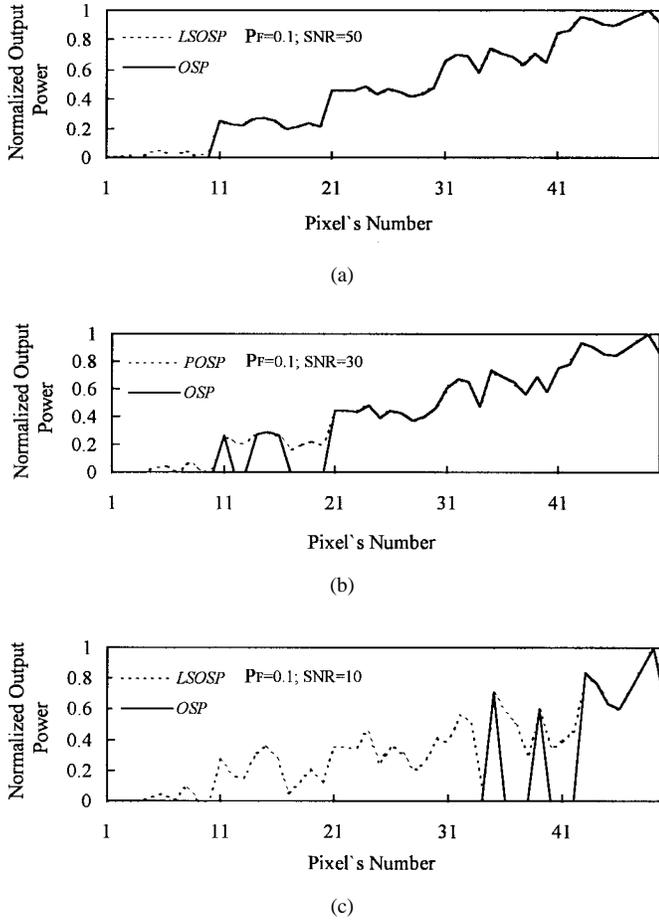


Fig. 4. Simulation results for false alarm probability = 0.1 for data set 2. (a) Normalized output power versus pixel's number for SNR = 50:1. (b) Normalized output power versus pixel's number for SNR = 30:1. (c) Normalized output power versus pixel's number for SNR = 10:1.

1% as shown in the row 1% of data set 3 of Table IV. In this case, we turn to the locally optimal N-P's detection theory to derive a locally optimal detector for weak signatures. Equations (53) and (59) are the powers of a locally optimal N-P's detectors for LSOSP and OSP respectively given that a false alarm probability is fixed at level a . In this experiment, 100 pixels are simulated as follows. Pixel one contains 0.01% of the desired signature, pixel two contains 0.02%, pixel three 0.03%, etc. The other two undesired signatures evenly share the remaining abundance. As the pixel number increases by 1, the abundance of the desired signature is increased by 0.01%. So, the hundredth pixel contains the largest abundance of the desired signature, which is 1%. The ROC curves are plotted for OSP and LSOSP in Fig. 6 with abundance ranging from 0.01% to 1% for 100 pixels, Fig. 6(a) for data set 1 with SNR's = 0.1, 1, 5, 10. 6(b) for data set 2 with SNR's = 0.1, 1, 5, 10 and 6(c) for data set 3 with SNR's = 0.1, 1, 5, 10. It should be noted that the diagonal line in the ROC plot represents the case that the false alarm probability is equal to the detection power, which is the lower bound to the detection power. This can be seen from (53) or (59). The first term on the right in (53) is the false alarm probability a and the second term is dominated by two factors, $\hat{\alpha}_p$ (or α_p) and a . If $\hat{\alpha}_p = 0$ (or $\alpha_p = 0$), then the

