[3] S. Spitz, "Seismic trace interpolation in the F–X domain," *Geophysics,* vol. 56, pp. 785–794, June 1991.
[4] J. Claerbout, *Earth Sounding Analysis.* Boston, MA: Blackwell, 1992.
[5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Signal Processing Part II: Control Theory and Applications,* vol. 23. New York: Springer-Verlag, 1990.
[6] S. Haykin, *Adaptive Filter Theory,* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.
[7] S. Kay, *Modern Spectral Estimation.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[8] A. Cordova, W. Gautschi, and S. Ruscheweyh, "Vandermonde matrices on the circle: Spectral properties and conditioning," *Numer. Math.,* vol. 57, no. 6-7, pp. 577–591, 1990.

Further Results on Relationship Between Spectral Unmixing and Subspace Projection

Chein-I Chang

***Abstract*—A recent short communication [1] showed that an orthogonal subspace projection (OSP) classifier developed for hyperspectral image classification in [2] was equivalent to a maximum likelihood estimator (MLE) resulting from a standard method of linear unmixing. It further concluded that the MLE subsumed the OSP classifier in spite of a constant difference in their magnitudes. Coincidentally, the equivalence of the OSP approach to linear unmixing was also derived in [3] and [4] by using the least-squares estimation with the same abundance estimate given by the MLE. In this communication, we show, on the contrary, that the MLE can be viewed as an *a posteriori* version of the OSP classifier and, thus, belongs to a family of OSP-based classifiers. More importantly, we further show that the constant produced by the MLE determines abundance estimation and has nothing to do with classification. As a result, it only alters the abundance concentration of the classified pixels, but not classification results.**

***Index Terms*—Least-squares estimate, maximum likelihood estimator (MLE), orthogonal subspace projection (OSP), spectral unmixing.**

I. Introduction

In a recent paper [2], Harsanyi and Chang developed an orthogonal subspace projection (OSP) approach to hyperspectral image classification that has shown promise in HYperspectral Digital Imagery Collection Experiment (HYDICE) data exploitation [5], [6]. For example, a family of OSP-based methods have been developed and presented in [6] since the OSP was introduced. Unfortunately, most of them were not published in the literature. Therefore, the OSP approach seems to be limited to the community that involves the HYDICE program. The potential and usefulness of the OSP has not been recognized outside this group. Based on the fact that the OSP in [2] assumed complete knowledge about signature abundance, a recent communication [1] showed that a maximum likelihood estimator (MLE) resulting from linear unmixing could be used for abundance estimation. It then concluded that the OSP approach was equivalent to a standard method of the linear unmixing and the MLE subsumed the OSP classifier. Interestingly, this equivalence was also derived by Harsanyi [3, chapter 4] before the OSP was published. Nevertheless, the OSP and the MLE were developed based on different principles. Here, we will show that using the concept of *a posteriori* model proposed in [4] the MLE can be implemented as an *a posteriori* OSP classifier with the unknown signature abundance replaced by its least-squares estimate or maximum likelihood estimate. Since the MLE is simply a scaled version of the OSP classifier by a constant β, they both generate the same classification feature vector and, therefore, produce the same classification results. In the light of this interpretation, the MLE can be viewed as one of the OSP-based classifiers. However, it is the β that gives rise to an extra estimation error for the MLE. Although [1] also noted the close relationship of β to the abundance estimation error, it unfortunately did not go in depth to exploit the implication and significance of the β. In this communication, we provide an error analysis to show how the β is related to the estimation error performance. Accordingly, only the fraction of the abundance in the classified pixels is affected by the β. This well explains why the OSP classifier could be used for classification of real image data without assuming *a priori* knowledge about abundance as it should [2].

Manuscript received April 28, 1997.

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@umbc2.umbc.edu).

Publisher Item Identifier S 0196-2892(98)02859-9.

II. Error Analysis

Let $\boldsymbol{r}$ be an $l \times 1$ column vector and denote a hyperspectral image pixel vector, where l is the number of spectral bands. Assume that M is an $l \times p$ signature matrix denoted by $(\boldsymbol{m}_1 \;\; \boldsymbol{m}_2 \;\; \cdots \;\; \boldsymbol{m}_p)$ where $\boldsymbol{m}_j$ is an $l \times 1$ column vector represented by the jth spectral signature (reflectance) within the pixel vector $\boldsymbol{r}$ and p is the number of signatures. We also let α be a $p \times 1$ abundance column vector given by $(\alpha_1 \;\; \alpha_2 \;\; \cdots \;\; \alpha_p)^T$, where α_j denotes the fraction of the jth signature present in $\boldsymbol{r}$.

