# Detection Algorithms for Hyperspectral Imaging Applications

D. Manolakis

7 February 2002

## Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

*LEXINGTON, MASSACHUSETTS*

AOA 399 744

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Plans and Programs Directorate
Contracted Support Management

Massachusetts Institue of Technology
Lincoln Laboratory

# Detection Algorithms for Hyperspectral Imaging Applications

*D. Manolakis*
*Group 97*

Project Report HTAP-8

7 February 2002

Lexington                                           Massachusetts

# ABSTRACT

Detection and identification of military and civilian targets from airborne platforms using hyperspectral sensors is of great interest. Relative to multispectral sensing, hyperspectral sensing can increase the detectability of pixel and subpixel size targets by exploiting finer detail in the spectral signatures of targets and natural backgrounds. A multitude of adaptive detection algorithms for resolved and subpixel targets, with known or unknown spectral characterization, in a background with known or unknown statistics, theoretically justified or ad hoc, with low or high computational complexity, have appeared in the literature or have found their way into software packages and end-user systems. The purpose of this report is twofold. First, we present a unified mathematical treatment of most adaptive matched filter detectors using common notation, and we state clearly the underlying theoretical assumptions. Whenever possible, we express existing ad hoc algorithms as computationally simpler versions of optimal methods. Second, we present a comparative performance analysis of the basic algorithms using theoretically obtained performance characteristics. We focus on algorithms characterized by theoretically desirable properties, practically desired features, or implementation simplicity. Sufficient detail is provided for others to verify and expand this evaluation and framework. A primary goal is to identify best-of-class algorithms for detailed performance evaluation. Finally, we provide a taxonomy of the key algorithms and introduce a preliminary experimental framework for evaluating their performance.

## ACKNOWLEDGMENTS

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

# 1. INTRODUCTION

Remote sensing is defined as the acquisition of information about a distant object without coming into physical contact with it. This is made possible by exploiting the fact that the materials making the various objects reflect, absorb, and emit electromagnetic radiation (photons) in ways characteristic of their chemical composition. If we measure the energy of this radiation as a function of the wavelength, we can obtain a *spectral signature* or simply *spectrum* which can be used to uniquely characterize any given material. The measurement, analysis, and interpretation of such spectra is the subject of spectroscopy. The combination of spectroscopy and imaging technologies and methods to acquire spectral information over large areas is known as *imaging spectroscopy*. The principle underlying imaging spectroscopy is illustrated in Figure 1.
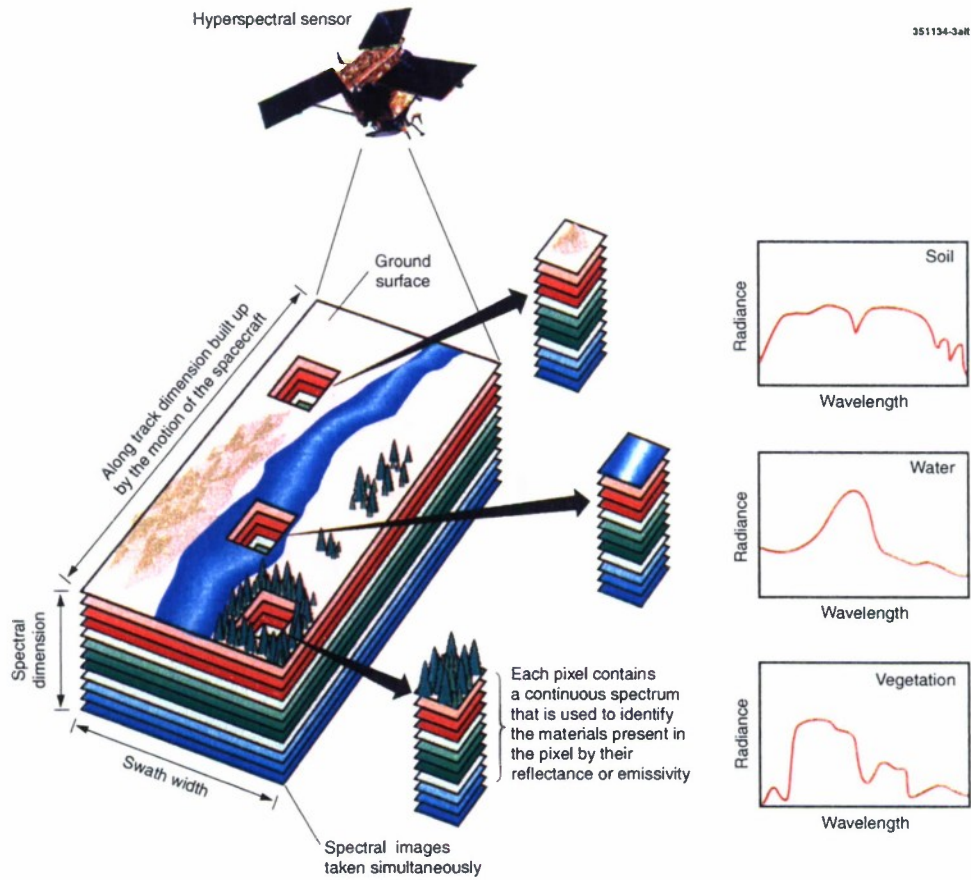


*Figure 1. Principle of imaging spectroscopy*

There are four sampling operations involved in this process: spectral, spatial, radiometric, and temporal. *Hyperspectral* sensors sample the portion of the electromagnetic spectrum that extends form the visible part (0.4-0.7 $\mu$m) to near-infrared (about 2.4 $\mu$m) in hundreds of narrow contiguous bands about 10 nm wide. Figure 2 shows the evolution of the technology from the past low resolution multispectra sensors to the future high resolution ultraspectra ones. Such high spectral resolution preserves important aspects of the spectrum

(e.g., shape of narrow absorption bands) and makes possible to differentiate between different materials on the ground. The spatial resolution or ground pixel size varies from meters to tens of meters and basically is a function of flight altitude, which in turn depends on the kind of platform (spaceborn versus airborn). Radiometric resolution is determined by the number of bits used to describe the radiance measured by the sensor at each spectral channel. Finally, the time interval between successive passes from the same location detrmines the temporal sampling rate.
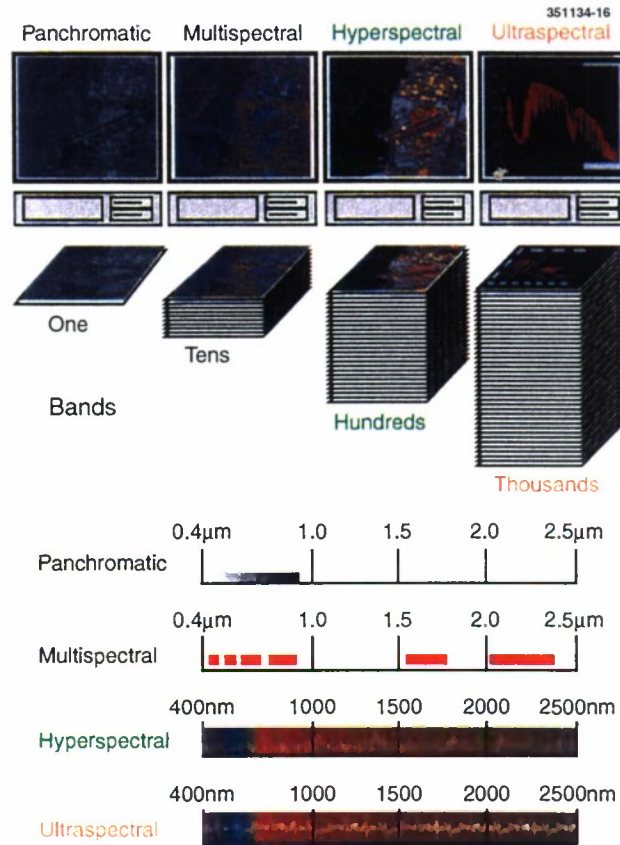


*Figure 2. Evolution of remote sensing spectroscopy with respect to spectral resolution (Adapted from Multispectral Imagery Reference Guide).*

As a result of spatial and spectral sampling, the fundamental hyperspectral data structure is a data cube, whose face is a function of the spatial coordinates and its depth is a function of spectral band (or wavelength). For every band, we have an image of the surface covered by the field of view of the sensor, whereas for each image pixel we have a spectrum characterizing the materials within the pixel. The nature and organization of the collected data is illustrated in Figure 3.

As a result of their fine spectral resolution, HSI sensors provide a significant amount of information about the physical and chemical composition of the materials occupying the pixel surface, as well as the characteristics of the atmosphere between the sensor and the surface during the data collection. Figure 4 summarizes various applications of hyperspectral imaging sensors in terms of the information which can be
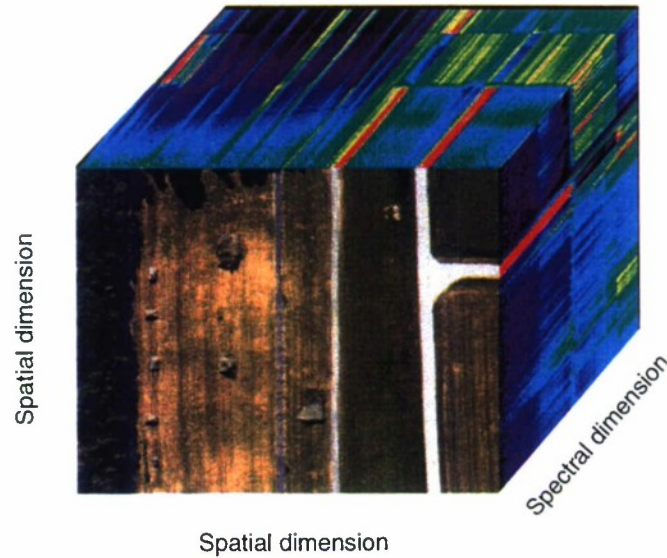
*Figure 3. Imaging spectrometry data cube illustrating the simultaneous spatial and spectral character of the data.*

extracted from different spectral bands. In this report, we focus on the detection of materials using their spectral signatures.

## 1.1 ATMOSPHERIC COMPENSATION

Due to the effects of the illumination source and the atmosphere, the "raw" radiance spectra obtained by an HSI sensor *cannot* be directly compared to either laboratory spectra or "raw" spectra collected at other times or places.

To overcome this obstacle, we work with the reflectance spectrum, which indicates the portion of incident energy which is reflected as a function of wavelength. Hence, the properties of the illuminating source and the effects of the propagating atmosphere are removed, and the shape of the reflectance curve is characteristic of the materials in the observed pixel. Once the data have been corrected for the effects of the atmospheric absorption and scattering, the resulting reflectance spectrum for every pixel, can be compared to spectra of known materials available in "spectral libraries".

## 1.2 SPECTRAL VARIABILITY AND MIXED PIXELS

The basic task underlying many HSI applications is to identify different materials based on their reflectance spectrum. In this respect, the concept of a *spectral signature*, which uniquely characterizes any given material, is highly attractive and widely used. However, spectra observed in the natural world do not exhibit a deterministic signature. The spectra observed from samples of the same material will never be identical, even in laboratory experiments, due to variations in the material surface. The amount of variability is more profound in remote sensing applications due to variations in atmospheric conditions, sensor noise, material composition, location, surrounding materials, and other factors. To make matters worse, totally

3

# Spectral Region Utility
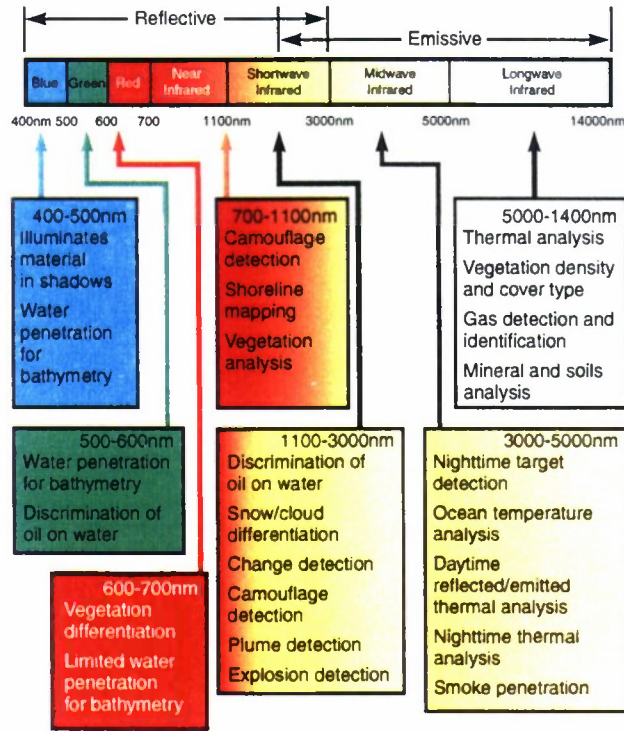


*Figure 4. Applications of hyperspectral image exploitation according to the utility of different spectral bands (Courtesy of SITAC).*

different material types can have very similar spectra.

Despite these difficulties, practical experience has shown that many materials of interest can be identified on the basis of their spectral characteristics. However, the ambiguity introduced by inherent variability of spectral signatures has important implications into the exploitation of HSI data for both civilian and military applications.

Despite the intrinsic spectral variability and the occasional lack of identifiability, the concept of spectral signature is widely used in remote sensing spectroscopy. In this paper, we assume that different materials are spectrally separable and focus on the problems introduced by the inherent variability of spectral signatures.

Another significant complication arises from the interplay between the spatial resolution of the sensor and the spatial variability present in the ground scene. The sensor integrates the radiance from all materials within the ground surface "seen" by the sensor as an image pixel. Therefore, depending on the spatial resolution of the sensor and the spatial distribution of surface materials within each ground pixel, the result is a HSI data set comprised of "pure" and "mixed" pixels. Mixed pixels present an additional challenge to HSI

data exploitation because their spectral signatures do not correspond to any single well-defined material.

Dealing with spectral signature variability and spectral compositions in mixed pixels, are among the most challenging problems in HSI data exploitation, both theoretically and practically.

## 1.3   BACKGROUND CLASSIFICATION AND TARGET DETECTION

There are two major applications that rely upon the ability to separate materials based on their spectral signatures: *background classification* and *target detection.*

The main objective of background classification is to automatically assign all pixels in an HSI data cube into land cover classes or themes, which has led to the term thematic mapping. The user has the task to up-front determine the number and type of classes as well as to quantitatively characterize these classes using spectral libraries or training data and ground truth information. Practical experience has shown that, the design of a good classifier requires a sufficient amount of training data for each background class. Clearly, for background classification, the natural criterion of performance is the minimization of the probability of missclassification errors.

In target detection applications, the main objective is to search the pixels of an HSI data cube for the presence of a specific material (target). Conceptually, at least at a theoretical level, target detection can be viewed as a two-class classification problem. However, there are some fundamental practical differences, that have a great impact upon the development and evaluation of practical algorithms for detection versus classification applications. In surveillance applications, the size of the objects (targets) we are searching for constitutes a very small fraction of the total search area. Therefore, the target class will be either empty or sparsely populated. On the other hand, the general "no-target" class includes almost all pixels in the cube and is the union of the different specific background classes. We shall use the term "background" to refer to all non-target pixels of a scene. Usually, targets are man-made objects with spectra that differ from the spectra of natural background pixels.

To summarize, the key aspects of detection applications are:

1. the target class is either empty or sparsely populated

2. the amount of available a priori information, for the spectral characterization of target class, varies depending upon the application from complete to none

3. the background class is densely populated

4. there is insufficient a priori information for adequate spectral characterization of the background class.

The sparseness of the target class implies that there are not sufficient data to train a statistical classifier or statistically evaluate the performance of a target detector. On the other hand, the heavy population of the background class, in conjunction with the emptiness of the target class, allows the use of the "unclassified" HSI cube to statistically characterize the background. In detection applications, where the target probability is very small, minimization of the error probability is not a good criterion of performance, because it can be minimized by classifying every pixel as background. For this reason, we typically seek to maximize

the probability of detection while keeping the probability of false alarm under a certain predefined value (Neyman-Pearson criterion).

The amount of a priori information about the spectral signature of the target, depends on the requirements of the specific application. A priori information about spectral signatures is available in libraries as reflectance spectra. Therefore, to look for targets with known spectral signatures, the data must first be converted into reflectance, a procedure that may lead to spectral distortions since it generally depends upon assumptions and measurements about atmospheric conditions. If we have no a priori information about the target or we are constrained to work with radiance, the most reasonable approach is to look for pixels whose spectral content is "significantly" different from the spectral content of the local background. The detection of targets with unknown anomalous spectra is known in the hyperspectral literature as *anomaly detection* and is discussed in a companion paper.

From an applications point of view, there are different types of spectral searches that can be implemented:

- determine whether a reflectance spectrum from the HSI cube exists in the spectral library (material identification)

- determine whether a spectrum from a library is present in the HSI cube (target detection)

- identify groups of spectrally similar pixels from the same cube (spectral classification)

- identify pixels with anomalous spectra compared to their local background (anomaly detection)

In this paper, we discuss algorithms for target detection, for the case where there is a priori knowledge available about the target spectral signature. In the radar and communication areas, detection algorithms are used to decide whether the received waveform consists of "noise only" or "signal masked by noise". Typically, there is sufficient a priori information about the transmitted signal and the performance of the system is limited by additive noise. In contrast, data obtained by HSI sensors exhibit a high signal-to-noise ratio (SNR) and the performance of detection algorithms is limited by target variability rather than noise. In mixed pixels, the target spectrum can be severely masked by dominating background spectra. In this case, detection performance is significantly limited by "background interference." This intuitively obvious observation leads to a fundamental partitioning of detection algorithms into two classes:

- Detection algorithms for full-pixel targets: in this case detection performance is mainly determined by target and background variability.

- Detection algorithms for sub-pixel targets: in this case, besides the variability of target and background spectra, detection performance is affected by background interference.

## 1.4  DESIGN AND EVALUATION OF TARGET DETECTORS

The mathematical framework for the design and evaluation of detection algorithms is provided by the area of statistics known as (binary) hypothesis testing. There are several approaches for the systematic design

of detection algorithms. However, it is well known that detectors based on the likelihood ratio (LR) test have certain advantages. First, in some sense, LR tests minimize the risk associated with incorrect decisions. Second, the LR leads to detectors which are optimum for a wide range of performance criteria, including the maximization of separation between target and background pixels.

Most HSI data processing techniques start with the idea that an observed spectrum can be considered as a vector in a multidimensional space, where the number of dimensions equals the numberof spectral bands, $L$ . Taking into consideration spectral variability and receiver noise, the observations provided by the sensor can be modelled, for the purpose of theoretical analysis, as random vectors with specific probability distributions. Given an observed spectrum, $x$, the LR is given by the ratio of the conditional probability density functions

$$\Lambda(x) \triangleq \frac{p(x \mid \text{signal present})}{p(x \mid \text{signal absent})} \tag{1}$$

If $\Lambda(x)$ exceeds a certain threshold $\eta$, then the "signal present" hypothesis is selected as true. If $\Lambda(x)$ is larger than the threshold, the "signal present" hypothesis is accepted. Basically, the LR test accepts as "true" the most "likely" hypothesis.

A practical question of paramount importance to a system designer is where to set the threshold to keep the number of detection errors (target misses and false alarms) small. Indeed, there is always a compromise between choosing a low threshold to increase the probability of (target) detection $P_D$ and a high threshold to keep the probability of false alarm $P_{FA}$ low. For any given detector, the trade-off between $P_D$ and $P_{FA}$ is described by the *Receiver Operating Characteristic* (ROC) curves, which plot $P_D(\eta)$ versus $P_{FA}(\eta)$ as a function of threshold $-\infty < \eta < \infty$. Clearly, any systematic procedure to determine ROC curves or the threshold requires specifying the distribution of the observed spectra $x$ under each of the two hypotheses.

If the conditional densities in (1) are completely known (simple hypotheses), the detector specified by the LR has the highest possible $P_D$ for any value of $P_{FA} \leq \alpha$ (Neyman-Pearson Lemma). Hence, the ROC curve of the optimum Neyman-Pearson detector provides an upper bound for the ROC of any other detector. It is also possible to choose the threshold in a way that leads to the minimization of the probability of detection errors (both misses and false alarms). This leads to the well known Bayes detector or classifier, which is widely used in pattern classification applications. It should be emphasized that the Bayes and Neyman-Pearson detectors are specified by the same LR function; they only differ in the selection of the threshold.

In most practical situations, the conditional probability densities in (1) depend on some unknown target and background parameters (composite hypotheses). Therefore, the ROC curves of any detector depend on the unknown parameters. In this case, is almost impossible to find a detector whose ROC curves remain an upper bound for the whole range of the unknown parameters (uniformly most powerful (UMP) detector).

An intuitively appealing and widely used approach, in the case of unknown density parameters, is to replace the unknown parameters in the LR (1) with their maximum likelihood estimates. In general, there are no optimallity properties associated with the resulting Generalized LR (GLR), $\Lambda_G(x)$. However, in practice, the GLR leads to detectors that seem to work well in several applications.