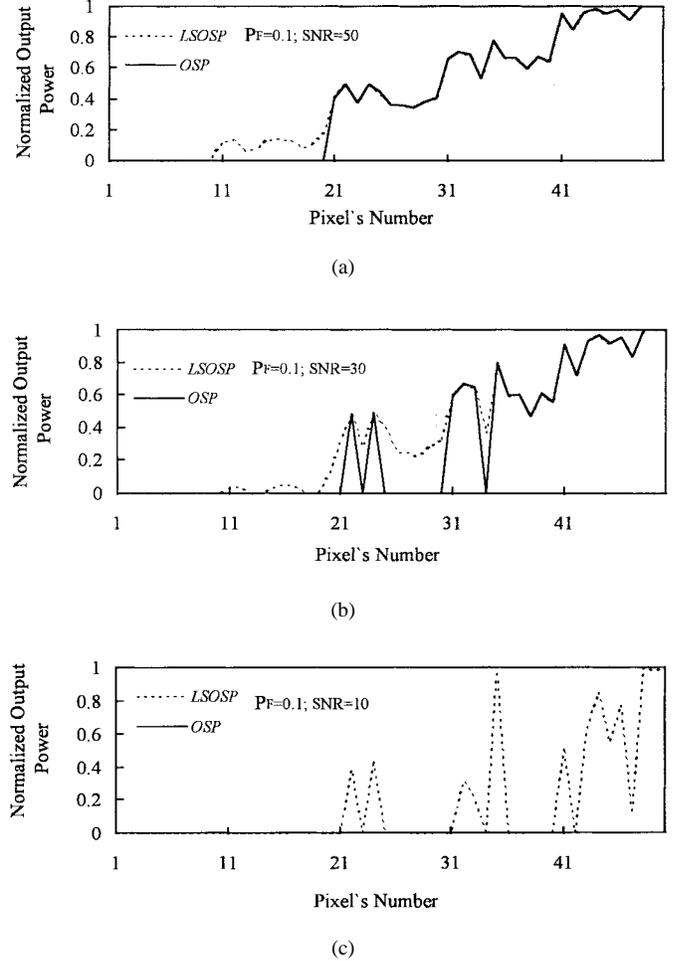


Fig. 5. Simulation results for false alarm probability = 0.1 for data set 3. (a) Normalized output power versus pixel's number for SNR = 50:1. (b) Normalized output power versus pixel's number for SNR = 30:1. (c) Normalized output power versus pixel's number for SNR = 10:1.

ROC curve is the diagonal line. When $a \rightarrow 0$ or $a \rightarrow 1$, the second term on the right in (53) or (59) always approaches 0, in which case, $P_D(\hat{\delta}_{LNP}; \hat{\alpha}_p) = a$ or $P_D(\hat{\delta}_{LNP}; \alpha_p) = a$ as expected. So, in the case of the weak signature detection, we may anticipate that the corresponding ROC curves will be slightly above this diagonal line as shown in Fig. 6, which indicates that the locally optimal detection power will not be as good as the case of strong signature detection. In Fig. 6, we can see that LSOSP significantly improve OSP in all three data sets, particularly, for data set 1, Fig. 6(a). It is also shown in Fig. 6 that the spectral similarity between signatures plays an important role in detection performance and the detection power is directly proportional to the SNR as well as the false alarm probability a .

VII. CONCLUSIONS

In this paper, a ps least squares orthogonal subspace projection (LSOSP) method is presented which is based on a ps linear spectral mixture model. It is different from the OSP in [7] derived from a pr linear spectral mixture model. The advantage of LSOSP is that it estimates the desired