A linear spectral mixture model for $\boldsymbol{r}$ is described by

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

where $\boldsymbol{n}$ is an $l \times 1$ column vector representing an additive white Gaussian noise with zero mean and variance $\sigma^2 I$ and I is the $l \times l$ identity matrix. We can further rewrite model (1) as

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

where U is the undesired spectral signature matrix made up of a set of $p-1$ undesired signatures $U = (\boldsymbol{m}_1 \;\; \boldsymbol{m}_2 \;\; \cdots \;\; \boldsymbol{m}_{p-1})$ and $\boldsymbol{d} = \boldsymbol{m}_p$ is the desired target signature.

Based on model (2), an OSP projector was developed in [2] and given by

$$P = I - UU^{\#} \tag{3}$$

where $U^{\#} = (U^T U)^{-1} U^T$ is the pseudo-inverse of U and an OSP classifier can be designed by

$$q^T \boldsymbol{r} = \boldsymbol{d}^T P \boldsymbol{r} \quad \text{for any pixel vector } \boldsymbol{r}. \tag{4}$$

Now suppose that α is an unknown constant needed to be estimated. From [2], [3]–[4], and [7], the least-squares estimate of α was given by

$$\hat{\alpha} = (M^T M)^{-1} M^T \boldsymbol{r}. \tag{5}$$

Equation (5) turns out to be exactly the same form derived by the MLE in [1]. In particular, the estimate of abundance α_p, denoted by $\hat{\alpha}_p$, can be expressed by

$$\hat{\alpha}_p = \beta(\boldsymbol{d}^T P)\boldsymbol{r} \tag{6}$$

which is identical to the first equation of (31) in [4], with β given by

$$\beta = (\boldsymbol{d}^T P \boldsymbol{d})^{-1}. \tag{7}$$

It should be noted that $\hat{\alpha}_p$ in (6) should be $\hat{\alpha}_p(\boldsymbol{r})$, which is a function of an observation vector $\boldsymbol{r}$. For simplicity, the $\boldsymbol{r}$ is omitted from the notation.

Equation (6) shows that the least-squares estimate derived for α_p in [3] and [4] is the same one obtained by the MLE for α_p, despite the fact that the former used the subspace projection approach, while the latter used the maximum likelihood estimation approach. Comparing (6) to (4), the only difference is the constant β appearing in (6) but not in (4). In [1], it was noted that the β is closely related to the mean-square prediction error of α_p and could be useful if it is compared against the theoretical predictions afforded by the prediction error matrix $\sigma^2(M^T M)^{-1}$. In this communication, we will further exploit the role of the β in the estimation of abundance using a receiver operating characteristics (ROC) analysis within the Neyman–Pearson detection theory setting.

As shown previously, (4) was derived from the model given by (1) that assumes complete knowledge about the model, including α. This model was referred to as *a priori* or Bayes model in [4]. Since no estimation is required for model (1), there is no β included in (4). On the other hand, (6) was obtained by estimating the unknown α. As a result, a constant β is introduced in (6) at the expense of inaccurate estimation of α. It is a direct result of the error from estimating α. Therefore, the β provides an important measure of the estimated abundance present in the classified pixels. Unfortunately, this point was not addressed in [1] and [3].

In what follows, we will show that the estimation error of abundance is indeed a function of the β. The idea is to formulate the error as the noise considered in a standard signal detection model. By appealing for the Neyman–Pearson detection theory, the detection power can be interpreted as the effectiveness of an abundance estimation technique and evaluated by the ROC curve.

From the last equation of (31) in [4], (6) can be represented by

$$\hat{\alpha}_p = \alpha_p + \hat{n} \tag{8}$$

where

$$\hat{n} = \beta \boldsymbol{d}^T P \boldsymbol{n} \approx N\left(0, \sigma^2_{\mathrm{OSP}}\right) \tag{9}$$

with

$$\sigma^2_{\mathrm{OSP}} = \sigma^2 (\beta \boldsymbol{d}^T P)(\beta \boldsymbol{d}^T P)^T = \sigma^2 \beta \tag{10}$$

and $N(0, \sigma^2_{\mathrm{OSP}})$ is a Gaussian distribution with zero mean and variance σ^2_{OSP}.

It is worth noting that (8) was also derived in [3], [4], [7], and [8]. The $\hat{n}$ in (9) is the noise resulting from the estimation error. As can be seen in (9), there appears a constant β in the noise variance. Now, if $\beta = 1$, (8) is reduced to model (1), where the $\hat{n}$ in (9) becomes the same noise in model (1). In this case, the OSP approach can be thought of as *a priori* version of linear unmixing. For unknown abundance α, the *a priori* OSP classifier given by (4) can be implemented as *a posteriori* OSP classifier using the following *a posteriori* model proposed in [4]

$$\boldsymbol{r} = M\hat{\alpha} + \hat{\boldsymbol{n}} \tag{11}$$

where $\hat{\boldsymbol{n}} = \boldsymbol{r} - M\hat{\alpha}$ is the noise estimate and the true unknown α is replaced with the estimate $\hat{\alpha}$ by (5).