Practical target detection systems should function automatically, that is, without operator intervention. This requires an automatic strategy to set a "proper" detection threshold. A high false alarm rate wastes

processing and reporting resources and may result to system overloading. Therefore is critical to keep the false alarm rate constant at a desirable level by using a Constant False Alarm Rate (CFAR) processor. The task of a CFAR algorithm is to provide detection thresholds that are relatively immune to noise and background variation and allow target detection with a constant false alarm rate.

## 1.5    FRAMEWORK FOR DETECTION ALGORITHM TAXONOMY

The key factors that determine the taxonomy of hyperspectral target detection algorithms are: the type of models used for spectral (target or background) variability , the composition of the pixel under test (pure or mixed), and the model used to describe mixed pixels.
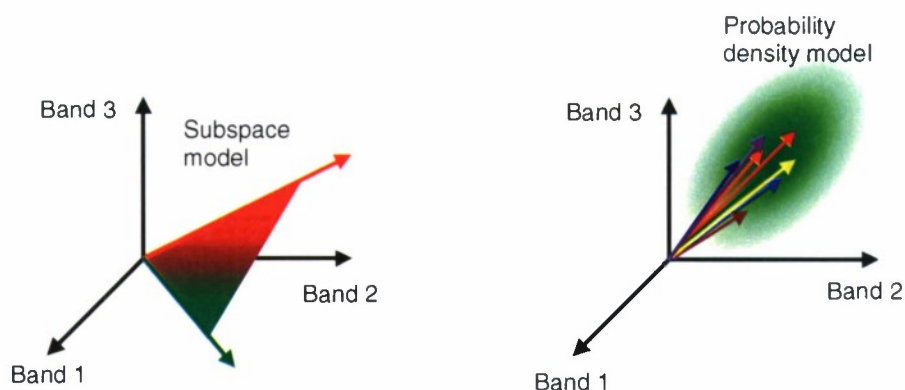


*Figure 5. Pictorial illustration of subspace and probabilistic models for the description of spectral variability in the spectral space.*

The observed spectral radiance data, or derived apparent surface reflectance data, can be viewed as scattering of points in an $L-$dimensional Euclidean space, where $L$ is the number of spectral bands. Each spectral band is assigned to one axis of the space, all axes being mutually orthogonal. Therefore, the spectrum of each pixel can be viewed as a vector. The tip od this vector corresponds to a point, whose Cartesian coordinates are the values at each spectral band (see Figure 5). Spectra without variability correspond to a single fixed point, whereas the tip of vector corresponding to spectra with variability can be anywhere within a certain volume of the spectral space. Depending on how we specify thsi space, there are two widely used ways to describe spectral variability.

The *geometric approach* restricts the spectrum vector to vary in an $M$-dimensional subspace of the data space ($M < L$). The observed spectrum is described by

$$x = \sum_{k=1}^{M} a_k s_k = Sa \tag{2}$$

The vectors $s_k$ or equivalently the matrix $S$, which define the variability subspace, can be (a) endmembers determined from spectral libraries or the data, or (b) vectors obtained with statistical techniques (for example,

the eigenvectors of the data correlation matrix). Clearly, the variability increases as $M$ increases from one to $L$.

The *statistical approach* provides a probability distribution model for the description of the spectral variability. Usually, first- and second-order moments (mean vector and covariance matrix) are employed under a multivariate normal distribution assumption. Clearly, variability is related to the spread of the distribution, and the highest variability is obtained for a uniform distribution over the data space.

For full pixel targets, there is no significant interaction between target and background other than secondary illumination, shading, etc. Hence, the spectrum observed by the sensor is produced either by the target spectrum or the background spectrum. In both cases, the observed spectrum is corrupted by additive sensor noise. The sensor noise is assumed insignificant or accounted for by the target and background distributions.

For subpixel targets, both the spectrum of the target and the spectrum or spectra of the background contribute to the observed mixed pixel spectrum. There are two widely used models for modelling subpixel targets.

The most widely used spectral mixing model is the linear mixing model [1] (LMM), which assumes that the observed pixel spectrum is generated by a linear combination of a small number of unique constituent deterministic spectral signatures known as "endmembers". The mathematical formulation of the LMM is given by

$$x = \sum_{k=1}^{M} a_k s_k + w = Sa + w \tag{3}$$

where $s_1, s_2, \ldots, s_M$, are the $M$ endmember spectra, assumed linearly independent, $a_1, a_2, \ldots, a_M$ are the corresponding abundances, and $w$ is an additive noise vector. Endmembers may be obtained from spectral libraries, in-scene spectra, or using geometrical techniques. We point out that the enforcement of positivity ($a_k \geq 0$) and additivity ($a_1 + a_2 + \cdots + a_M = 1$) constraints makes the LMM a replacement model. Spectra satisfying the LMM with both sets of constraints are confined in an $L$-dimensional simplex [1] studied by the mathematical theory of convex sets.

If the endmember spectra are randomly and independently drawn from multivariate normal distributions, we have the stochastic mixing model [53, 60].

The choice of a pixel composition assumption (pure or mixed pixel), the selection of a model to account for spectral variability (subspace or probability distribution), and the selection of a mixing procedure leads to different types of target detection algorithms. The detection problem is typically formulated as a binary hypothesis test with two competing hypotheses: background only ($H_0$) or target and background ($H_1$). Since the two hypotheses contain unknown parameters (for example, covariance matrix of the background) that have to be estimated from the data, the detector has to be adaptive, and it is usually designed using the generalized likelihood ratio test (GLRT) approach.

Most detection algorithms for full pixel and subpixel targets have been obtained by describing spectral variability using the multivariate normal distribution or a subspace model. Mixed pixels are usually modelled using the LMM. A target detection algorithm based on the stochastic mixing model, known as finite target matched filter, is discussed in [53, 60].

Finally, we note that, in several practical applications we do not have adequate *a priori* information about the desired target. In such cases, it is possible to design algorithms that look for spectra which deviate from the local background (anomaly detection). The type of the statistical model used for the background leads to different anomaly detection algorithms. Use of a multivariate normal distribution model leads to the RX algorithm [45, 63] which is one of the most widely used algorithms for anomaly detection. Recently, a new algorithm [58] has been developed, which fuses the local statistics used by the RX algorithm and the clustering statistics obtained using stochastic expectation maximization (SEM) to improve detection performance. More details about anomaly detection algorithms can be found in [57].

Table 1 provides a conceptual taxonomy of target detection algorithms according to the adopted signal model and the available a-priori information. We note that both target and background variability may be modelled using either a multivariate normal distribution or a subspace model. However, most practical algorithms for target detection use subspace models to account for target variability.

*TABLE 1. Taxonomy of algorithms for target detection.*

| A-priori information | Quantity | Matched filter | Clairvoyant | Adaptive | Anomaly detection |
|---|---|---|---|---|---|
| Target subspace | $S$ | Known | Known | Known | Unknown |
| Target abundance | $a$ | Known | Unknown | Unknown | Unknown |
| Background and noise covariance | $\Gamma$ | Known | Known | Unknown | Unknown |
| Background subspace and noise variance | $B, \sigma_w^2$ | Known | Known | Unknown | Unknown |

The rest of the report is organized as follows. Section 2 provides a brief description of the HSI data sets used for the experimental investigations and the various practical issues regarding the implementation and evaluation of target detection algorithms. The mathematical modelling of spectral variability for target and background pixels is the subject of Section 3. In Section 4, we discuss detection algorithms for full-pixel (or resolved) targets. Detection of sub-pixel targets is presented in Section 5. The linear mixing model and the estimation of its parameters using the principle of least squares is discussed in Section 6. The application of the linear mixing model to sub-pixel target detection is the subject of Section 7. Finally, we present a taxonomy of target detection algorithms in Section 8 and a brief conclusion in Section 9.

## 2. DATA DESCRIPTION AND PRACTICAL CONSIDERATIONS

Airborne hyperspectral imagery data collected by the HYDICE [48] sensor at the U.S. Army Aberdeen Proving Grounds on 24 August 1995 will be used to illustrate various issues and algorithms discussed in this report. HYDICE collects calibrated (after processing) spectral radiance data in 210 wavelengths spanning 0.4 to 2.5 $\mu m$ in nominally 10 nm wide bands. Figure 6 shows a single band ($\lambda = 0.565 \mu m$) image of the Run 07 data collected at 9:27 AM local time under clear conditions from an altitude of 3 km. The spatial resolution of the imagery is approximately 1.5 meters.



*Figure 6. Run 07 data and three regions used for statistical analysis and detection algorithm evaluation.*

The implementation of hyperspectral target detection algorithms in a real-world environment involves confrontation with many "practical details" and challenges that result from the violation of the theoretical assumptions used for the derivation of the various algorithms.

To illustrate various issues regarding detection algorithms, we shall use the Forest Radiance I data

(see Figure 6) collected with the HYDICE sensor. We have selected the three areas outlined in Figure 6 to investigate three different types of background: grass (G), trees (T), and mixed grass-road (GR). The first two scenes are relatively homogeneous, whereas the third scene consists of a non-homogeneous background including different types of grass and roads. Also considered were classes resulting from a supervised classification process performed to isolate spectrally similar (not necessarily spatially adjacent) pixels. Data from two classes selected from this analysis were labelled "Class 2 Grass" and "Class 9 Tree". Table 2 summarizes the classes analyzed and their sample sizes (number of pixels).

*TABLE 2. Classes selected for statistical analysis.*

| Class name | Selection Technique | Sample Size |
|------------|--------------------|-------------|
| Grass | Spatially adjacent | 7,760 |
| Tree | Spatially adjacent | 8,232 |
| Mixed | Spatially adjacent | 7,590 |
| Class 2 Grass | Supervised classification | 27,351 |
| Class 9 Tree | Supervised classification | 25,872 |

The data were analyzed in units of calibrated spectral radiance for the characterization of joint statistics (Section 3.3) and surface reflectance for the modelling of target detection statistics (Section 4.4). While hyperspectral data is often transformed to apparent surface reflectance through an atmospheric compensation algorithm, such transformations are usually linear, and as such, do not affect the statistical distributions. Also, for the multivariate analyses examples, only 144 of the 210 channels were used to avoid regions of low signal-to-noise ratio (water vapor bands). While it is known that data artifacts exist in the calibrated spectral radiance due to an interpolation process used to replace "dead" pixels, only the data selected by the supervised classification technique had any type of screening applied for anomalous pixels. The other classes represent typical data as it would be processed by an unsupervised automated algorithm.

Regarding the characterization of in-scene targets, single-pixel or multi-pixel, we can identify pure or mixed pixel with a certain level of confidence, only. Therefore, it is useful to label the pixels in the vicinity of a target as full, mixed, shaded, and guard pixels, and distinguish among such pixels when we compare different detectors. This concern, has lead to the development of target masks (see Figure 7) by an elaborate manual process. Target masks are part of the HYDICE "Canonical Data Sets" developed at Lincoln Laboratory.

Figure 8 shows the reflectance spectra of the pixels specified by the target mask in Figure 7. The mean value of the full pixel spectra, which is used as the target template by detection algorithms, is shown by a thick line in all plots.

A target detector maps the multidimensional test pixel spectrum $x$ into a scalar detection statistic $y = D(x)$ (see Figure 9a). The detection threshold is determined by the user or a CFAR processor. The distribution of the detection statistics of the background and target classes (see Figure 9b) determine the performance of the detector using ROC curves. The ability to properly determine a threshold or achieve CFAR

12
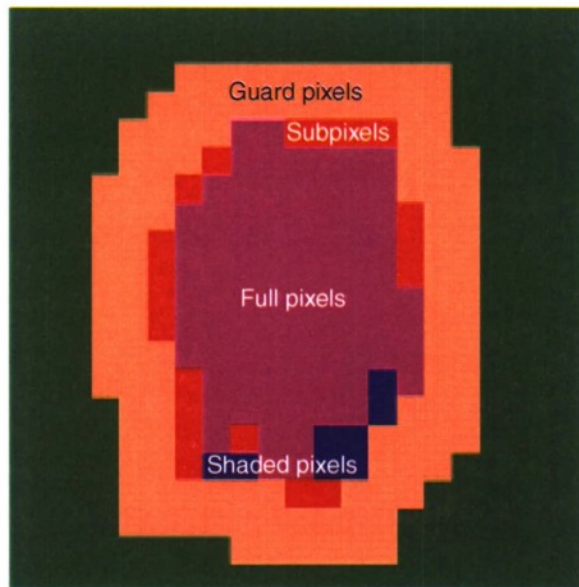
*Figure 7. Example of a target mask, illustrating the various types of pixels identified in the canonical data sets.*

operation depends upon the accurate modelling of the background detection statistics. The performance evaluation of detection algorithms in practice is challenging due to the limitations imposed by the limited amount of target data. As a result the establishment of accurate ROC curves is quite difficult. Indeed, it is well known, that as a rule of thumb, the minimum number $N$ of pixels used to estimate a probability $P$ should be $10/P$ or more preferable $100/P$. In this report, we shall compare the various algorithms in terms of their ability to operate in CFAR mode and the enhancement of the separation between targets and background they provide. Sensor and environmental noise do not seem to be significant factors.

The CFAR property depends on the capability to accurately model the detection statistics of the background pixels for a given algorithm. In this respect, we use a quantile-quantile (Q-Q) plot [17] to compare the empirical distribution of the background detection statistics to the theoretically expected one.

Taking into consideration that the number of target pixels is much smaller than the number of background pixels, a useful way to represent the output of any target detector is shown in Figure 11. The key idea is to represent the background response by its histogram and the response of the various "target" pixels by stems, whose location indicates the magnitude of the detector output and the type of the pixel. The enhancement of target visibility is characterized by the separation between the upper limit of the background histogram and the full target pixel with the smallest detection statistics response.

13
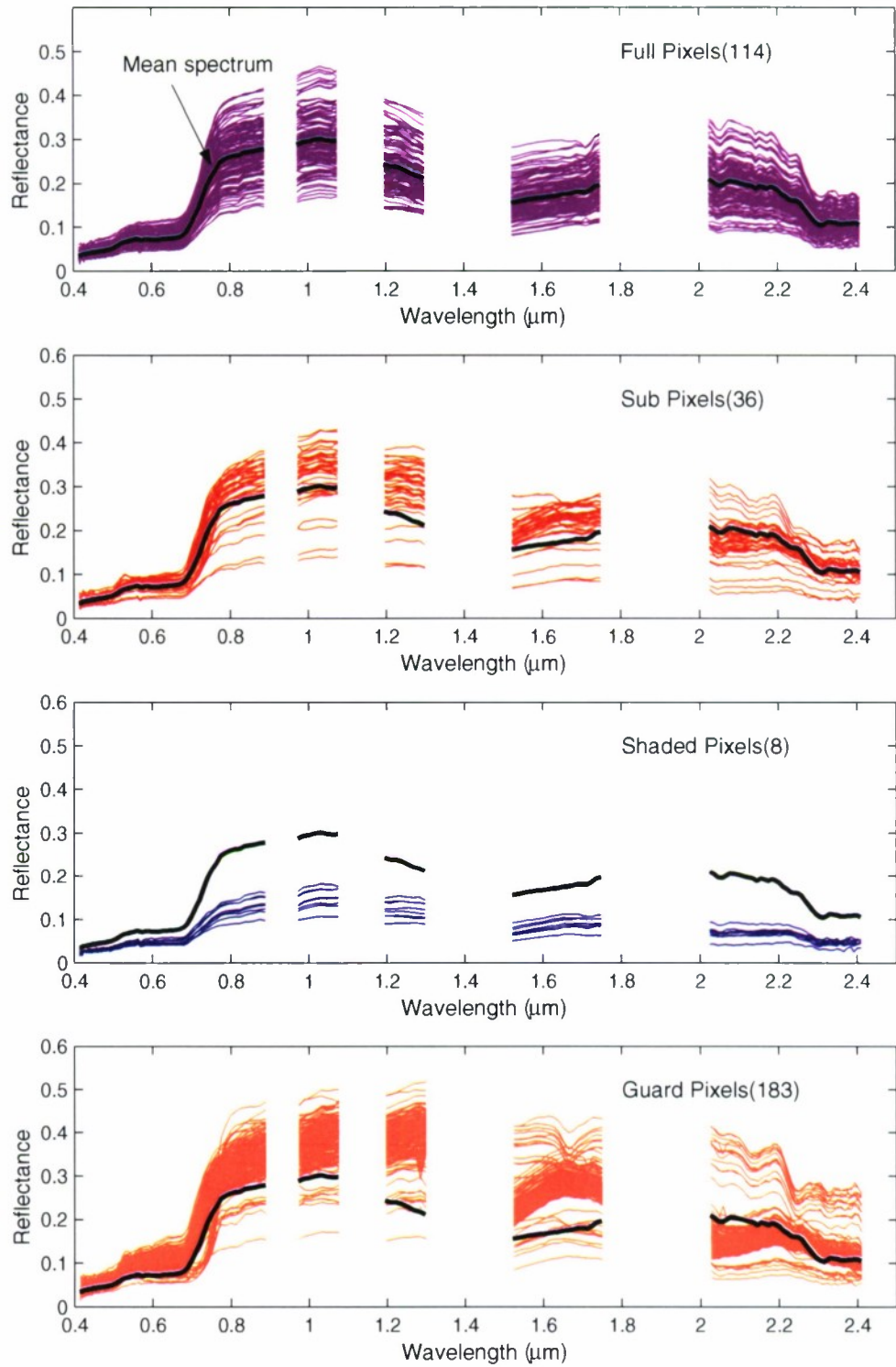
*Figure 8. Spectra of the pixels specified by a target mask. The thick line, shown in all plots, it is the mean spectrum of the full target pixels.*
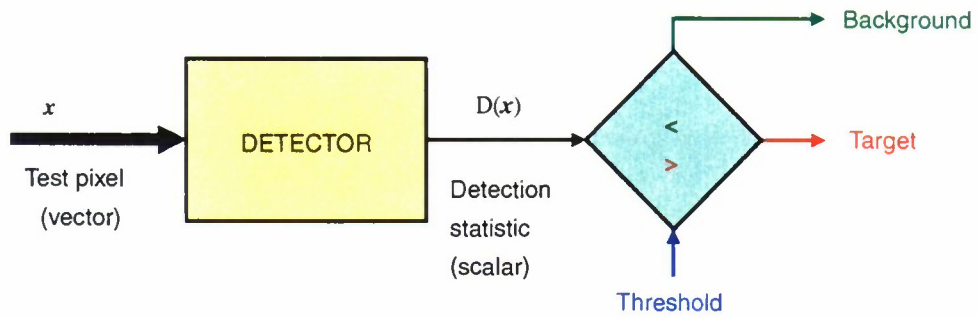
(a)
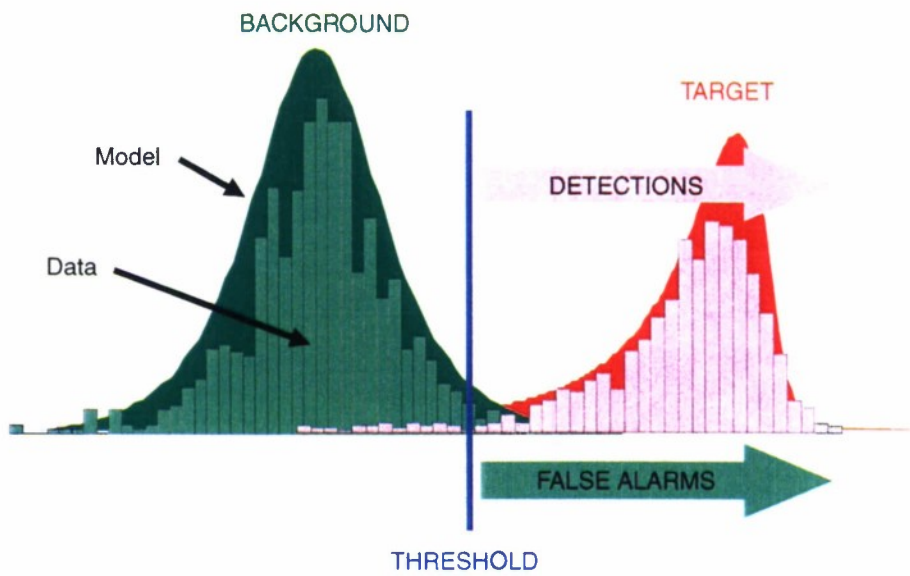


(b)

*Figure 9. (a) Every detector maps the spectrum **x** of the test pixel (multidimensional vector) into a scalar detection statistic y = D(**x**). (b) Modelling the background detection statistic is important for CFAR operation.*

(a)                                                  (b)
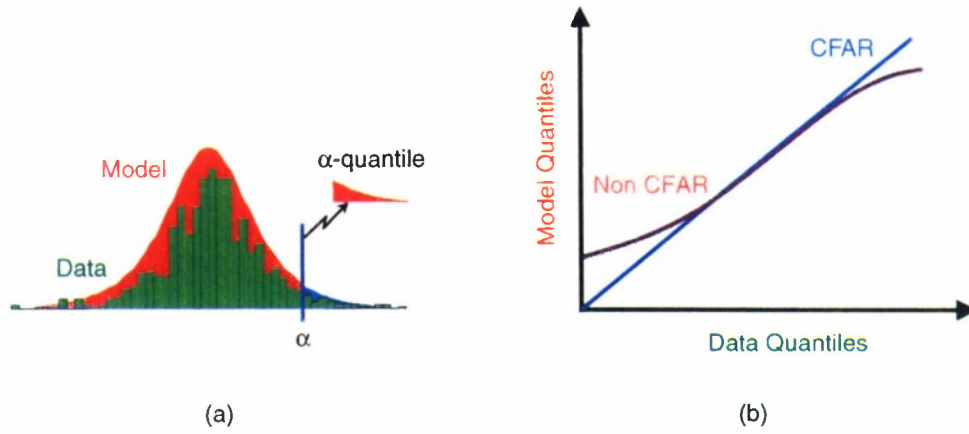
*Figure 10. Use of Q-Q plots for evaluating background detection statistic for potential CFAR operation.*
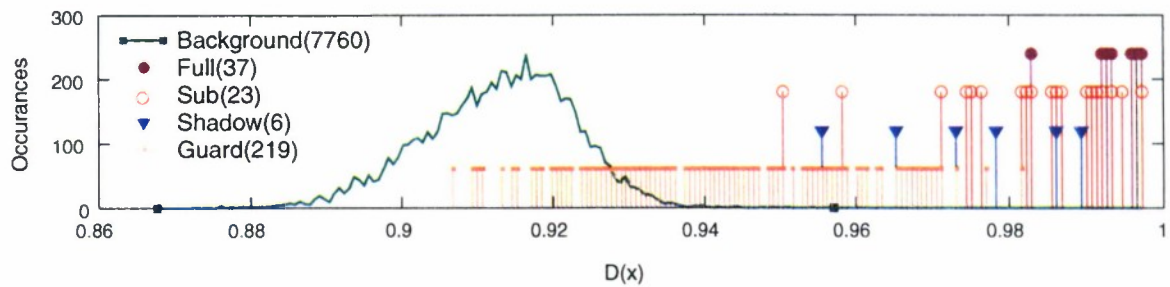


*Figure 11. Representation of the response of a target detector to background and different types of target pixels.*

16

# 3.  MODELING SPECTRAL VARIABILITY

In examining the statistical properties of the data, several groupings, or classes, were considered. Three regions are identified in the white boxes in Figure 6 describing three classes that were selected by their spatial proximity. In the lower right is a "Grass" region, the middle top is a "Tree" region, and on the left is a "mixed" region. These regions define the pixels selected for three of the classes considered. Also considered were two classes resulting from a supervised classification process performed to isolate spectrally similar (not necessarily spatially adjacent) pixels.