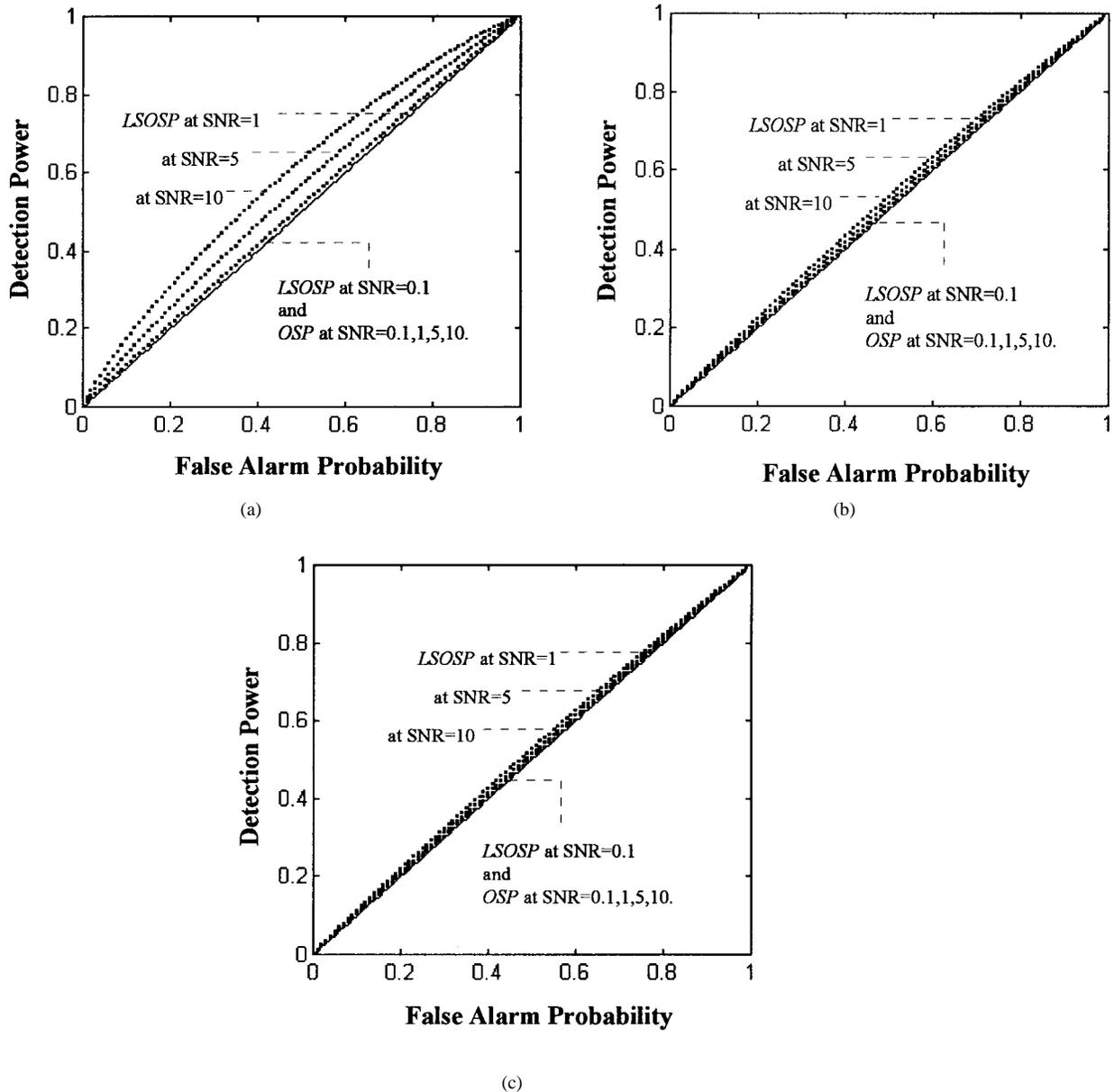


Fig. 6. ROC curves for locally optimal detection. (a) Data set 1. (b) Data set 2. (c) Data set 3.

signature abundance by projecting an observation into the signature space while suppressing noise. As a result, the effects incurred by noise are greatly reduced so as to improve the OSP performance developed in [7]. The comparative performance of OSP and LSOSP is evaluated by the ROC analysis through the N-P detection theory. The computer simulations demonstrate that LSOSP does perform better than OSP in all cases. More details on comparative analysis of OSP and LSOSP can be found in [19], [20]. Furthermore, in the case where a signature is very weak and close to zero or the SNR is very small, a locally optimal N-P's detector is also developed for OSP and LSOSP. Finally, some further studies on subspace projection methods are currently under investigation. For example, an analysis of the estimation errors resulting from LSOSP is investigated in [21]. Also, an oblique subspace projection approach is proposed in [19], [20], [22] as

an alternative to LSOSP where the oblique subspace projection is not necessarily orthogonal as is LSOSP.

ACKNOWLEDGMENT

The authors would like to thank Prof. D.A. Landgrebe of Purdue University who provided the FSS data and the anonymous reviewers' suggestions which greatly improved this paper's quality and presentation.

REFERENCES

- [1] G. Vane and A. F. H. Goetz, "Terrestrial imaging spectroscopy," *Remote Sensing Environ.*, vol. 24, no. 1, pp. 1-29, 1988.
- [2] D. A. Landgrebe, "A perspective on the analysis of hyperspectral data," *Proc. IGARSS '93 Symp.*, pp. 1362-1364, 1993.
- [3] W. M. Porter and H.T. Enmark, "A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS)," *Imaging Spectroscopy II*, G. Vane, Ed. Bellingham, WA: Society of Photo-Optical Instrumentation Engineering, vol. 834, pp. 22-31, 1987.