If we define P_F and P_D to be the false alarm probability and the detection power (or the probability of detection), respectively, it is easily shown from [4] and [7] that

$$P_D = 1 - \Phi\left(\Phi^{-1}(1 - P_F) - \frac{\alpha_p}{\sigma\sqrt{\beta}}\right). \tag{12}$$

Based on (12), we can plot an ROC curve of P_D versus P_F and use it to evaluate the effectiveness of the estimator $\hat{\alpha}_p$. For a detailed discussion, refer to [4], [7], and [8]. Equation (12) describes the precise relationship between the β and its associated estimation error, which was not derived in [1] and [3].

From (7), the β is also determined by the correlation between the desired signature $\boldsymbol{d}$ and the OSP projector P. Let the inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ be defined by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y}$. The β given by (7) can be expressed as the inner product of $\boldsymbol{d}$ and $P\boldsymbol{d}$ by

$$\beta = (\boldsymbol{d}^T P \boldsymbol{d})^{-1} = (< \boldsymbol{d}, P\boldsymbol{d} >)^{-1} \tag{13}$$

where $P\boldsymbol{d}$ is the projection of $\boldsymbol{d}$ via the OSP projector P. By virtue of (13), the β is inversely proportional to the correlation between $\boldsymbol{d}$ and $P\boldsymbol{d}$. The less the correlation between $\boldsymbol{d}$ and U, the larger the magnitude $\boldsymbol{d}^T P \boldsymbol{d}$. Thus, the smaller the β, the greater the P_D. This further implies that the estimation method is better. From (13) we also note that

$$0 \le \boldsymbol{d}^T P \boldsymbol{d} \le \boldsymbol{d}^T \boldsymbol{d} \Rightarrow \boldsymbol{d}^T \boldsymbol{d} \le \beta \le \infty. \tag{14}$$

In particular, if $\boldsymbol{d}$ is normalized, i.e., $\boldsymbol{d}^T \boldsymbol{d} = 1$, then $\beta \ge 1$. Equation (14) shows that the OSP classifier specified by (4) produces the maximum detection power and provides an upper bound to (12). This makes sense since (4) assumes *a priori* knowledge about α. So, the only error resulting from the OSP classifier is the classification error, and there is no estimation error involved in (4). On the other hand, any means used to estimate α will generate a constant β, which is a quantitative measure of the estimation error. Increasing the β in (12) will decrease the detection power P_D. This indicates that the error resulting from the MLE not only contains a classification error produced by (4), but also an estimation error incurred by the β. Another interpretation of the β is to view the β as a differential measure of abundance in the classified pixels between the OSP classifier and MLE. A detailed study of this phenomenon along with experiments can be found in [8].

III. Conclusion

In this communication, we showed that the MLE derived in [1] can be considered to be one of the OSP-based methods used for hyperspectral image classification. This is because both the OSP classifier and the MLE produce the same classification feature vector $q^T = \boldsymbol{d}^T P$ given by (4). As a consequence, the resulting classification should be the same. However, there is a significant difference in their magnitudes specified by a constant β. This β is closely related to the error created by the estimation of true abundance. In order to assess the role of the β in abundance estimation, an error analysis was presented in the framework of Neyman–Pearson detection theory. Although the MLE generates the same classifier as does the OSP approach, it does not imply that the MLE and the OSP are the same approach. As a matter of fact, they are designed by completely different concepts. The MLE maximizes the prior probability density function of the observation data, which requires statistics of all orders. It is an *a priori* approach operating on an *a posteriori* model. On the other hand, the OSP classifier derived by a subspace projection approach requires only the second-order statistic of the noise, i.e., covariance matrix. When the noise is assumed to be additive white Gaussian, it is not surprising that they both arrive at the same classification feature vector since a Gaussian distribution can be completely determined by its first two order statistics. When the noise is not Gaussian, the MLE and the OSP may not produce the same classifier. Therefore, the MLE and the OSP approach should be considered to be different methods.

References

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

[2] J. 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.

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

[4] 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 Sensing*, vol. 35, pp. 127–139, Jan. 1997.

[5] *DARPA Spectral Exploitation Workshop*, Defense Advanced Res. Projects Agency, Annapolis, MD, July 1–2, 1996.

[6] *Spectroradiometric Working Group*, Fort Washington, MD, Feb. 3–4, 1997.

[7] X. Zhao, "Subspace projection to multispectral/hyperspectral image classification using linear spectral mixture modeling," M.S. thesis, Dept. Comput. Sci. Elect. Eng., Univ. Maryland, Baltimore County, Baltimore, 1996.

[8] C.-I Chang, "An application of oblique subspace projection approach to hyperspectral image classification," submitted for publication.