## 3.1  SINGLE WAVEBAND STATISTICS

Before discussing the joint, or multivariate, statistics, it is worthwhile to examine some of the scalar, or marginal, statistics. Figure 12 presents Q-Q plots for two of the classes and one spectral channel, showing the cumulative probability distribution as a function of the data value, overlaid with straight lines representing Gaussian data. As one can see, neither class would be well represented by the Gaussian assumption, although the class selected through the supervised classification process is slightly better matched to the Gaussian assumption. Additional insight into single waveband statistics can be found in [62]. Since, detection and classification applications mainly depend upon the joint distribution of the data,we are not going to further investigate single waveband statistics in this report.
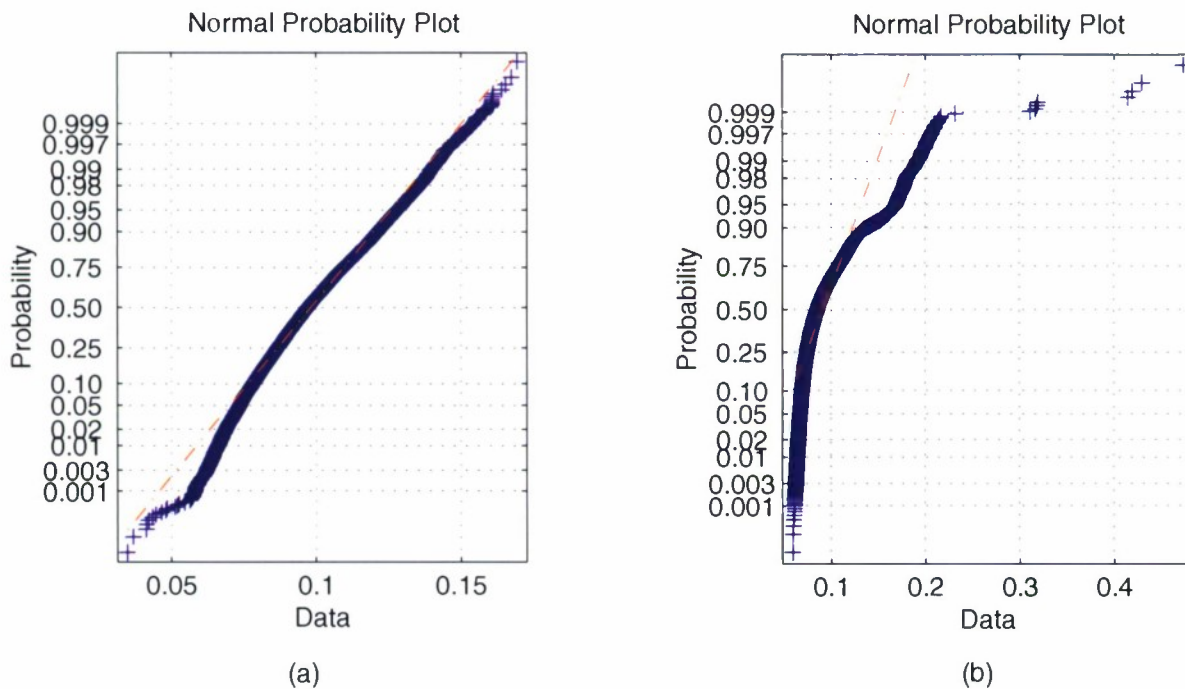


*Figure 12. Quantile-Quantile (Q-Q) plots of HYDICE data from band 39 ($\lambda = 0.565\mu m$) for (a) Class 2 Grass and (b) Mixed background data. Also shown are straight lines corresponding to the normal distributions.*

## 3.2  RANDOM VECTORS WITH ELLIPTICALLY CONTOURED (EC) DISTRIBUTIONS

The class of elliptically contoured (EC) distributions contains distributions that have similar features to the multivariate normal distribution, but which exhibit either longer or shorter tails than the normal. In this section, we consider the class of scale or compound mixtures of normal distributions, which is a subclass [10] of EC distributions.

A random real vector $x = [x_1 x_2 \ldots x_p]^T$ with mean $\mu$ and covariance matrix $\Gamma$ has an EC distribution if and only if its PDF has the form [10]

$$f_p(x) = (2\pi)^{-p/2} |\Gamma|^{-1/2} h_p(d) \tag{4}$$

where $d$ is a quadratic form (Mahalanobis distance) defined by

$$d = (x - \mu)^T \Gamma^{-1} (x - \mu) \tag{5}$$

and $h_p(d)$ is a positive, monotonically decreasing function for all $p$. We shall denote such a distribution using the shorthand notation $EC(\mu, \Gamma, h)$.

The multivariate normal distribution (MVN) is a special case with

$$h_p(d) = \exp\left(-\frac{1}{2} d\right) \tag{6}$$

In addition, the EC class includes the multivariate $t$, the multivariate Cauchy, and the double exponential distributions. The discrete normal (Gaussian) mixture, which is widely used in supervised or unsupervised pattern recognition algorithms, is a special case of EC distributions. The normal mixture PDF is a simple finite weighted sum of normal PDFs and can be used for the approximation of many other ECDs. The common feature of all ECDs that they all have elliptical contours of equiprobability. Many of the properties of the multivariate normal distribution extend also to the class of EC distributions. Thus, all the marginals of an EC distribution, and the conditional distributions of some variables, given the values of the others, are also EC.

Any ECD, $x$, can be represented by

$$x = \Gamma^{1/2}(vz) + \mu \tag{7}$$

where

$$z \sim N(0, I) \tag{8}$$

and $v$ is a non-negative random variable independent of $z$. The type of an ECD is uniquely determined by the PDF $f_v(v)$ of $v$, which is known as the *characteristic PDF* of $x$. The random variable $v$ is normalized so the $E\{v^2\} = 1$, that is, to unit mean squared error. Clearly, $f_v(v)$ and $\Gamma$ can be specified independently.

If we have an analytical expression for $f_v(v)$, the class of admissible $h_p(d)$ functions is obtained by

$$h_p(d) = \int_0^\infty v^{-p} \exp\left(-\frac{1}{2} \frac{d}{v^2}\right) f_v(v) dv \tag{9}$$

18

The PDF of Mahalanobis distance $d$ is given by

$$f_p(d) = \frac{1}{2^{p/2}\Gamma(p/2)} d^{p/2-1} h_p(d) \tag{10}$$

and can be used to *uniquely* reduce the multidimensional PDF modeling problem to an equivalent one-dimensional one. If $x(n), \ldots, x(N)$ form a random sample from a $N(\mu, \Gamma)$ population, then

$$d(n) = [x(n) - \mu]^T \Gamma^{-1} [x(n) - \mu] \tag{11}$$

for $n = 1, \ldots, N$ correspondingly form a random sample from a $\chi_p^2$ distribution with $p$ degrees of freedom.

If the PDF $f_v(v)$ is not available in closed form, we can generate SIRV's and determine $h_p(d)$ by expressing $x$ in spherical coordinates $r, \theta$, and $\phi_1, \phi_2, \ldots, \phi_{p-2}$. The PDF of $r$ is given by

$$f_p(r) = \frac{r^{p-1}}{2^{p/2-1}\Gamma(p/2)} h_p(r^2) u(r) \tag{12}$$

The angles $\theta$ and $\phi_k$ are independent of the radius $r$ and they do not affect the type of the SIRV.

One of the most important implications of (10) is that the multivariate PDF for any ECD is uniquely determined by the univariate PDF of the Mahalanobis distance (5). As a result, the multivariate PDF identification problem is reduced to a simpler univariate one. This result provides the cornerstone for our investigations in the statistical characterization of HSI data. Several researchers [43,44] have developed libraries and generation techniques for ECDs specified by their characteristic PDF or $h_p(d)$ that can be applied to our objective of modelling unknown hyperspectral backgrounds.

### 3.3 MULTIPLE (JOINT) WAVEBAND STATISTICS

Since the distribution of multivariate data in the observation space is inherently "sparse", it is impractical to use goodness-of-fit tests to evaluate the capacity of the multivariate normal distribution or ECDs to characterize HSI data. In practice, we can assess multivariate normality using techniques from the following categories [17]: (a) Univariate techniques for evaluating marginal normality, (b) Multivariate techniques for evaluating joint normality, and (c) Techniques that use unidimensional views (projections) of multivariate data. These techniques can be extended for the characterization of elliptical [10,43,44] distributions.

If we know the mean and covariance of a multivariate random sample, we can check for normality by comparing the distribution of the Mahalanobis distance (11) against a chi-squared distribution. However, in practice, we have to estimate [21,29] the mean and covariance from the available data. To this end, we use the maximum likelihood estimates of the mean

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x(n) \tag{13}$$

and the covariance matrix

$$\hat{\Gamma} = \frac{1}{N} \sum_{n=1}^{N} [x(n) - \hat{\mu}][x(n) - \hat{\mu}]^T \tag{14}$$

19

to compute the Mahalanobis distance

$$d(n) = \frac{1}{N-1}[x(n) - \hat{\mu}]^T \hat{\Gamma}^{-1}[x(n) - \hat{\mu}] \tag{15}$$

for all pixels $1 \le n \le N$. If $x(n) \sim \mathcal{N}(\mu, \Gamma)$, it can be shown that $d(n)$ is distributed according to the beta distribution

$$p_d(d) = \frac{1}{B(a,b)}d^{a-1}(1-d)^{b-1}, \quad 0 \le d \le 1 \tag{16}$$

where the parameters

$$a = \frac{p}{2} \quad \text{and} \quad b = \frac{N-p-1}{2} \tag{17}$$

When the sample size, $N$, is greater than about $10p$, the difference between the true beta distribution and the chi-squared approximation, $\chi_p^2$, is negligible for our purposes. Therefore, $\chi^2$ probability plots provides a good graphical method for testing multivariate normality.

For EC random vectors, the distribution of the Mahalanobis distance completely characterizes their joint PDF. Therefore, to test whether an $EC(\mu, \Gamma, h)$ fits the data, we follow three steps: (a) compute the Mahalanobis distances (15), (b) compute the expected PDF (10), and (c) compare the two distributions using a probability plot. However, for detection applications, where the key goal is accurate prediction of the probability of false alarm, it is more interesting to compare the empirical and theoretically predicted probabilities of false alarms. It is interesting to note that the Mahalanobis distance (15) is the statistic used by the popular RX algorithm [45, 58] for anomaly detection.

From the multitude of ECDs discussed in statistics literature, the EC $t$-distribution [10] has been shown to provide a good model for many HSI data sets [40]. This distribution is defined by

$$t_L(x; \mu, C, v) = \frac{\Gamma[(L+v)/2]}{\Gamma(v/2)(vL)^{L/2}|C|^{1/2}} \left[1 + \frac{1}{v}(x-\mu)^T C^{-1}(x-\mu)\right]^{\frac{L+v}{2}} \tag{18}$$

where

$$E\{x\} = \mu \tag{19}$$

$$Cov\{x\} = \frac{v}{v-2}C \tag{20}$$

where $v$ is the number of degrees of freedom and $C$ the scale matrix. The Mahalanobis distance is distributed as

$$\frac{1}{L}(x-\mu)^T C^{-1}(x-\mu) \sim F_{L,v} \tag{21}$$

where $F_{L,v}$ is the F-distribution with $L$ and $v$ degrees of freedom. The integer $v$ controls the tails of the distribution: $v = 1$ leads to the multivariate Cauchy distribution (heavier tails), whereas as $v \to \infty$ the EC $t$-distribution approaches the multivariate normal distribution (lighter tails).

20

When the covariance matrix is estimated from the data, the distribution of the Mahalanobis distance can be approximated by (21) when the number $N$ of pixels used to estimate the covariance matrix is much larger than the number of bands $L$.

To study the joint (among bands) distribution of HSI data sets, we compute the probability of exceedance of Mahalanobis distance for various HYDICE data sets. The obtained empirical distributions are compared to the theoretical chi-squared and F distributions corresponding to the multivariate normal and $t$ distributions. To reduce the effects of spatial inhomogeneity, we divide the data cube into rectangular blocks as shown in Figure 13. The distribution of the Mahalanobis distance is shown in Figure 14 for all blocks plus the three spatially determined classes.
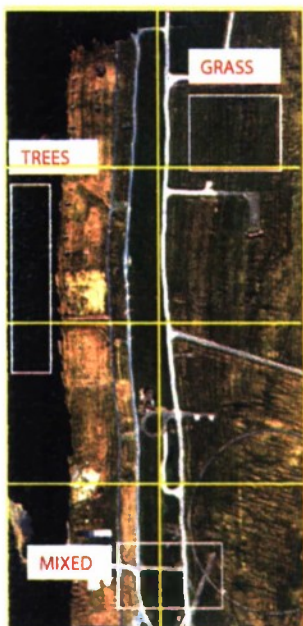


- HYDICE (HYperspectral Digital Imagery Collection Experiment)
  - Airborne sensor
- 210 spectral bands
  - 399-2501 nm
  - Channel widths ~ 3 – 11 nm
  - Spatial resolution, 1m x 1 m

Classes Selected for Statistical Analysis

| Class Name | Selection Technique | Sample Size |
|---|---|---|
| Grass | Spatially Adjacent | 7,760 |
| Trees | Spatially Adjacent | 8,232 |
| Mixed | Spatially Adjacent | 7,590 |
| Blocks 1-8 | Spatially Adjacent | 25,000 |

*Figure 13. Division of data cube into rectangular blocks to reduce spatial inhomogeneity.*

Another way to reduce spatial inhomogeneity is to model each class obtained by supervised or unsupervised classification separately. To this end, we use the classes shown in Figure 15, which have been derived by elaborate processing techniques by a group at the Remote Sensing Laboratory, Purdue University. The distribution of the Mahalanobis distance for the five most populated classes is shown in Figure 16.

The results in Figures 14 and 16 indicate that, typically, the joint distribution of HYDICE data cubes cannot be accurately modelled by the multivariate normal distribution. However, the elliptically contoured multivariate $t$-distribution provides a promising model. We note that the $t$-distribution tends to the normal distribution when the number of degrees of freedom increases.
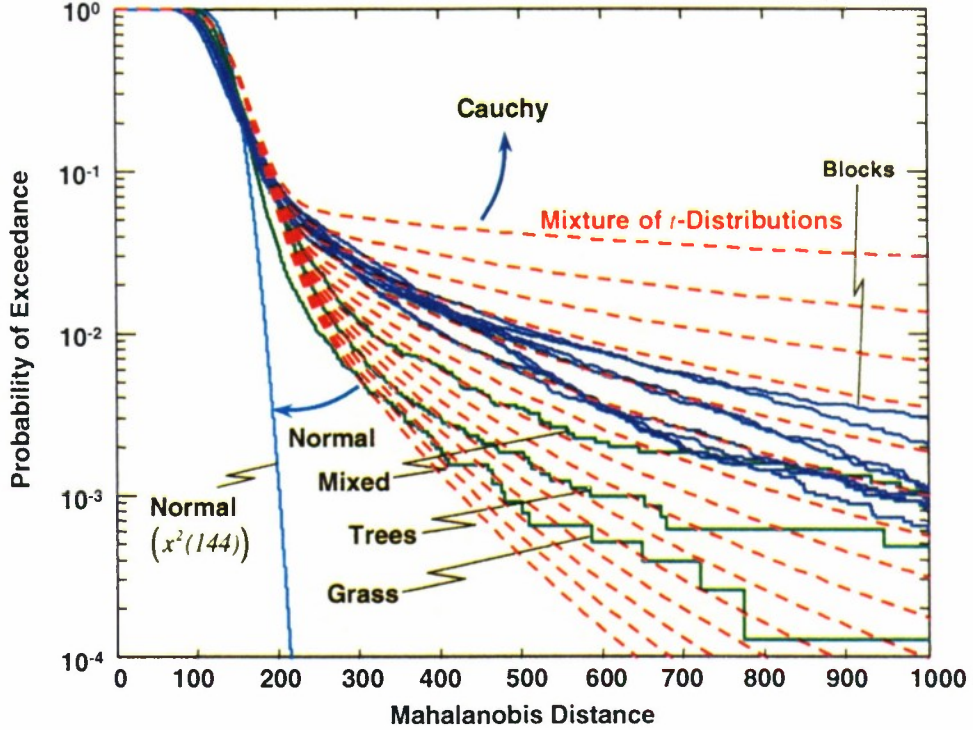
*Figure 14. Modelling the distribution of the Mahalanobis distance for the HSI data blocks shown in Figure 13.*

## 3.4 SUBSPACE OR LOW-RANK DATA MODELLING

To understand the nature of target variability, we compute the angle between all possible pairs of target spectra for the targets shown in Figure 17. The resulting matrix is displayed pictorially as an image, which show the angular variability of the target spectrum. Furthermore, we show the histogram of the angle between each target pixel spectrum and the mean spectrum. The clustering of the spectra about the mean indicates that target spectra populate a lower dimensional subspace of the $L$-dimensional data space. To investigate the nature of this subspace, we can use the singular value decomposition technique.