- [4] M. O. Smith, P. E. Johnson and J. B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis," *J. Geophys. Res.*, vol. 90, supplement, pp. C797-C804, 1985.
- [5] J. B. Adams, and M. O. Smith, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res.*, vol. 91, pp. 8098-8112, July 1986.
- [6] B. Hapke, "Bidirectional reflectance spectroscopy. 1. Theory," *J. Geophys. Res.*, vol. 86, pp. 3039-3054, July 1981.
- [7] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach" *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 779-785, July 1994.
- [8] J. C. Harsanyi, "Detection and Classification of Subpixel Spectral Signatures in Hyperspectral Image Sequences," Ph.D. Dissertation, Department of Electrical Engineering, University of Maryland Baltimore County, MD, Aug. 1993.
- [9] J. W. V. Miller, J. B. Farison, and Y. Shin, "Spatially invariant image sequences," *IEEE Trans. Image Processing*, vol. 1, pp. 148-161, Apr. 1992.
- [10] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* MA: Addison-Wesley, ch. 9, 1991.
- [11] J. Capon, "On the asymptotic efficiency of locally optimum detectors," *IRE Trans. Inform. Theory*, vol. IT-7, pp. 67-71, Apr. 1961.
- [12] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [13] A. D. Whalen, *Detection of Signals in Noise*. San Francisco, CA: Academic, 1971.
- [14] Y. E. Shimabukuro and J. A. Smith, "Least squares mixing models to generate fraction images derived from multispectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, pp. 16-20, Jan. 1991.
- [15] A. R. Gillespie, M. O. Smith, J. B. Willis, S. C. Fischer, A. F. III, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images," Owens Valley, California," in *Proc. 2nd AVIRIS Workshop*, JPL Publ., CA 4-5, pp. 243-270, June 1990.
- [16] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, ch. 6, 1973, pp. 221-225.
- [17] J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*, 2nd ed. Berlin, Germany: Springer-Verlag, 1993, pp. 255-258.
- [18] L. L. Biehl *et al.*, "A crops and soils data base for scene radiation research," in *Proc. Machine Processing Remotely Sensed Data Symp.*, 1982, pp. 169-177.
- [19] X. Zhao, "Subspace Projection Approach to Hyperspectral Image Classification Using Linear Spectral Mixture Model," Master Thesis, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, MD, May, 1996.
- [20] C.-I Chang *et al.*, "Least squares subspace approach to mixed pixel classification for hyperspectral images," submitted for publication.
- [21] C.-I Chang, "Error analysis of least squares subspace projection approach to linear mixing problems," submitted for publication.
- [22] T.-M. Tu *et al.*, "An oblique subspace projection approach to mixed pixel classification in multispectral/hyperspectral images," submitted for publication.



Te-Ming Tu was born in Kaohsiung, Taiwan, in 1959. He received the B.S.E.E. degree from Chung Cheng Institute of Technology in 1986, the M.S.E.E. degree from Sun Yat-Sen University in 1989, and is currently pursuing the Ph.D. degree at the National Cheng Kung University.

Since 1981 he has been employed by the R.O.C. Army Command, working on communication engineering. He was a Teaching Assistant from 1986 to 1989 and an Instructor from 1991 to 1993 within the Department of Electrical Engineering at Chung

Institute of Technology. His research interests include remote sensing, medical image, neural networks, and statistical pattern recognition.



Chin-Hsing Chen was born in Tainan, Taiwan, ROC, in 1958. He studied at the National Taiwan University, where he received the B.S. degree in electrical engineering in 1980. He received the M.S. and Ph.D. degrees in electrical and computer engineering from UCSB in 1983 and 1987, respectively.

After serving as a Communication Officer in the Chinese Marine Corps for two years, he returned to graduate studies at the University of California at Santa Barbara where he worked on research projects joined by Santa Barbara Research Center and Delco

Electronics. In 1988, he became an Associate Professor of the Department of Electrical Engineering in National Cheng Kung University, Taiwan, ROC. His current interests include medical imaging, remote sensing, robot vision, fractal and chaos, wavelet transform, and systolic circuit implementation.

Dr. Chen has published 110 papers and has given more than 65 technical presentations at meeting of professional societies in more than 15 countries. He has served as a section chairman and a member of various organization committees for several international conferences. In 1992, he was granted a Visiting Researchship at Academia Sinica by the National Science Council of ROC. He has received several awards for students who received best papers and excellent papers under his supervision.



Chein-I Chang (S'81-M'87-SM'92) received the B.S. degree from Soochow University, Taipei, Taiwan, in 1973, the M.S. degree from National Tsing Hua University, Hsinchu, Taiwan in 1975, and the M.A. degree from SUNY at Stony Brook in 1977, all in mathematics. He received the M.S. and MSEE degrees from the University of Illinois at Urbana-Champaign in 1980 and 1982 and the Ph.D. degree in electrical engineering from the University of Maryland, College Park in 1987.

He was a visiting Assistant Professor from January 1987 to August 1987, an Assistant Professor from 1987 to 1993, and is currently an Associate Professor in the Department of Computer Science and Electrical Engineering at the University of Maryland, Baltimore County. He was a visiting specialist in the Institute of Information Engineering at National Cheng Kung University, Taiwan, ROC from 1994 to 1995. His research interests include information theory and coding, signal detection and estimation, multispectral/hyperspectral remote sensing, neural networks, and pattern recognition.

Dr. Chang is a member of SPIE, INNS, Phi Kappa Phi, and Eta Kappa Nu.