The background subspace matrix $S_b$ can be estimated from the HSI cube using statistical or geometrical techniques [28]. To describe the ideas behind some of these methods, consider a matrix $X^T$ of size $L \times N$, where each column contains the spectrum of an HSI pixel $x(n)$, that is

$$X^T \triangleq [x(1)\, x(2) \ldots x(N)] \tag{22}$$

where $N$ is the total number of pixels. Background characterization for the detection of low-probability targets can be done using the eigenvectors [11, 56] of the HSI cube correlation matrix $\hat{R}_x = X^T X/N$ or equivalently the singular vectors [61] of the data matrix $X^T$. In the first case, matrix $S_b$ is formed by the first $Q$ significant eigenvectors of $R_x$. In the second case, $S_b$ is formed by the first $Q$ significant left-singular

- Supervised classification resulted in 12 background classes

| Class | #Samples |
|---|---|
| Grass 1 | 57,274 |
| Grass 2 | 96,359 |
| Grass 3 | 59,438 |
| Tree 1 | 16,732 |
| Tree 2 | 16,668 |
| Bushes | 1,816 |
| Shaded Trees | 25,496 |
| Shaded Road | 7,448 |
| Road 1 | 575 |
| Road 2 | 8,049 |
| Road 3 | 4,066 |
| Road 4 | 5,648 |

- Classes were screened to be unimodal
- Fine class distinctions apparent

**Classification Result**

- Unlabelled
- Grass 1
- Grass 2
- Grass 3
- Tree 1
- Tree 2
- Bushes
- Shaded Trees
- Shaded Road
- Road 1
- Road 2
- Road 3
- Road 4

**Natural Color**

*Figure 15. Classes, classification results, and natural color image for the analyzed HYDICE data cube.*

vectors of $X^T$. The SVD decomposition implies the following approximation

$$X^T = U \Sigma V^T = \sum_{k=1}^{r} \tilde{\sigma}_k u_k v_k^T \Rightarrow \hat{X}^T = \sum_{k=1}^{Q} \tilde{\sigma}_k u_k v_k^T \tag{23}$$

which is optimum in the sense of least squares error. It can be shown that

$$\sum_{n=1}^{N} \left\| x(n) - \hat{x}(n) \right\|^2 = \sum_{k=Q+1}^{r} \tilde{\sigma}_k^2, \tag{24}$$

where $\tilde{\sigma}_k$ are the singular values of $X$. To obtain a good estimate of the background, the spectrum of interest should not be included among the significant eigenvectors or singular vectors. It should be stressed at this point that there is no one-to-one correspondence between the estimated $S_b$ and spectral properties. This is not a problem for detection applications as long as $S_b$ provides a good statistical approximation of the background and there is no leakage from the target subspace to the background subspace. The rarity of the target is a very important requirement in this respect; however, it is helpful to remove target-like pixels (that is pixels with large projections $\left\| S_t^T x(n) \right\|$ onto the target subspace) before computing the eigendecomposition or the SVD. Target leakage can be also reduced by excluding eigenvectors or singular vectors with large projections $\left\| S_t^T u_k \right\|$ onto the target subspace [61].

Figure 18 illustrates that a three dimensional subspace can be used to accurately model the spectral variability of the selected target. Indeed, $Q = 3$ singular vectors can be used to reconstruct 45 target spectra with a squared error of the order of $10^{-3}$ percent.

23
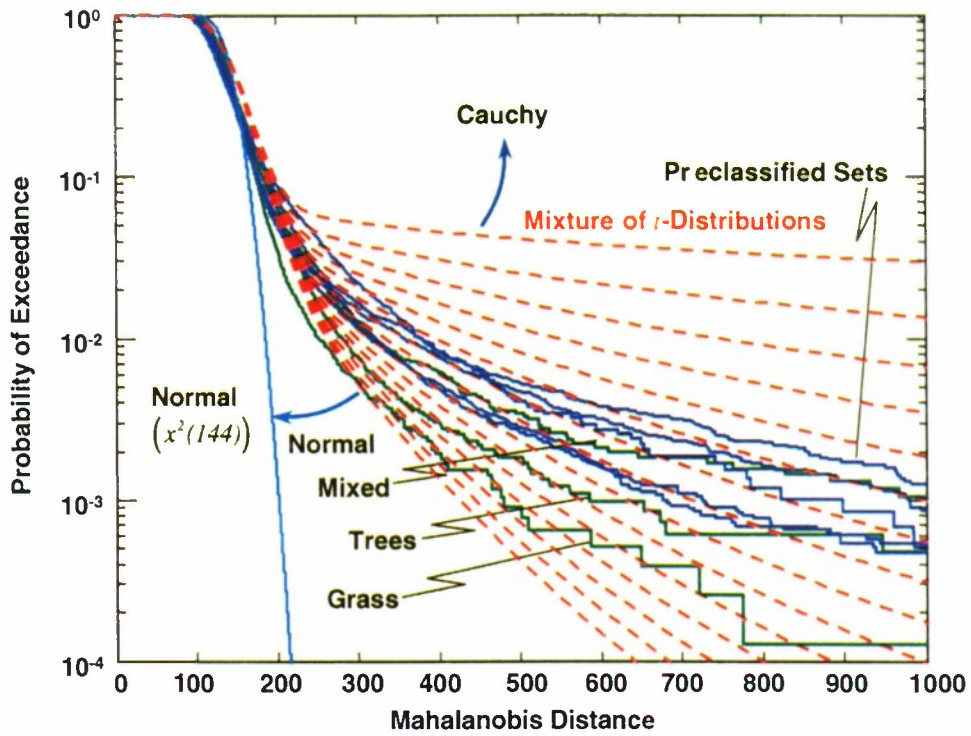
*Figure 16. Modelling the distribution of the Mahalanobis distance of the HSI data classes shown in Figure 15.*
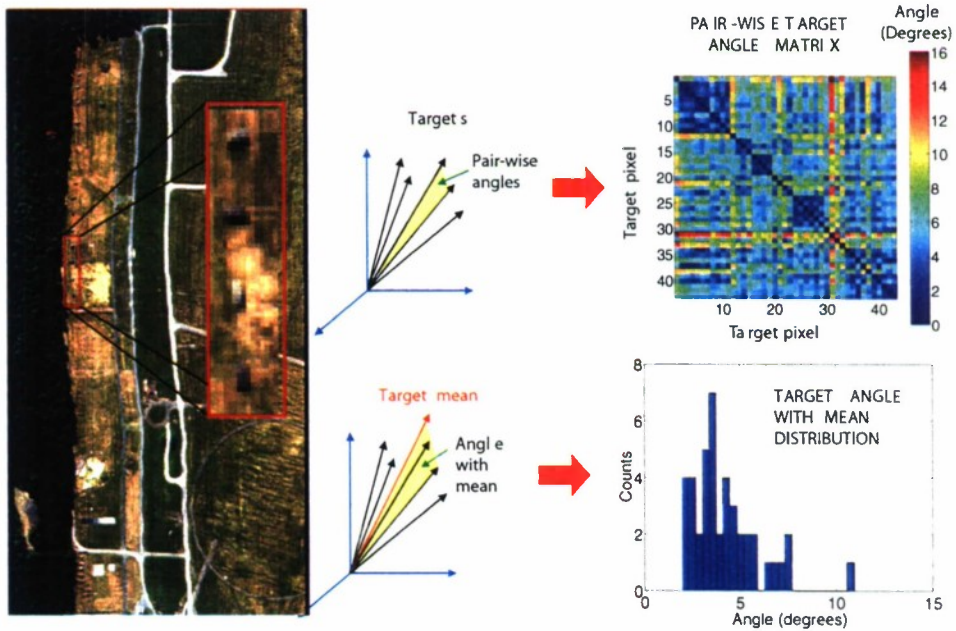


*Figure 17. Illustration of the angular distribution of target pixel spectra.*

24
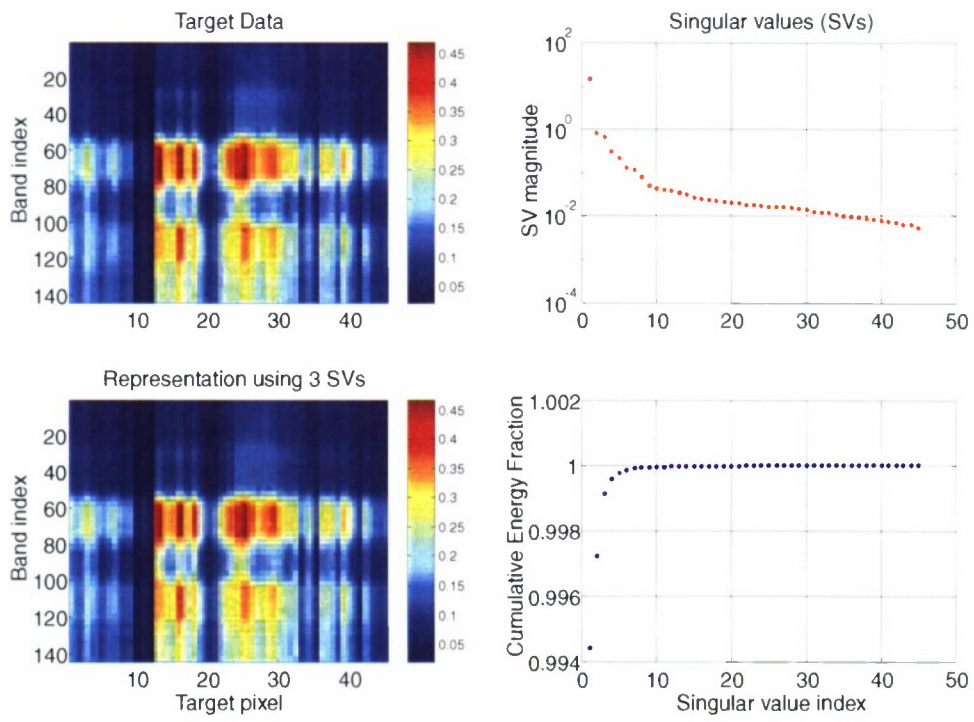
*Figure 18. Illustration of target subspace approximation using the singular value decomposition.*

# 4. DETECTION ALGORITHMS FOR FULL-PIXEL TARGETS

In this section, we assume that target pixels are completely filled by the material-of-interest, that is, we focus attention to full-pixel or resolved targets. In this case, the detection process is complicated by the spectral variabilities of the target and background classes. We can think of the totality of target spectra as constituting a target class and those from the background as being the background class. Let $R$ be the entire $L$-dimensional space in which the point of $L$-band spectrum $x$ falls. In order to make a decision, we should divide the region $R$ into two regions, $R_t$ and $R_b$, by some optimum method. A pixel is assigned to the target class if its spectrum $x$ falls in region $R_t$ or to the background class if $x$ falls in $R_t$. We shall pictorially illustrate the various concepts and algorithms using a hypothetical sensor with two spectral bands. However, due to the geometrical framework, the results and their interpretation hold for spectra produced by HSI sensors with a much larger number of bands. This process is illustrated in Figure 19 for $L = 2$ bands. Clearly, meaningful decision making is possible if the observed target spectra differ to some extent from the observed background spectra. Usually, the two classes overlap and even the best detector will result in misclassification errors. In general, the decision boundary will be a curve corresponding to a nonlinear detector. We can also make a decision by processing the spectrum vector $x$ by a system which calculates a scalar $y = D(x)$ and then comparing $y$ to a scalar threshold. Usually, the function $D(x)$ is obtained using the LR or GLR approaches. This system, which can be linear or nonlinear, is known as classifier, discriminant function, statistic, filter, or detector. We shall interchangeably use the terms filter or detector since they are widely used in the engineering literature.
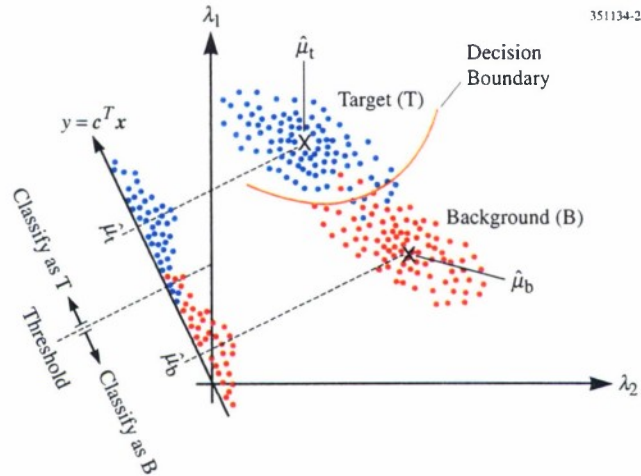


*Figure 19. Illustration of the feature space for two-class classification using two spectral bands.*

We discuss two approaches. First, we shall use the LR to obtain detectors without any structural constraints, that is, detectors with arbitrary decision surfaces in $R^L$. Second, we focus on the design of detectors with hyperplane decision surfaces. These linear detectors project the data onto a line specified by their coefficient vector with the objective to increase class separation.

## 4.1 LIKELIHOOD RATIO DETECTORS

Since statistical decision procedures, based on normal probability models, are simple and often lead to good performance, we shall model the target and background spectra as multivariate normal vectors. A random vector $x$ follows a multivariate normal distribution with mean vector $\mu \triangleq E\{x\}$ and covariance matrix $\Gamma \triangleq E\{(x - \mu)^T (x - \mu)\}$, denoted by $x \sim N(\mu, \Gamma)$, if its probability density function is given by

$$p(x) = \frac{1}{(2\pi)^{L/2} |\Gamma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Gamma^{-1}(x-\mu)} \tag{25}$$

where $|\Gamma|$ represents the determinant of matrix $\Gamma$.

Consider the detection problem specified by the following hypotheses

$$\begin{aligned} H_0 &: x \sim N(\mu_b, \Gamma_b) \quad \text{(Target absent)} \\ H_1 &: x \sim N(\mu_t, \Gamma_t) \quad \text{(Target present)} \end{aligned} \tag{26}$$

where the target and background classes follow multivariate normal distributions with different mean vectors and covariance matrices. Since the probability densities are completely specified under each hypothesis, we can design a Neyman-Pearson detector. Indeed, computing the natural logarithm of the LR (1) leads to the quadratic detector

$$y = D(x) = \frac{1}{2}(x - \mu_b)^T \Gamma_b^{-1}(x - \mu_b) - \frac{1}{2}(x - \mu_t)^T \Gamma_t^{-1}(x - \mu_t) \tag{27}$$

which compares the Mahalanobis distances of the observed spectrum from the centers of the two classes. The required threshold $\eta$ is determined from

$$P_{FA} = \int_\eta^\infty p(y|H_0) dy = \alpha \tag{28}$$

where $\alpha$ is the desired probability of false alarm. As a result of the quadratic mapping, the distribution of the random variable $y$ (detector output) is not normal, which makes the performance evaluation of the detector difficult.

If the target and background classes have the same covariance matrix, that is, $\Gamma_t = \Gamma_b \triangleq \Gamma$, the quadratic terms in (27) disappear, and the likelihood ratio detector (27) becomes

$$y = D(x) = (\mu_t - \mu_b)^T \Gamma^{-1} x \tag{29}$$

In this case, we have a linear detector

$$y = c^T x = \sum_{k=1}^{L} c_k x_k \tag{30}$$

which is specified by the coefficient vector

$$c = \Gamma^{-1}(\mu_t - \mu_b) \tag{31}$$

The output $y$ is now normally distributed because it is a linear combination of normal random variables. This result simplifies the evaluation of the detector and the computation of detection thresholds using (28).

A further simplification occurs when the observed spectra have uncorrelated components with equal variances, that is, $\Gamma = \sigma^2 I$. In this case, we have

$$y = D(x) = \frac{1}{\sigma^2} \left( \|x - \mu_t\|^2 - \|x - \mu_b\|^2 \right) \tag{32}$$

which is the well known Euclidean distance classifier. Equivalently, this can be expressed as

$$y = \frac{1}{\sigma^2} (\mu_t - \mu_b)^T x = \frac{1}{\sigma^2} (\mu_t{}^T x - \mu_b{}^T x) \tag{33}$$

which compares the correlation of the input spectrum with the mean of target and background spectra.

## 4.2   LINEAR (MATCHED FILTER) DETECTORS

When $\Gamma_t \neq \Gamma_b$, the resulting Neyman-Pearson detector is nonlinear. However, many times, either for simplicity or necessity, the design of a linear detector is the only available option. Optimum linear detectors can be formed by

$$c = (\kappa_1 \Gamma_b + \kappa_2 \Gamma_t)^{-1} (\mu_t - \mu_b) \tag{34}$$

where $\kappa_1 \Gamma_b + \kappa_2 \Gamma_t$ should be positive definite. It has been shown (see [3], Section 6.10) that the constants $\kappa_1$ and $\kappa_2$ can be chosen to (a) minimize the total error rate, (b) minimize one error rate with the level of the other error rate specified, or minimizing the maximum of the two error rates.

Since the output of the detector is normally distributed when $x$ is multivariate normal, detection performance is determined by the means $(\mu_b, \mu_t)$ and variances $(\sigma_b^2, \sigma_t^2)$ of $y$, under the two hypotheses. It can be shown [14] that minimization or maximization of any criterion $J(\mu_b, \mu_t, \sigma_b^2, \sigma_t^2)$ yields

$$c = [\delta \Gamma_b + (1 - \delta) \Gamma_t]^{-1} (\mu_t - \mu_b) \tag{35}$$

where

$$\delta = \frac{\partial J / \partial \sigma_b^2}{\partial J / \partial \sigma_b^2 + \partial J / \partial \sigma_t^2} \tag{36}$$

is a weighting factor $0 \leq \delta \leq 1$. We note that the functional form of $J$ affects the matched filter only through $\delta$. The criterion $J = (\mu_t - \mu_b)^2 / (\sigma_t^2 + \sigma_b^2)$ leads to $\delta = 1/2$, $J = (\mu_t - \mu_b)^2 / \sigma_b^2$ leads to $\delta = 1$, whereas if $J = (P_t \mu_t^2 + P_b \mu_b^2) / (P_t \sigma_t^2 + P_b \sigma_b^2)$ we obtain $\delta = P_b$.

The most important implication of (27) and (34) is that, when $\Gamma_t = \Gamma_b = \Gamma$, both lead to the same linear detector

$$c_{\text{MF}} = \kappa \Gamma^{-1} (\mu_t - \mu_b) \tag{37}$$

where $\kappa$ is a normalization constant. Furthermore, for $\kappa = 1$ this is identical with the LR detector (31) for equal covariance matrices. This detector, which is known as *Fisher Linear Discriminant* [13], is widely

29

used in pattern recognition applications. We shall use the term *matched filter (MF)*, which is more widely used in the communications and signal processing literature. There, the matched filter (37) is usually derived by maximizing the cost function

$$J(c) \triangleq \frac{[E\{y|H_1\} - E\{y|H_0\}]^2}{\text{var}\{y|H_0\}} = \frac{[c^T(\mu_t - \mu_b)]^2}{c^T \Gamma c} \tag{38}$$

which measures the distance between the means of two normal distributions in units of the common variance. The maximum, which is obtained by substituting (37) into (38), is

$$J_{\max} = \Delta^2 \triangleq (\mu_t - \mu_b)^T \Gamma^{-1}(\mu_t - \mu_b) \tag{39}$$

which is the Mahalanobis squared distance [23] between the means of the target and background distributions. The filter (37) with $\kappa = 1/\Delta^2$ minimizes the output variance $c^T \Gamma c$ subject to the linear constraint $c^T \mu_t = 1$.

**Geometrical interpretation** The output of a linear detector is the projection of the observation $x$ along the direction of the parameter vector $c$. We wish to determine the direction which provides the "best" separability between the two classes. The direction of the optimum matched filter for a white or spherical background distribution is along the direction of $\mu_t - \mu_b$. For elliptical backgrounds this direction is modified by the transformation $\Gamma^{-1}$. If we use the square-root decomposition $\Gamma = \Gamma^{1/2}\Gamma^{T/2}$ and we remove the mean of the background from the observed spectrum, the output of the matched filter is given by

$$y = \kappa \left[ \Gamma^{-1/2}(\mu_t - \mu_b) \right] \left[ \Gamma^{-1/2}(x - \mu_b) \right] \triangleq \kappa(\Delta\tilde{\mu})^T(\Delta\tilde{x}) \tag{40}$$

The transformation $\Delta\tilde{x} \triangleq \Gamma^{-1/2}(x - \mu_b)$, known as whitening or spherizing, creates a random vector with identity covariance matrix. The output of the matched filter is the projection of the whitened difference $\Delta\tilde{x}$ along the direction of $\Delta\tilde{\mu} \triangleq \Gamma^{-1/2}(\mu_t - \mu_b)$. The operation of the matched filter in the original spectral space and the whitened space are illustrated in Figure 20. If we choose $\kappa = 1/\left\| \Delta\tilde{\mu} \right\|^2 = 1/\Delta^2$, the output of the detector is normalized so that $y = D(\mu_t) = 1$. If we set $\kappa = 1/\left( \left\| \Delta\tilde{\mu} \right\| \left\| \Delta\tilde{x} \right\| \right)$, we obtain a nonlinear processor which provides the cosine of the angle between the vectors $\Delta\tilde{\mu}$ and $\Delta\tilde{x}$. In the original space, this is the Mahalanobis angle between the vectors $\Delta\mu$ and $\Delta x$. The length of a "spectral" vector increases or decreases when the overall illumination increases or decreases, but its angular orientation remains fixed. Equivalently, the shape of the spectrum remains the same but its amplitude changes proportionally to the illumination. For this reason, angle distances play an important role in HSI data processing.

**Detection Statistic** The properties of a given detector are specified by the distribution of its output, $y = D(x)$. In the first case, we assume that the observation $x$ is the only random quantity used to compute the output of the detector. This is the case, when we know the required mean vectors and covariance matrices. It is known, that if $x \sim N(\mu, \Gamma)$, then $y = c^T x \sim N(c^T \mu, c^T \Gamma c)$ for any non-null vector $c$. Therefore, when the mean and covariance matrices of the target and background classes are known, the output of the detector follows [3] the univariate normal distributions $N(c^T \mu_t, c^T \Gamma_t c)$ and $N(c^T \mu_b, c^T \Gamma_b c)$ for the target and background classes, respectively. These distributions can be used to determine the probabilities of detection and false alarm for any threshold $\eta$. We note that, the normalization constant $\kappa = 1/\Delta$ implies $y \sim N(0, 1)$, which simplifies the selection of the threshold. For deterministic targets $\Gamma_t = 0$ and $\mu_t = s$, which implies $x = s$ with probability one.
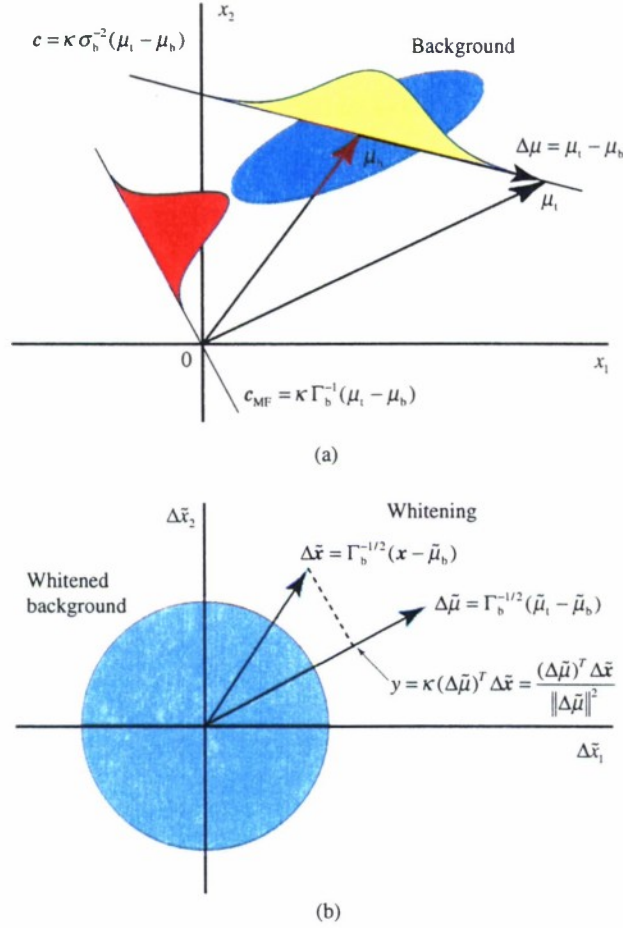
*Figure 20. Illustration of matched filter operation (a) Projection in the "best" direction (b) Background whitening followed by projection onto the $\Delta\tilde{\mu}$ vector.*

## 4.3  ADAPTIVE MATCHED FILTERS

The matched filter detector (37) requires the mean vector and the covariance matrix of the target and background distributions. Furthermore, *the resulting detector is optimum (in the Bayes or Neyman-Pearson sense) only when the target and background classes follow multivariate normal distributions with the same covariance matrix, an unlikely situation for real-world HSI data.* In practical applications, these quantities are unavailable and have to be estimated from the available data. Under the assumption of low-probability targets, we can use the available data $x(n), n = 1, 2, \ldots, N$, to determine the maximum likelihood estimates

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x(n) \simeq \hat{\mu}_{\mathrm{b}} \tag{41}$$

31

$$\hat{\Gamma} = \frac{1}{N} \sum_{n=1}^{N} [x(n) - \hat{\mu}][x(n) - \hat{\mu}]^T \simeq \hat{\Gamma}_b \tag{42}$$

of the mean vector and covariance matrix of the background. Unfortunately, there is usually not sufficient training data to determine the mean and covariance of the target. Typically, we use a target spectral signature $s_t$ from a library or the mean of a small number of known target pixels observed under the same conditions. The resulting *adaptive matched filter (AMF)* is given by

$$y = \frac{s^T \hat{\Gamma}_b^{-1} x}{s^T \hat{\Gamma}_b^{-1} s} \tag{43}$$

where usually the data cube mean is removed from the target and test pixel spectra.

If we know the "true" covariance matrix $\Gamma_b$, the output $y$ is distributed as $y \sim N(y_0, (s^T \Gamma_b^{-1} s)^{-1})$, where $y_0 = E\{y\}$. When the required means and covariances are estimated from the data, the resulting estimates are random quantities. If we treat them as constant, we can determine the class-conditional distribution of the detector output as in the known statistics case. However, the correct approach is to treat the estimated means and covariances as random and determine the unconditional distribution of $y = D(x)$. Unfortunately, the derivation of unconditional distributions is a very difficult problem, even under the most simplified assumptions. An extremely complicated expression for Fisher's linear discriminant has been obtained by Sitgreaves [55].

The distribution $p_y(y)$, when we use the sample covariance matrix $\hat{\Gamma}_b$, has been determined in a paper by Richmond [46] to be

$$p_y(y) = \frac{s^T \Gamma_b^{-1} s}{\pi} \frac{N_b - L + 2}{N_b + 1} \, {}_1F_1 \left( \frac{N_b - L + 3}{2}; \frac{N_b + 2}{2}; -\frac{1}{2}(y - y_0)^2 s^T \Gamma_b^{-1} s \right) \tag{44}$$

where

$$_1F_1(a; b; x) = \sum_{\beta=0}^{\infty} \frac{\Gamma(a + \beta)\Gamma(b)x^\beta}{\Gamma(a)\Gamma(b + \beta)\beta!} \tag{45}$$

is the confluent hypergeometric function and $\Gamma(v)$ denotes the Gamma function. Recall that $L$ is the number of spectral bands and $N_b$ is the number of pixels used to estimate the covariance matrix of the background. The background pixels are assumed to come from a multivariate normal distribution $x \sim N(\mu_b, \Gamma_b)$. Figure shows $p_y(y)$ for $L = 64$ and $L = 144$ spectral bands for various values of $N_b$. We see that, as the number of pixels used to estimate the covariance matrix of the background increases, the distribution approaches the normal distribution curve (see [34] for more details). This is expected because as $N_b \to \infty$, the sample covariance matrix $\hat{\Gamma}_b \to \Gamma_b$, the true covariance matrix of the background. However, as the number of bands increases, the tails of the distribution deviate more from the tails of the normal distribution. This differs from what we would expect by invoking the central limit theorem.

**Constrained Energy Minimization (CEM)** A similar detector, termed the constrained energy minimization (CEM) algorithm [19], can be obtained by minimizing the total energy of the filter output

$$E = \frac{1}{N} \sum_{n=1}^{N} y^2(n) = c^T \left[ \frac{1}{N} \sum_{n=1}^{N} x(n)x^T(n) \right] c = c^T \hat{R} c \tag{46}$$

32
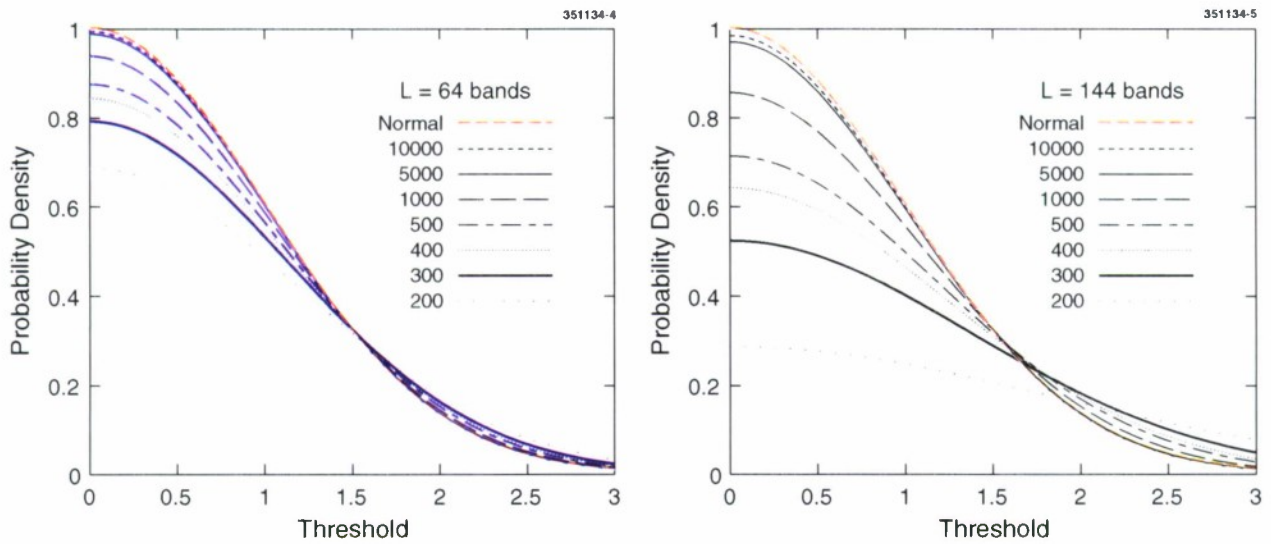
*Figure 21. Probability density function of $p_y(y)$ for various values of the number $N_b$ of the background pixels. The threshold is measured from the ideal value $y_0$ in units of $(s_t^T \Gamma_b^{-1} s_t)^{-1/2}$.*

subject to the constrain $c^T s = 1$. The matrix $\hat{R}$ is the sample correlation matrix of the data cube. The solution is

$$\hat{c}_{\text{CEM}} = \frac{\hat{R}^{-1} s}{s^T \hat{R}^{-1} s} \tag{47}$$

assuming that the sample correlation matrix is invertible. The minimum of the cost function is given by $E_{\text{CEM}} = 1/s^T \hat{R}^{-1} s$.

**Spectral similarity measures** The widely used spectral angle mapper [54] (SAM) algorithm is given by

$$D_{\text{SAM}}(x) \triangleq \frac{s^T x}{(s^T s)^{1/2} (x^T x)^{1/2}} \tag{48}$$

Clearly, SAM is the cosine of the angle between the test and target spectra and is always between zero and one because all spectra vectors have positive components. We note that SAM is usually defined in the remote sensing community as the angle between two vectors, instead of the cosine of the angle.

## 4.4   DISTRIBUTION OF UNIVARIATE MATCHED FILTER DETECTION STATISTICS

The heavy tails in the univariate distribution of the Mahalanobis distance imply heavy tails in the multivariate distribution of the data. Therefore, heavy tails may appear not only in the quadratic Mahalanobis distance, but in other linear and quadratic statistics employed in several widely used [35, 36] target detection techniques. In this section, for illustration purposes, we shall investigate the statistic of the popular matched

33
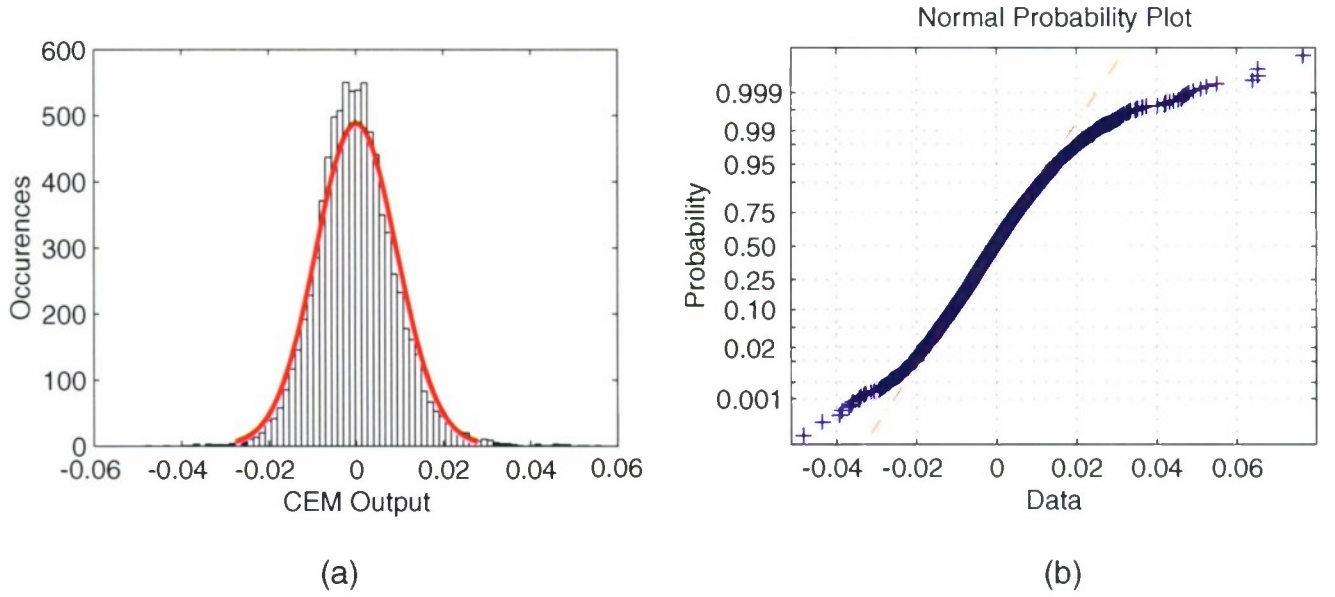
*Figure 22. Matched filter output statistic (a) histogram and fitted normal density curve and (b) normal probability plot for the tree data set.*

filter [36] algorithm termed Constrained Energy Minimization [12] (CEM). The algorithm is given by

$$D(x) = \frac{(s - \hat{\mu})^T \hat{\Gamma}^{-1}(x - \hat{\mu})}{(s - \hat{\mu})^T \hat{\Gamma}^{-1}(s - \hat{\mu})} \tag{49}$$

where $s$ is the desired target spectral signature and $x$ is the test pixel. If $x \sim \mathcal{N}(\mu, \Gamma)$, and $N$ is large compared to the number of bands $L$, $T(x)$ should be normally distributed. Figure 22 shows the histogram of the matched filter output for the tree data set and the corresponding normal probability plot. Clearly, the normal probability plot indicates the existence of heavy tails. Similar plots, with smaller or larger deviations from normality, have been produced using the other data sets. Hence, using a normal distribution to predict the probability of false alarm will provide optimistic estimates.

The family of symmetric $\alpha$-stable (S$\alpha$S) distributions provides a good model for data with impulsive behavior. They are characterized by a parameter $\alpha$ (characteristic exponent) that takes values in the range $0 < \alpha \leq 2$. The value $\alpha = 1$ leads to the Cauchy distribution and the value $\alpha = 2$ to the Gaussian. The stable distributions result from the central limit theorem if we remove the finite variance constraint. The only stable distribution with finite second-order moments is the normal distribution. Since $\alpha$-stable distributions follow from the central limit theorem they are invariant under linear transformations. Since there is no closed-form expression for their probability density function, S$\alpha$S random variables are specified [2, 39, 50] by their characteristic function (that is, the Fourier transform of the PDF)

$$\Phi(\xi) = \exp(j\mu\xi - |\sigma\xi|^{\alpha}) \tag{50}$$

where $\alpha$ is the characteristic exponent, $\sigma$ is a scale parameter, and $\mu$ is a location parameter. The heaviness of the tails increases as $\alpha$ increases from 1 (Cauchy) to 2 (Gaussian). The estimation of the parameters of a

stable distribution from data is a challenging due to the presence of "spikes". The published compilation [2] provides a comprehensive review of statistical techniques for stable distributions from a practical perspective. The estimation method used in this paper [30] is based on the use of the characteristic function.

Figure 23 shows the probability of false alarm when the matched filter detector is used for different scenes as well as superimposed theoretical curves obtained using the family of S$\alpha$S distributions for various values of $\alpha$. It can be seen that the tails of the empirical $P_{FA}$ curves can be modelled by the heavier tails of the stable distribution.



*Figure 23. Matched filter output statistics and their modelling using stable distributions.*

# 5. DETECTION ALGORITHMS FOR SUB-PIXEL TARGETS

By definition, subpixel targets occupy only part of the pixel area. The remaining part is filled with one or more materials, which will be collectively referred to as background. In this section, we discuss the detection of compact and isolated subpixel size objec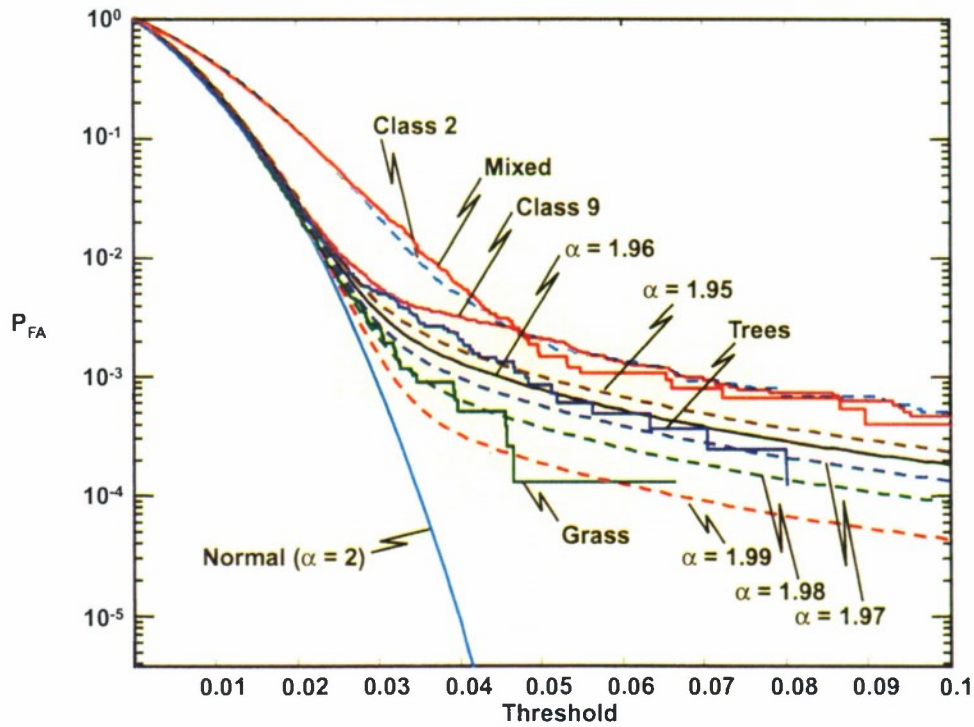ts characterized by a known spectral signature with or without variability. As a result of this area mixing, the observed spectral signature can be modelled reasonably well as a linear combination of the target and background spectra. Furthermore, there is always an additive amount of noise from various sources (sensor, atmosphere, etc).

The choice of the mathematical model used to describe the variability of target and background spectra (subspace versus statistical), leads to different families of subpixel target detection algorithms. The variability of the target spectral signature is always described using a subspace model $Sa$. If the columns of $S$ are endmembers, the vector $a$ provides their abundances and should satisfy the constraints of the linear mixing model. Otherwise, $a$ simply determines the position of a target in the column space of $S$.

The variability of the background can be described using either a subspace model (structured background) or a statistical distribution (unstructured background). Therefore, the type of the background model drives the development of subpixel detection algorithms.

Any matched filter-based target detection algorithm requires the specification of a spectral signature of interest, which can be obtained from within the HSI cube or from a spectral library. Usually, targets are modelled using a single spectrum. However, due to changes in atmospheric conditions, sensor geometry, surface defects and films, a target spectral signature can exhibit significant variability. If we have available a multitude of target spectrum observations, say $s_k$, $k = 1, 2, \ldots, N_t$, we can model target variability and use it in the detection algorithm to improve robustness. We can account for this variability statistically or geometrically. A statistical description could use the mean and covariance of the available target signatures $s_1, s_2, \ldots, s_{N_t}$. A geometrical description involves finding an orthogonal base that spans, with sufficient accuracy, the subspace spanned by the set of vectors $s_1, s_2, \ldots, s_{N_t}$. These basis vectors constitute the columns of the target subspace matrix $S_t$. It has been shown [61] that the target subspace for a material can be described by a small number of basis vectors under a variety of scenes and illumination conditions.

## 5.1 UNSTRUCTURED BACKGROUND MODELS

Unstructured background models assume that the additive noise has been included in the background $b$, which in turn is modelled by a multivariate normal distribution with mean $\mu_b$ and covariance matrix $\Gamma_b$, that is $b \sim N(0, \Gamma)$ (for simplicity, we drop the subscript $_b$ from this point forward). The competing hypotheses are

$$H_0 : \quad x = b, \qquad \text{Target absent}$$
$$H_1 : \quad x = Sa + b, \quad \text{Target present}$$

(51)

Hence, $x \sim N(0, \Gamma)$ under $H_0$ and $x \sim N(Sa, \Gamma)$ under $H_1$. In addition, we assume that we have access to a set of training background pixels $x(n), n = 1, 2, \ldots, N$, which are independently and identically distributed (IID). The test pixel $x$ and the training pixels are assumed statistically independent. Since, $x(n) \sim N(0, \Gamma)$, we can use these pixels to obtain the maximum likelihood estimate of the covariance matrix. Since HSI data have a non zero mean, we usually remove the estimated mean to comply with this model.

37

Using the generalized likelihood ratio approach, Kelly [25, 26] obtained the following detector

$$
D_K(x) = \frac{x^T \hat{\Gamma}^{-1} S (S^T \hat{\Gamma}^{-1} S)^{-1} S^T \hat{\Gamma}^{-1} x}{N + x^T \hat{\Gamma}^{-1} x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_K \tag{52}
$$

where $\hat{\Gamma}$ is the MLE of the covariance matrix (42). Although there is no optimallity test associated with the GLR approach [24], it leads to the design of useful, practical detectors. The threshold parameter $\eta_K$ determines both the probability of detection, $P_D$, and the probability of false alarm, $P_{FA}$.

The matrix $S$ contains the available a-priori variability information about the target. This information decreases as we increase the number of columns $P$ (dimensionality of the target subspace) of $S$ and becomes minimum when $P = L$. In this case, we simply know that we are looking for deterministic targets that lie in the data subspace. Since the matrix $S$ has full rank and therefore is invertible, (52) leads to the following detector

$$
D_A(x) = x^T \hat{\Gamma}^{-1} x \underset{H_0}{\overset{H_1}{\gtrless}} \eta_A \tag{53}
$$

which was derived by Kelly [27] using the approach presented here and by Reed and Yu [45] using a multivariate analysis of variance formulation. Basically, $D_A(x)$ estimates the Mahalanobis distance of the test pixel from the mean of the background, which is zero for demeaned data. Algorithm (53), which has the CFAR property, is used extensively for anomaly detection in multispectra and hyperspectral data [57].

A key assumption in the derivation of (52) was that the covariance matrix of the background is the same under the two hypotheses. However, for subpixel targets the amount of background covered area is different under the two hypotheses. Therefore, it is more appropriate to use the following hypotheses

$$
\begin{aligned}
H_0 : \quad & x = b, & \text{Target absent} \\
H_1 : \quad & x = Sa + \sigma b, & \text{Target present}
\end{aligned} \tag{54}
$$

which implies that $x \sim N(0, \Gamma)$ under $H_0$ and $x \sim N(Sa, \sigma^2 \Gamma)$ under $H_1$. In other words, the background has the same covariance structure but different variance. This variance is directly related to the fill factor of the target, that is, the percentage of the pixel area occupied by the target object. The GLR approach leads to the following Adaptive Coherence/Cosine Estimator (ACE) detector [31], [32]

$$
D_{ACE}(x) = \frac{x^T \hat{\Gamma}^{-1} S (S^T \hat{\Gamma}^{-1} S)^{-1} S^T \hat{\Gamma}^{-1} x}{x^T \hat{\Gamma}^{-1} x} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ACE} \tag{55}
$$

which can be obtained from (52) by removing the number $N$ of background training pixels from the denominator.

If we use the adaptive whitening transformation $\tilde{x} \triangleq \hat{\Gamma}^{-1/2} x$, where $\hat{\Gamma} = \hat{\Gamma}^{1/2} \hat{\Gamma}^{1/2}$ is the square-root decomposition of the estimated covariance matrix, the ACE can be expressed as

$$
D_{ACE}(x) = \frac{\tilde{x}^T \tilde{S} (\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T \tilde{x}}{\tilde{x}^T \tilde{x}} = \frac{\tilde{x}^T P_{\tilde{S}} \tilde{x}}{\tilde{x}^T \tilde{x}} \tag{56}
$$

where $\tilde{S} \triangleq \hat{\Gamma}^{-1/2} S$ and $P_{\tilde{S}} \triangleq \tilde{S}(\tilde{S}^T \tilde{S})^{-1} \tilde{S}^T$ is the orthogonal projection operator onto the column space of $\tilde{S}$. Since $P_{\tilde{S}}^2 = P_{\tilde{S}}$, we can write (55) as

$$D_{\text{ACE}}(x) = \frac{\left\| P_{\tilde{S}} \tilde{x} \right\|^2}{\|\tilde{x}\|^2} = \cos^2 \theta \tag{57}$$

which shows that $D_{\text{ACE}}(x)$ is equal to the cosine of the angle between the test pixel and the target subspace into the whitened coordinate space. This is illustrated in Figure 24.
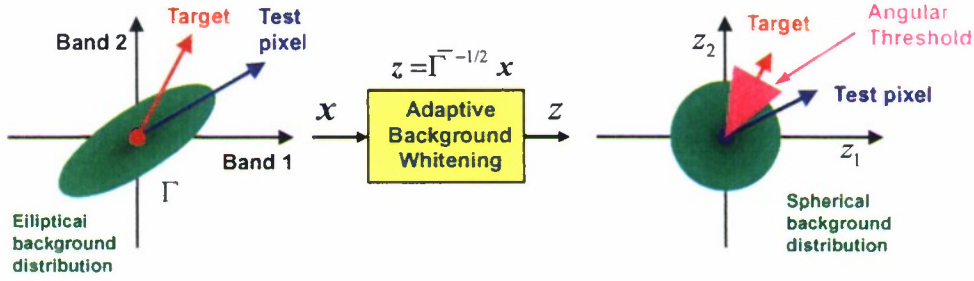


*Figure 24. Illustration of GLRT detectors.*

Using the whitening transformation, the anomaly detector (53) can be expressed as $D_{\text{A}}(x) = \tilde{x}^T \tilde{x}$ which is the Euclidean distance of the test pixel from the background mean in the whitened space (see Figure 25). We note that, in the absence of a target direction, the detector uses the distance from the center of the background distribution.
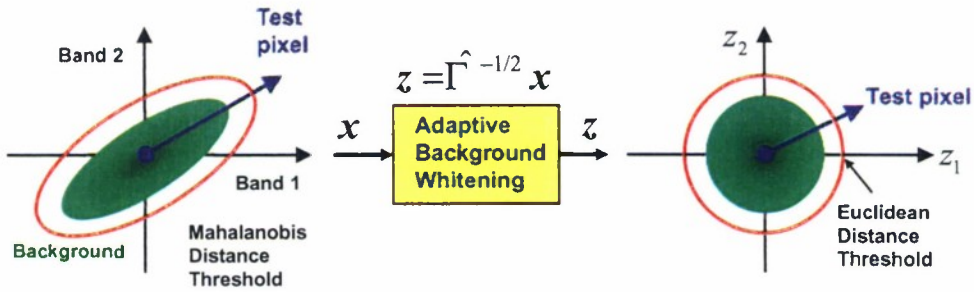


*Figure 25. Illustration of detection of targets with unknown spectral signatures (anomaly detection).*

For *targets without variability*, we have $P = 1$ and the target subspace $S$ is specified by the direction of a single vector $s$. Then, the formulas for the previous GLR detectors are simplified to

$$D(x) = \frac{(s^T \hat{\Gamma}^{-1} x)^2}{(s^T \hat{\Gamma}^{-1} s)(\psi_1 + \psi_2 x^T \hat{\Gamma}^{-1} x)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{58}$$

where $(\psi_1 = N, \psi_2 = 1)$ for the Kelly detector and $(\psi_1 = 0, \psi_2 = 1)$ for the ACE. Kelly's algorithm was derived for real-valued signals and was applied to multispectral target detection in [63]. The one-dimensional

version of ACE, has been derived in [9], [52]. Finally, we note that, if ($\psi_1 = N$, $\psi_2 = 0$), we obtain the adaptive matched filter (AMF) detector [49], [7].

Determining the distribution of the various GLRT detectors is an elaborate process [25], [27], [45], [47], [32]. It turns out that the distribution of the different detector outputs involves a non-central F-distribution. The non-centrality parameter is the theoretical signal-to-interference plus noise ratio

$$\text{SINR}_\text{o} = (Sa)^T \Gamma^{-1}(Sa) \tag{59}$$

The performance of all GLRT detectors depends only on the dimensional integer numbers $L$, $P$, $N$ and the optimum $\text{SINR}_\text{o}$ parameter. Under the $H_0$ hypothesis (target absent) $\text{SINR}_\text{o} = 0$, the output distribution becomes central , and the probability of false alarm depends only on the parameters $L$, $P$, $N$. Therefore, all GLRT detectors discussed in this section have the CFAR property. For a remote sensing system with high SNR, the detection performance depends on the fraction of the target occupying a pixel, since this determines the SINR. Figure 26 illustrates the effects of target dimensionality on detection performance for Kelly's GLRT detector. We see that as $P$ increases (that is as a-priori information about the target decreases), detection performance deteriorates. The worst performance is obtained when $P = L$, which corresponds to the anomaly detector (53). Figure 27 illustrates performance as a function of the number of training pixels. Clearly, performance improves as $N$ increases, that is as the estimate of the interference covariance matrix becomes more accurate. In both figures, we have included the curve indicating the performance of the optimum matched filter [24] (known target in noise with known covariance matrix) for comparison.
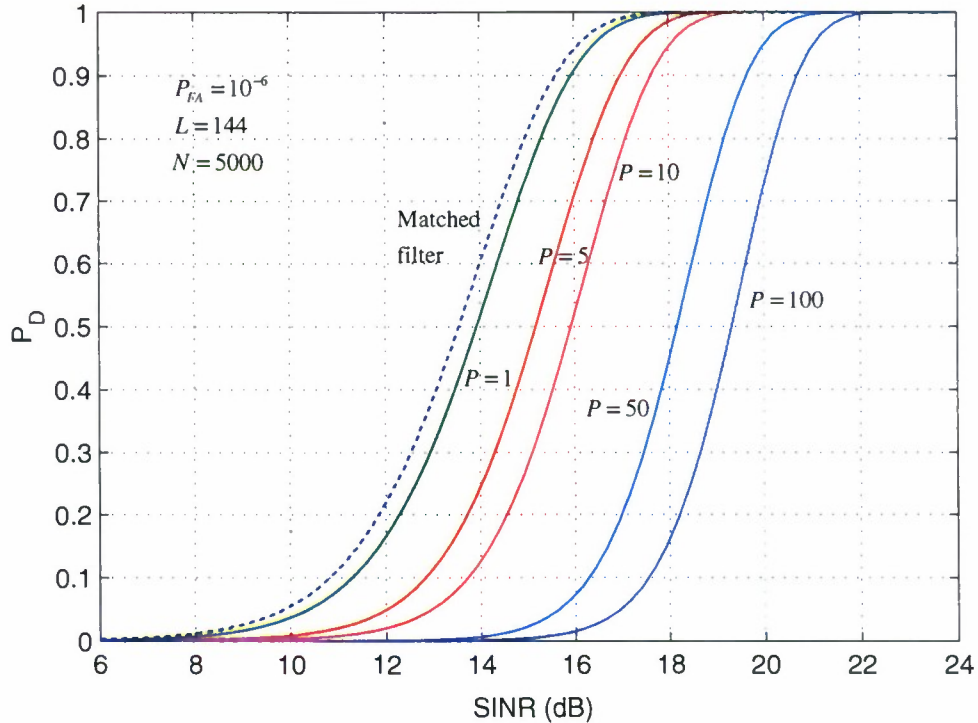


*Figure 26. Probability of detection as a function of SINR illustrating the effect of target subspace dimensionality.*
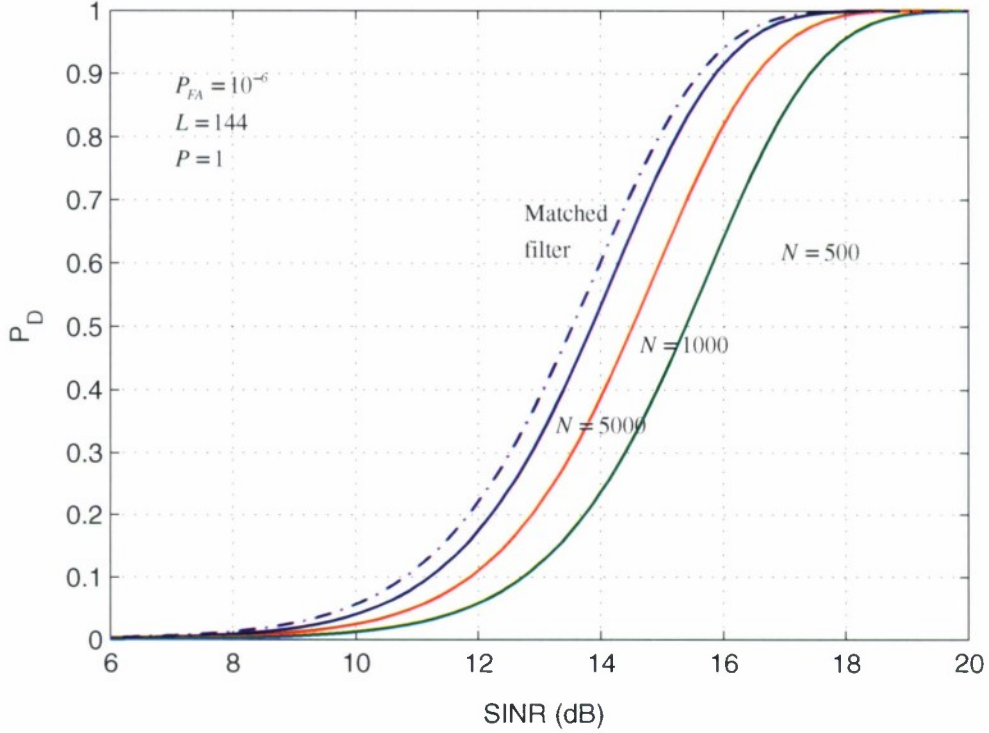
*Figure 27. Probability of detection as a function of SINR illustrating the effect of the number of training pixels.*

The key assumptions used for the detectors using the covariance matrix of the background are (a) the background is homogeneous and can be modelled by a multivariate normal distribution, (b) the background spectrum interfering with the test pixel spectrum has the same covariance matrix with the background training pixels, (c) the test and training pixels are independent, and (d) the target and background spectra interact in an additive manner (additive instead of replacement model).

## 5.2 BACKGROUNDS WITH LOW-RANK COVARIANCE MATRICES

The linear mixing model (3) provides a quite accurate model of the background with $w \sim N(0, \sigma_w^2 I)$ and $M = Q$ (number of background endmembers). This leads to the following structured covariance matrix

$$\Gamma = \sum_{k=1}^{Q} a_k^2 s_k s_k^T + \sigma_w^2 I \triangleq BAB^T + \sigma_w^2 I \tag{60}$$

where $A \triangleq \mathrm{diag}\{a_1^2, a_2^2, \ldots, a_Q^2\}$. By performing the eigendecomposition of $\Gamma$, we get

$$\Gamma = U \Lambda U^T = [U_b \, U_n] \begin{bmatrix} \Lambda_b & 0 \\ 0 & \Lambda_n \end{bmatrix} \begin{bmatrix} U_b^T \\ U_n^T \end{bmatrix} \tag{61}$$

where $\Lambda_b$ contains the $Q$ largest eigenvalues of $\Gamma$ in descending order and $U_b = [u_1, u_2, \ldots, u_Q]$ the corresponding orthonormal eigenvectors. $\Lambda_n = \sigma_w^2 I$ and $U_n$ contains the remaining orthogonal eigenvectors

41

corresponding to the eigenvalue $\sigma_w^2$. Clearly, the eigenvectors in $U_b$ are different from the endmembers in $B$; however, they span the same linear space. The range space of $U_b$ is called the background subspace and its orthogonal complement, known as the noise subspace, is spanned by the columns of $U_n$.

The practical implication of (60) is that the estimated background covariance matrix has low numerical rank, that is, there are only a few large eigenvalues and a large number of much smaller ones. However, there is no clear cut transition to help the separation between background and noise subspaces. When the number $N$ of background pixels used to estimate the covariance matrix is much larger than the number of bands $L$, the obtained matrix is usually invertible. However, as $N$ becomes equal to or less than $L$, invertibility is a concern and various techniques have been proposed to mitigate the problem. This situation arises when we process a small local window at a time to assure that a statistically homogeneous area is used to estimate the background covariance matrix. The widely used method of diagonal loading [5], uses the estimate $\hat{\Gamma}_d = \hat{\Gamma} + \delta I$ where the scalar $\delta$ is known as the loading factor. It is suggested in [8] that $\lambda_{min} < \delta < \sigma_w^2$, where $\lambda_{min}$ is the minimum background eigenvalue. If $\hat{\Gamma} = \hat{U}\hat{\Lambda}\hat{U}^T$, other approaches include: (a) the low-rank approximation [15] $\hat{\Gamma}_Q = \hat{U}\hat{\Lambda}_Q\hat{U}^T$, $\hat{\Lambda}_Q \triangleq \text{diag}\{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \lambda_Q, 0, \dots, 0\}$, (b) the fast maximum likelihood approach [4], [59] $\hat{\Gamma}_{FML} = \hat{U}\hat{\Lambda}_{FML}\hat{U}^T$, $\hat{\Lambda}_{FML} \triangleq \text{diag}\{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \lambda_Q, \delta^2, \dots, \delta^2\}$ with $\lambda_{min} < \delta < \sigma_w^2$, and the cross-spectral metric approach [18], which chooses the $Q \ll L$ eigenvectors so as to maximize the quantity

$$s^T\hat{\Gamma}^{-1}s = \sum_{k=1}^{L} \frac{(\hat{u}_k^T s)^2}{\hat{\lambda}_k} \tag{62}$$

which is related to the theoretical $\text{SINR}_o$ (59). More details regarding the effects of these approximations to hyperspectral target detection are given in [33].

To understand the effect the low-rank of the background covariance matrix has upon the operation of a detector, assume that $\Gamma = U\Lambda U^T + \sigma_w^2 I$, where $\Lambda$ contains the eigenvalues of $\Gamma$ in decreasing order and $U$ the corresponding orthonormal eigenvectors. The inverse $\Gamma^{-1}$ can be expressed as

$$\Gamma^{-1} = \frac{1}{\sigma_w^2}(I - P) \tag{63}$$

where

$$P = \sum_{k=1}^{L} \frac{\lambda_k}{\lambda_k + \sigma_w^2} u_k u_k^T \tag{64}$$

If $\lambda_k \gg \sigma_w^2$ for $k \leq Q$, we obtain $P \simeq \sum_{k=1}^{Q} u_k u_k^T \triangleq U_Q U_Q^T$. Hence, $\Gamma^{-1} \simeq (I - U_Q U_Q^T)/\sigma_w^2$. Since the columns of $U_Q$ are orthonormal, $P_{U_Q}^{\perp} = I - U_Q U_Q^T$ is the orthogonal subspace projector onto the background subspace spanned by the columns of $B$ or $U_Q$. This result provides a connection between the unstructured and structured background detection algorithms discussed in Section 7.

# 6. THE LINEAR MIXING MODEL

The basic premises of linear mixture modeling are that within a given scene : (a) the surface is dominated by a small number of materials with relatively constant spectra (end members), (b) most of the spectral variability within the scene results from varying proportions of the end members, and (c) the mixing relationship is linear if the end members are arranged in spatially distinct patterns [1].

In the linear mixing model (LMM), the spectrum of a mixed pixel is represented [1] as a linear combination of component spectra (end members). The weight of each end member spectrum (abundance) is proportional to the fraction of the pixel area covered by the end member. If there are $L$ spectral bands, the spectrum of the pixel and the spectra of the end members can be represented by $L$-dimensional vectors. Therefore, the general equation for mixing by area is given by

$$x = \sum_{k=1}^{M} a_k s_k + w \triangleq Sa + w \tag{65}$$

$$S \triangleq [s_1 s_2 \dots s_M] \tag{66}$$

$$a \triangleq [a_1 a_2 \dots a_M]^T \tag{67}$$

where $x$ is the spectrum of the mixed pixel, $s_k$ are the spectra of the end members, $a_k$ are their abundances, $M$ is the number of the end members, and $w$ is an $L$-dimensional error vector accounting for lack-of-fit and noise effects. Physical considerations dictate the following constrains

$$a_k \geq 0 \quad \text{(non-negativity constraint)} \tag{68}$$

$$\sum_{k=1}^{M} a_k = 1 \quad \text{(additivity constraint)} \tag{69}$$

which can be enforced, if necessary, to guarantee meaningful parameter values. If we wish to allow for unmodeled end members, we can replace the $(=)$ sign with the $(\leq)$ sign. Fitting a linear mixing model involves two steps: (a) end member identification and (b) abundance estimation. Although there are algorithms where the two steps are interwoven, the objectives of this discussion are better served by keeping the two steps distinct.

If we know the end members $s_k$, unmixing can be viewed either as a linear estimation problem or as a linear model fitting problem. Furthermore, in detection and classification applications we do not need to determine explicitly estimates of $a$; we can use a measure of the quality of the estimator or a measure of the model "goodness-of-fit" to make decisions. This is the approach used in this report.

## 6.1 ESTIMATION OF THE UNCONSTRAINED MODEL PARAMETERS

Suppose that we know the end members $s_k$, and we wish to estimate the abundance, $a$, given the observed pixel $x$. If $\hat{a}$ is an estimate of $a$, then $\hat{x} = S\hat{a}$ is an estimate of $x$ with a corresponding error
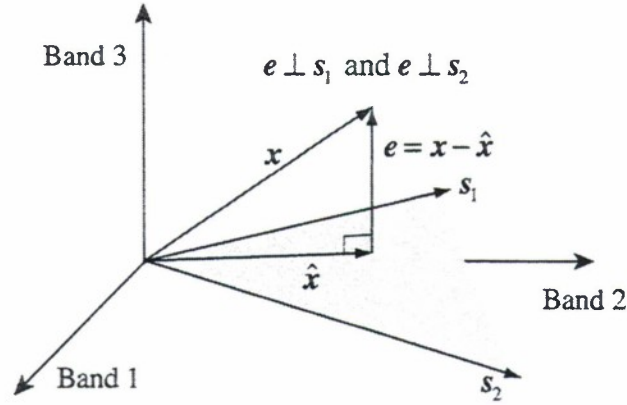
*Figure 28. Geometrical illustration of LS estimation of linear mixing model.*

$e = x - \hat{x} = x - S\hat{a}$. The *unconstrained least squares estimate (LSE)* $\hat{x}$ of $x$ is uniquely determined by the foot of the perpendicular from the tip of $x$ to the hyperplane specified by the columns of $S$ (see Figure 28). Thus

$$S^T e = S^T (x - \hat{x}) = 0 \tag{70}$$

which is known as the orthogonality principle.

If the columns of $S$, that is the end members $s_k$, are linearly independent (this is possible when $M \leq L$) there exists a unique vector $\hat{a}$ such that $\hat{x} = S\hat{a}$. Therefore, we have $S^T (x - S\hat{a}) = 0$ or

$$(S^T S)\hat{a} = S^T x \tag{71}$$

which are known as *normal equations*. In this case, the inverse of $(S^T S)$ exists and the LSE estimate of $x$ is given by

$$\hat{x} = S\hat{a} = S(S^T S)^{-1} S^T x = P_S x \tag{72}$$

where

$$P_S \triangleq S(S^T S)^{-1} S^T \tag{73}$$

is known as the *projection* matrix. The error or residual vector isobtained by

$$e = x - \hat{x} = P_S^{\perp} x \tag{74}$$

where

$$P_S^{\perp} \triangleq I - P_S \tag{75}$$

is the orthogonal projection error matrix. The model goodness-of-fit is assessed by the length of $e$, using the sum-of-squared errors (SSE)

$$\begin{aligned} \text{SSE}(\hat{a}) &\triangleq \|e\|^2 = x^T x - \hat{x}^T \hat{x} \\ &= x^T (I - P_S) x = x^T P_S^{\perp} x \end{aligned} \tag{76}$$

44

obtained by using the Pythagorean theorem. These ideas are pictorially illustrated in Figure 28.

If we assume that $w$ has mean $E\{w\} = 0$ and covariance $\Gamma_w = \sigma_w^2 I$, it can be shown that $\hat{a}$ is the best linear unbiased estimator (BLUE) of $a$ and $\hat{\sigma}_w^2 = \text{SSE}(\hat{a})/(L - M)$ is the best quadratic unbiased estimator (BQUE) of $\sigma_w^2$ [22].

If we further assume that $w \sim \mathcal{N}(0, \sigma_w^2 I)$, that is the error is normally distributed with zero mean and covariance $\Gamma_w = \sigma_w^2 I$, we can determine the *maximum likelihood estimates (MLEs)* of $a$ and $\sigma_w^2$ and their sampling distributions. The likelihood function is

$$\mathcal{L}(a, \sigma_w^2; x) = \frac{1}{(2\pi\sigma_w^2)^{L/2}} \exp\left\{ -\frac{(x - Sa)^T (x - Sa)}{2\sigma_w^2} \right\} \tag{77}$$

which is a function of $a$ and $\sigma_w^2$ with the data vector $x$ playing the role of a parameter. The probability density function $p(x; a, \sigma_w^2)$ is also given by (31), but now $x$ is the variable and $a$ and $\sigma_w^2$ are considered as parameters. Maximizing the monotonic log-likelihood function

$$\ell(a, \sigma_w^2; x) = \ln \mathcal{L}(a, \sigma_w^2; x)$$

gives

$$(S^T S)\hat{a} = S^T x \tag{78}$$

$$\hat{\sigma}_w^2 = \frac{1}{L} \text{SSE}(\hat{a}) \tag{79}$$

and

$$\mathcal{L}(\hat{a}, \hat{\sigma}_w^2; x) = \left[ \frac{L/2\pi}{\text{SSE}(\hat{a})} \right]^{L/2} \exp\left( -\frac{L}{2} \right) \tag{80}$$

which show that for $w \sim \mathcal{N}(0, \sigma_w^2 I)$ the LSE and MLE are the same (hence the use of the same symbol).

The statistical properties of the unconstrained LMM parameter estimates are summarized in the following proposition (see [22, 24, 42]).

**Proposition 1** *Consider the LMM $x = Sa + w$, where $w \sim \mathcal{N}(0, \sigma_w^2 I)$ and $S$ is an $L \times M$ matrix of rank $M$.*

*1. The minimum variance unbiased estimators of $a$ and $\sigma_w^2$ are given by*

$$\hat{a} = (S^T S)^{-1} S^T x \quad \text{and} \quad \hat{\sigma}_w^2 = \frac{1}{L - M} \text{SSE}(\hat{a}) \tag{81}$$

*2. The estimates $\hat{a}$ and $\hat{\sigma}_w^2$ are independently distributed as*

$$\hat{a} \sim \mathcal{N}(a, \sigma_w^2 (S^T S)^{-1}) \tag{82}$$

*and*

$$\hat{\sigma}_w^2 \sim \frac{\sigma_w^2}{L - M} \chi_{L-M}^2(0) \tag{83}$$

*where $\chi_K^2(0)$ denotes a central chi-square distribution with $K$ degrees of freedom.*

3. *The estimates $\hat{x}$ and $e$ are independently distributed as*

$$\hat{x} \sim \mathcal{N}(Sa, \sigma_w^2 P_S) \quad \text{and} \quad e \sim \mathcal{N}(0, \sigma_w^2 P_S^{\perp}) \tag{84}$$

## 6.2   ESTIMATION OF THE LINEARLY CONSTRAINED MODEL PARAMETERS

The additivity constraint $\sum_{k=1}^{M} a_k = 1$ is a special case of a more general linear constraint that can be expressed in matrix form by a consistent set of linear equations

$$Ga = g \tag{85}$$

where $G$ is a $P \times M$ matrix ($P < M$) of rank $P$ and $g$ is a $P \times 1$ vector. For example, we notice that

$$\begin{aligned} G &= [1\ 1\ \dots\ 1] \quad \text{and} \quad g = 1 \Rightarrow \quad \text{Additivity constraint} \\ G &= [I_P \mid 0] \qquad \text{and} \quad g = 0 \Rightarrow \quad a_k = 0, 1 \le k \le P \end{aligned} \tag{86}$$

which will be useful in estimation and detection applications. The properties of the linearly constrained LMM are summarized in the following proposition [22, 24].

**Proposition 2** *Consider the linearly constrained LMM $x = Sa + w$, where $w \sim \mathcal{N}(0, \sigma_w^2 I)$, and $Ga = g$ is a set of $P$ linear constraints. $S$ is an $L \times M$ matrix of rank $M$ and $G$ is an $P \times M$ matrix of rank $P$.*

1. *The constrained minimum variance unbiased estimators of $a$ and $\sigma_w^2$ are*

$$\hat{a}_G = A\hat{a} + Bg \quad \text{and} \quad \hat{\sigma}_{w.G}^2 = \frac{1}{L-M} \text{SSE}(\hat{a}_G) \tag{87}$$

*where*

$$A \triangleq I - BG \tag{88}$$

*and*

$$B \triangleq (S^T S)^{-1} G^T \left[ G(S^T S)^{-1} G^T \right]^{-1} \tag{89}$$

2. *The estimates $\hat{a}_G$ and $\hat{\sigma}_{w.G}^2$ are independently distributed as*

$$\hat{a}_G \sim \mathcal{N}(a, \sigma_w^2 A(S^T S)^{-1} A^T) \tag{90}$$

*and*

$$\hat{\sigma}_{w.G}^2 \sim \frac{\sigma_w^2}{L-M} \chi_{L-Q}^2(0)$$

*where $Q = M - P$ is the rank of matrix $A$.*

Since $\hat{a}$ minimizes SSE($a$), it is obvious that SSE($\hat{a}_G$) $\geq$ SSE($\hat{a}$). Also, it can be shown that for any vector $c$ we have var$\{c^T \hat{a}_G\} \leq$ var$\{c^T \hat{a}\}$, that is the variance of the estimate of any component of $\hat{a}$ is decreased by enforcing constraints [22].

The results in Propositions 1 and 2, which can be used to derive confidence intervals and hypothesis tests for the estimates of $a$ and $\sigma_w^2$, will be used to derive detection algorithms in the sequel. We note that, the enforcement of the positivity constraint (68) leads to a nonlinear optimization problem [16].

## 6.3 MODEL VALIDATION

In the previous discussion we have assumed the a-priori availability of the end member spectral matrix $S$. However, in practice end members should be determined from the image data. Brute force exhaustive search methods use geometrical techniques to check whether each pixel can be described as a mixture of spectra from previously tested pixels or whether it is another end member. However, for detection applications statistical end member identification methods are more appropriate (see Section 3.4). Experienced spectroscopists can identify end members using reference spectral libraries; however, this approach requires calibrated data. Note that end member spectra may or may not correspond to "pure" materials. In practice less than 10 end members [1] are usually adequate to model any given area (local modeling) of an image at a given spatial resolution.

The key requirements for "good" end members are: (a) Mixtures of end members should reproduce the spectra of other pixels, and (b) End members must be spectrally distinct from one another (collinear end members lead to numerically unstable unmixing methods).

There are several ways to validate a LMM in practice [1]:

1. Compute and display the "error-fit" image; it should have "small" values and no visible structure.

2. Compute, display, and analyze "fraction" images using topographical and other information.

3. Compute and display "fraction overflow" images. For inadequate models $a_k(n)$ may fall outside the range $0 \leq a_k(n) \leq 1$ of the data volume.

There are several reasons when a model does not fit the data or does not account for the spectral variation in the data (model inadequacy):

1. Some important part of the image, for example another end member, has been missed.

2. One or more of the used end members are inappropriate.

3. Part of the spectral variation in the image is caused by atmospheric, instrumental, or other effects that are extrinsic to the surface of the earth.

We can improve the model by changing the image end members or selecting one or more additional end members.

To evaluate the capacity of LMM to capture the spectral variability of the data across the scene, we fitted the LMM to each scene and computed the norm of the spectral residual vector for each pixel (see Section 6). The low energy of the residuals and the lack of significant structure at the resulting residual images are indications of good performance. The capacity of the LMM to reduce the correlation between different spectral bands is assessed using the correlation matrix of the original data and the corresponding residuals. Finally, the normality of the data and the resulting residuals is assessed using the beta Q-Q plot test described in Section 3.3. Figure 29 shows the results obtained for the mixed GR scene, which are quite representative of the results obtained for several other scenes. Careful examination of these and other results suggests that the LMM captures accurately the statistical variability of the background with the exception of the probability distribution. The Q-Q plots clearly indicate that both the original data and the corresponding residuals *cannot* be adequately described by a multivariate normal distribution. This analysis was intended to provide statistical support for this statement which is widely adopted by researchers in the area.
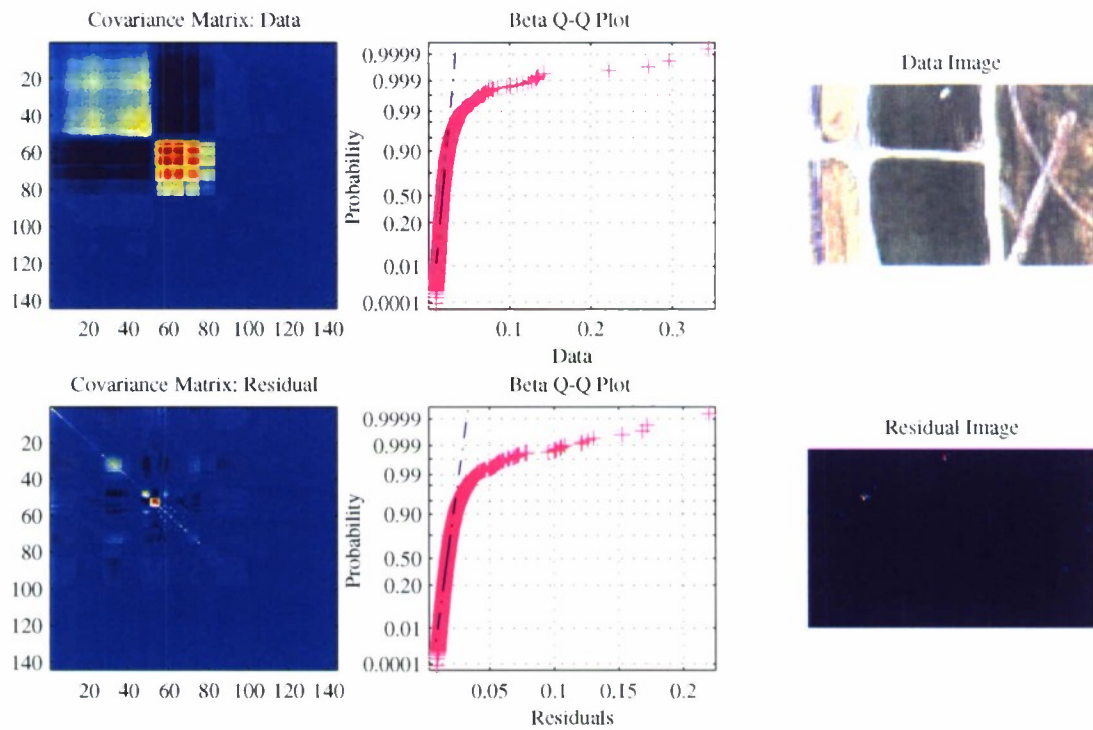


*Figure 29. Statistical analysis of the mixed scene background to evaluate model fit, band decorrelation, and normality.*

# 7.  SUBPIXEL TARGET DETECTION USING SUBSPACE BACKGROUND MODELS

In Section 5 we discussed algorithms for the detection of subpixel targets with or without variability. The variability of the background was modeled statistically using the mean and covariance matrix of target-free data. In this section, we discuss subpixel target detection algorithms which model background variability using a subspace model or the linear mixing model. Both approaches lead to the same mathematical detection algorithms. However, the specification of the background subspace matrix and the interpretation of algorithmic quantities are different. The algorithms discussed here are most useful when endmember spectra are available for both the target and background classes. Otherwise, the covariance-based detection algorithms provide the best practical approach to hyperspectral target detection.

When the background variability is modelled using a subspace model, the target detection problem involves choosing between the following competing hypotheses

$$
\begin{aligned}
H_0 &: \quad x = Ba_b + w, \quad w \sim N(0, \sigma_w^2 I) \quad \text{(Target absent)} \\
H_1 &: \quad x = Sa_t + Ba_b + w \quad\quad\quad\quad\quad \text{(Target present)}
\end{aligned}
\tag{91}
$$

where $S$ ($L \times P$) has to be specified by the user and $B$ ($L \times Q$) has to be determined from the data (see Section ?).

If we wish to determine whether a pixel $x$ contains some material (target) characterized by either a single spectral signature or a linear combination of spectral signatures, we should use of the LMM.

When the target is present, the spectrum of an observed pixel can be decomposed into two components as

$$
x = \sum_{k=1}^{M} a_k s_k + w = \underbrace{\sum_{k=1}^{P} a_k s_k}_{\text{Target } s_t} + \underbrace{\sum_{k=P+1}^{M} a_k s_k}_{\text{Background } s_b} + w
\tag{92}
$$

If $S_t$ and $S_b$ are the matrices formed by the first $P$ and the last $Q$ columns of $S$ ($M = P + Q$) we can write (92) in matrix form as

$$
x = Sa + w \triangleq S_t a_t + S_b a_b + w
\tag{93}
$$

where $a_t$ and $a_b$ consist of the first $P$ and the last $Q$ components of $a$, respectively. Since $s_t = S_t a_t$ we say that the target lies in an $P$-dimensional subspace $\langle S_t \rangle$ of $\mathbb{R}^L$ specified by the columns of $S_t$ (target subspace). Similarly, the $Q$-dimensional subspace $\langle S_b \rangle$ of $\mathbb{R}^L$, specified by the columns of $S_b$, is known as the background subspace.

When the target is absent, the spectrum of the observed pixel is adequately described by

$$
x = S_b a_b + w
\tag{94}
$$

which is a reduced order model. Therefore, to decide whether a given target is present or not, we can fit the full model (40) or the reduced model (94) to the test pixel spectrum and check which model provides a better fit according to a given criterion.

49

The LMM can handle both subpixel and resolved targets and assumes that the target and background subspace matrices $S_t$ and $S_b$ do not depend on the pixel location. There are two extreme special cases of interest: (a) If $P = 1$ we have $s_t = a_1 s_1$ which corresponds to a target having a spectral signature with known shape. (b) If $P = L$ the vectors $a_t$ and $s_t = S_t a_t$ have the same number of components. Therefore, we can think of $s$ as a deterministic $L \times 1$ target with unknown shape [35]. The information about the target decreases as $P$ increases from one to $L$. Therefore, we expect the detection performance to deteriorate with increasing $P$.

Given the target and background subspaces, the target detection problem involves choosing between the hypotheses

$$\begin{aligned} H_0 : & \quad x = S_b a_b + w, \quad w \sim \mathcal{N}(0, \sigma_w^2 I) \quad \text{(Target absent)} \\ H_1 : & \quad x = S_t a_t + S_b a_b + w = Sa + w \quad \text{(Target present)} \end{aligned} \tag{95}$$

which is equivalent to choosing between the "full" LMM $x = Sa + w$ and the "reduced" LMM $x = S_b a_b + w$.

Simple inspection shows that (91) and (95) have the same mathematical form; only the meaning of the involved quantities is different. Clearly, the quantities $a_t$ and $a_b$ can be interpretted as abundances only in the LMM formulation in (95). We next derive detection algorithms for the binary hypotheses in (95). However, the same algorithms can be used for the detection problem defined by (91) with an obvious change in notation. More details and experimental results regarding the performance of detection algorithms based on the LMM can be found in [38].

If we know $a_t$ and $\sigma_w^2$ we can design a Neyman-Pearson detector that maximizes the probability of detection $P_D$ for a given probability of false alarm $P_{FA}$ using the likelihood ratio [24, 42]

$$\text{LR}(x) \triangleq \frac{\mathcal{L}_1(\cdot)}{\mathcal{L}_0(\cdot)} = \frac{\mathcal{L}(a, \sigma_w^2; x)}{\mathcal{L}(a_b, \sigma_{w,b}^2; x)} \tag{96}$$

which will be larger than one whenever a target is present because the full model provides better fit than the reduced model.

When the parameters $a_t$ and $\sigma_w^2$ are unavailable, they should be estimated from the data. If we replace, in (96), the unknown parameters by their ML estimates, the result is known as generalized likelihood ratio. Since there is not any optimality associated with GLR detectors, we should determine their distribution and then use it to evaluate their performance.

## 7.1 CLAIRVOYANT DETECTOR (KNOWN NOISE VARIANCE)

If the variance $\sigma_w^2$ of the noise is known, we only need to determine the MLE of the abundance $a$ under the two hypotheses. However, due to the nature of the noise, the MLE is equivalent to the LSE (78). The full model for a test pixel $x$ is determined by the foot S of the perpendicular PS to the combined target and background subspace (hypothesis $H_1$), whereas the reduced model is specified by the foot B of the perpendicular PB from the pixel to the background subspace (hypothesis $H_0$) (see Figure 30).

Since PS⊥SB we have PB>PS or SSE($\hat{a}$) $\leq$ SSE($\hat{a}_b$). Intuitively, we should accept hypothesis $H_0$ when SSE($\hat{a}$) $\approx$ SSE($\hat{a}_b$), that is when the length SB or the angle $\phi$ are close to zero.

50
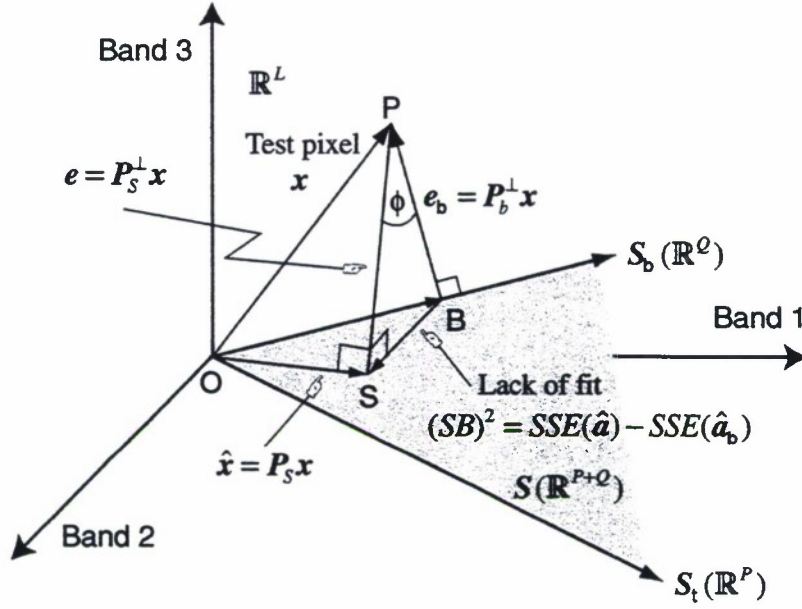
*Figure 30. Geometrical illustration of the operation of subspace matched filter detection.*

The GLR, obtained using (29), (31), and (96) is

$$\text{GLR}(x) = \exp\left\{\frac{\text{SSE}(\hat{a}_b) - \text{SSE}(\hat{a})}{2\sigma_w^2}\right\} \tag{97}$$

For convenience, we use the logarithmic test

$$D(x) = 2\ln \text{GLR}(x) = \frac{\text{SSE}(\hat{a}_b) - \text{SSE}(\hat{a})}{\sigma_w^2} \tag{98}$$

or equivalently

$$D(x) = \frac{x^T(P_b^\perp - P_S^\perp)x}{\sigma_w^2} = \frac{x^T P_b^\perp P_Z P_b^\perp x}{\sigma_w^2} \tag{99}$$

where we have used the property $P_b^\perp - P_S^\perp = P_Z P_b^\perp$ with $Z \triangleq P_b^\perp S_t$ (see Appendix 9).

To make a decision we need to compare $D(x)$ to a given threshold $\eta_0$ and decide $H_1$ (target present) when $D(x) \geq \eta_0$ and $H_0$ (target absent) otherwise. Since the threshold determines both $P_D$ and $P_{FA}$, we need to determine the probability distribution of $T(x)$.

The random vector $P_Z P_b^\perp x$ is distributed as

$$P_Z P_b^\perp x \sim \begin{cases} \mathcal{N}(0, \sigma_w^2 P_Z), & H_0 \\ \mathcal{N}(P_b^\perp S_t a_t, \sigma_w^2 P_Z), & H_1 \end{cases} \tag{100}$$

51

under the two hypotheses. In general, if $x \sim \mathcal{N}(\mu, \Gamma)$ then $x^T \Gamma^{-1} x \sim \chi_r^2(\mu^T \Gamma^{-1}\mu)$, where the noncentral chi-square distribution has noncentrality parameter $\mu^T \Gamma^{-1}\mu$ and $r = \mathrm{rank}(P_Z P_b^\perp)$ degrees of freedom [24]. Using this result, we can easily show that

$$D(x) = \frac{\left\| P_Z P_b^\perp x \right\|^2}{\sigma_w^2} \sim \begin{cases} \chi_P^2(0), & H_0 \\ \chi_P^2(\mathrm{SINR_o}), & H_1 \end{cases} \tag{101}$$

where

$$\mathrm{SINR_o} = \frac{(S_t a_t)^T P_b^\perp (S_t a_t)}{\sigma_w^2} = \frac{\left\| P_b^\perp (S_t a_t) \right\|^2}{\sigma_w^2} \tag{102}$$

is the noncentrality parameter of the chi-squared distribution. Since the term $P_b^\perp(S_t a_t)$ is the component of the target which is orthogonal to the background subspace, $\mathrm{SINR_o}$ can be interpreted as the theoretical SINR.

The false alarm and detection probabilities are

$$P_{\mathrm{FA}} = \Pr\left[\chi_P^2(0) \geq \eta_0\right] \tag{103}$$

$$P_{\mathrm{D}} = \Pr\left[\chi_P^2(\mathrm{SINR_o}) \geq \eta_0\right] \tag{104}$$

where $\eta_0$ is the detection threshold. The obtained detector has the constant false alarm rate (CFAR) operation property, because $\eta_0$ depends only on $P$. The probability of detection $P_{\mathrm{D}}$ cannot be maximized for a given $P_{\mathrm{FA}}$ because it depends on $a_t$, which is unavailable.

## 7.2 ADAPTIVE DETECTOR (UNKNOWN NOISE VARIANCE)

When $\sigma_w^2$ is unknown it has to be estimated from the available data. In this case we determine the GLR by replacing $a_t$ and $\sigma_w^2$ by their ML estimates under each hypothesis. Therefore, using (33) we obtain

$$\mathrm{GLR}(x) \triangleq \frac{\mathcal{L}(\hat{a}, \hat{\sigma}_w^2; x)}{\mathcal{L}(\hat{a}_b, \hat{\sigma}_{w,b}^2; x)} = \left[\frac{\mathrm{SSE}(\hat{a}_b)}{\mathrm{SSE}(\hat{a})}\right]^{L/2}$$

$$= \left(\frac{x^T P_b^\perp x}{x^T P_S^\perp x}\right)^{L/2} = (\cos\phi)^{-L} \tag{105}$$

which can be used to obtain a target detection algorithm. Since $\mathrm{SSE}(\hat{a}) \leq \mathrm{SSE}(\hat{a}_b)$, we have $\mathrm{GLR}(x) \geq 1$.

To make a decision we need to compare $\mathrm{GLR}(x)$ to a given threshold $\ell_0$ and decide $H_1$ when $\mathrm{GLR}(x) \geq \ell_0$ and $H_0$ otherwise. Since the threshold determines both $P_D$ and $P_{FA}$, we need to determine the probability distribution of $\mathrm{GLR}(x)$. Since PS$\perp$SB, the random vectors $e = P_S^\perp x$ and $\overrightarrow{\mathrm{BS}} = e_b - e = (P_b^\perp - P_S^\perp)x$ are orthogonal (that is uncorrelated). Furthermore, since $e$ and $\overrightarrow{\mathrm{BS}}$ are normal (as linear transformations of normal vector $x$) they are also independent. Therefore, in practice we use the ratio

$$\tilde{D}(x) \triangleq \frac{x^T(P_b^\perp - P_S^\perp)x}{x^T P_S^\perp x} = \frac{(\mathrm{BS})^2}{(\mathrm{PS})^2}$$

$$= (\tan\phi)^2 = [\mathrm{GLR}(x)]^{2/L} - 1 \tag{106}$$

52

whose distribution can be more easily determined due to the independence of numerator and denominator terms. Indeed, it can be shown [24,51] that $x^T(P_b^\perp - P_S^\perp)x/(\sigma_w^2 P) \sim x_P^2(\text{SINR}_o)$ and $x^T P_S^\perp x / [\sigma_w^2(L - P - Q)] \sim x_{L-P-Q}^2(0)$. Note that $(\text{BS})^2 = \|e_b\|^2 - \|e\|^2$ and $(\text{PS})^2 = \|e\|^2$. Since the two quadratic forms are independent

$$D(x) = \frac{(\text{BS})^2}{(\text{PS})^2} \frac{L - P - Q}{P} \sim F_{P,L-P-Q}(\text{SINR}_o) \qquad (107)$$

where $F_{P,L-P-Q}(\text{SINR}_o)$ is noncentral F distribution with $P$ numerator degrees of freedom, $L - P - Q$ denominator degrees of freedom, and noncentrality parameter $\text{SINR}_o$, see (102), which depends on the unknown variance $\sigma_w^2$. Notice that $\sigma_w^2$ is not required for the computation of the detection statistics because it cancels out. In the statistical literature $T(x)$ is denoted by $F(x)$ and is known as the F-test[1]. The detection test is given by

$$D(x) = \frac{(\text{BS})^2}{(\text{PS})^2} \frac{L - P - Q}{P} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_0 \qquad (108)$$

where the threshold $\eta_0$ is specified[2] by the required probability of false alarm

$$P_{\text{FA}} = 1 - F_{P,L-P-Q}(0, \eta_0) \qquad (109)$$

because $\text{SINR}_o = 0$ under $H_0$. Since the distribution of $T(x)$ under $H_0$ is known, we can set the threshold $\eta_0$ to attain CFAR operation. The probability of detection

$$P_D = 1 - F_{P,L-P-Q}(\text{SINR}_o, \eta_0) \qquad (110)$$

depends on the unknown $a_t$, $\sigma_w^2$ and therefore it cannot be maximized to obtain an optimum Neyman-Pearson detector.

It can be shown that, independently of the normality assumption, the detectors maximize the signal to interference ratio [24, 39]. Furthermore, it has been proved that "within the class of invariant detectors which have the same invariances as the GLR test, the GLR test is uniformly most powerful (UMP) invariant. This is the strongest statement of optimality one could hope to make for a detector" [51].

## 7.3  THEORETICAL PERFORMANCE

To get some insight into the performance of the adaptive GLRT subspace detector we consider a background with $Q = 5$ end members. The number of bands is $L = 144$ and the probability of false alarm is fixed at $P_{\text{FA}} = 10^{-6}$. The dimension of the target subspace is varied from $P = 1$ to $P = 9$ in steps of 2. Figure 31 shows plots of the probability of detection as a function of the SINR for fixed values of $P_{\text{FA}}$, $Q$, $P$, and $L$. The family of curves, which is parameterized with respect to $P$, shows that performance

---

[1]The GLRT$(x)$ and $D(x)$ tests are monotonically related and the distribution of the GLR can be expressed in terms of the beta distribution by exploiting its relationship with the F distribution [22, 23].

[2]With a slight abuse of notation, we use $F_{P,L-P-Q}(\text{SINR}_o, \eta_0)$ to denote the cumulative distribution function of the non-central $F$ random variable.

deteriorates as the dimensionality of the target subspace increases (that is as the a-priori information about the target decreases) as is expected. The figure also includes curves illustrating the performance of the Neyman-Pearson and clairvoyant detectors. We see that for a given probability of detection, there is a loss in SINR due to the need to estimate the target abundance (clairvoyant detector) plus the noise variance (adaptive detector). Figure 32 shows a similar family of curves parameterized with respect to the number of spectral
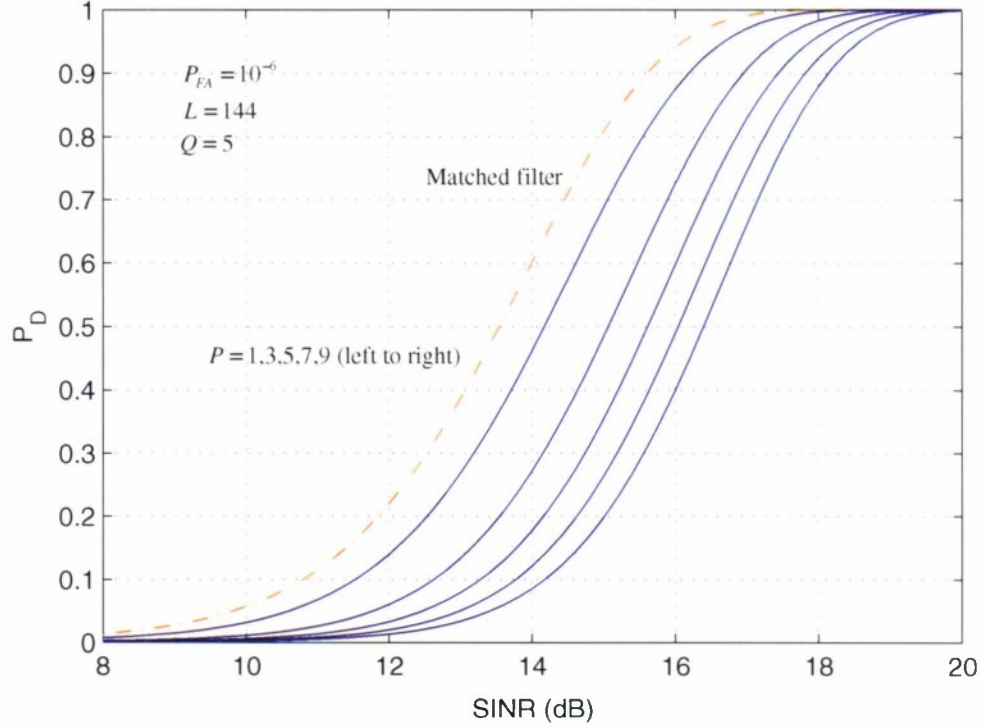


*Figure 31. Probability of detection as a function of SINR illustrating the effect of target subspace dimensionality.*

bands $L$, all other parameters being held fixed. As it is intuitively expected, performance improves as the number of band increases, that is as we supply more information into the detector. Clearly, the improvement is more dramatic when we start adding more bands and becomes less significant after a certain point. The performance of the Neyman-Pearson and clairvoyant detectors does not depend on the number of bands.

## 7.4 IMPLEMENTATION

An interesting geometrical interpretation can be obtained by considering the part $Z \triangleq P_b^\perp S_t$ of the target subspace that is orthogonal to the background subspace [51]. Indeed, using the identities $P_S^\perp = P_b^\perp P_Z^\perp P_b^\perp$, $P_b^\perp - P_S^\perp = P_Z P_b^\perp$, and (105), we have

$$\tilde{T}_{\text{AMSD}}(x) = \frac{x^T P_b^\perp P_Z P_b^\perp x}{x^T P_b^\perp P_Z^\perp P_b^\perp x} = \frac{\left\| P_Z P_b^\perp x \right\|^2}{\left\| P_Z^\perp P_b^\perp x \right\|^2} \tag{111}$$
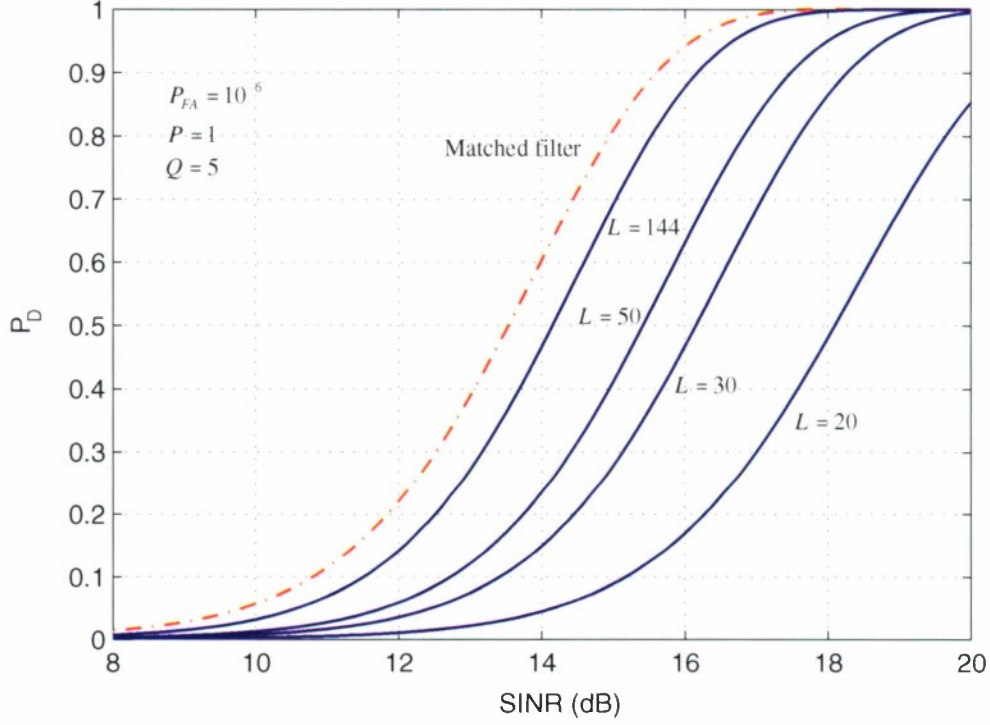
*Figure 32. Probability of detection as a function of SINR illustrating the effect of number of spectral bands.*

which suggests the computational structure shown in Figure 33.

## 7.5   ONE DIMENSIONAL TARGET SUBSPACES ($P = 1$)

When $P = 1$ the full model parameter vector takes the form $\boldsymbol{a} = [a_\mathrm{t}\, \boldsymbol{a}_\mathrm{b}^T]^T$, where $a_\mathrm{t}$ is the abundance of the target signature vector $\boldsymbol{s}_\mathrm{t}$. If we use the inversion lemma for partitioned matrices we can show that [23,39]

$$\hat{a}_\mathrm{t} = \frac{\boldsymbol{s}_\mathrm{t}^T \boldsymbol{P}_\mathrm{b}^\perp \boldsymbol{x}}{\boldsymbol{s}_\mathrm{t}^T \boldsymbol{P}_\mathrm{b}^\perp \boldsymbol{s}_\mathrm{t}} \sim \mathcal{N}(a_\mathrm{t}, \sigma_w^2\,(\boldsymbol{s}_\mathrm{t}^T \boldsymbol{P}_\mathrm{b}^\perp \boldsymbol{s}_\mathrm{t})^{-1}) \tag{112}$$

The detection test takes the form

$$D(\boldsymbol{x}) = \frac{\hat{a}_\mathrm{t}^2}{\mathrm{SSE}(\hat{\boldsymbol{a}})/(L - 1 - Q)} \boldsymbol{s}_\mathrm{t}^T \boldsymbol{P}_\mathrm{b}^\perp \boldsymbol{s}_\mathrm{t} \tag{113}$$

and it is distributed according to $F_{1,L-1-Q}(\mathrm{SINR}_0)$ which is the square of the $t_{L-1-Q}$ Student's t-distribution. Therefore, we can use the quantity

$$t(\boldsymbol{x}) = \hat{a}_\mathrm{t} \left[ \frac{(\boldsymbol{s}_\mathrm{t}^T \boldsymbol{P}_\mathrm{b}^\perp \boldsymbol{s}_\mathrm{t})}{\mathrm{SSE}(\hat{\boldsymbol{a}})/(L - 1 - Q)} \right]^{1/2} \tag{114}$$

55

*Figure 33. Computational structure for the implementation of the adaptive subspace detector for structured background.*

to develop a bidirectional test that takes into consideration the sign of $\hat{a}_t$. Clearly, detection algorithms based on $\hat{a}_t$ are not CFAR because $\sigma_w^2$ is unknown. Furthermore, because the distribution under $H_1$ depends on $a_t$ and $\sigma_w^2$, which are both unavailable the probability of detection cannot be maximized to attain Neyman-Pearson optimality.

## 7.6 ORTHOGONAL SUBSPACE PROJECTION (OSP) AND RELATED ALGORITHMS

The numerator of the abundance estimator (112) given by has been proposed in [20] as a detection statistic under the OSP name. However, it is better to use the normalized statistic

$$D_{\text{OSP}}(x) = \frac{s_t^T P_b^\perp x}{s_t^T P_b^\perp s_t} \tag{115}$$

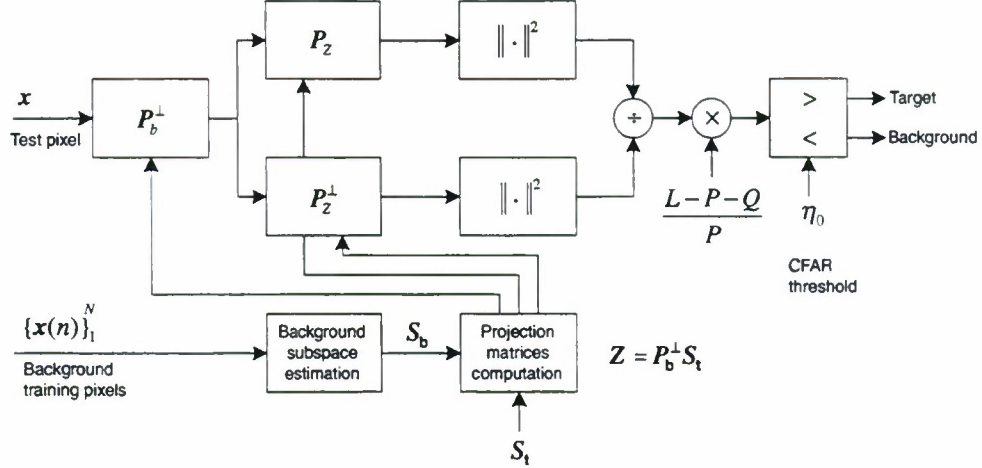because it corresponds to the abundance estimate for the wanted target. The resulting algorithm is not CFAR because the abundance of the target and the variance of the sensor noise are unknown. Clearly, the theoretically predicted Gaussian distribution of the target abundance contradicts with the physical constraint $0 \le a_t \le 1$.

The operation $e_b = P_b^\perp x$ removes from $x$ the part which belongs to the background $S_b$. The decision is based on the correlation $y = s_t^T e_b$ between the residual "target plus noise" and the target template $s_t$, which is proportional to the angle between those two vectors. Alternatively, we can write $y = c_{\text{bmf}}^T x$ where $c_{\text{bmf}} \triangleq P_b^\perp s_t$ is the background-based matched filter. If $S_b = 0$ (no background), we have $c_{\text{bmf}} = s_t$, that is the typical matched filter for white noise. Some additional subspace projection algorithms can be expressed as functions of $y = s_t^T P_b^\perp x$ and some other quantities [6]. However, if we adopt the signal model (95), the GLRT detector given by (113) follows naturally. This detector maximizes the SINR for any noise distribution and is CFAR in the case of Gaussian noise. Thinking of the optimum detector as a whitening filter followed by a matched filter for white noise and trying to separately design these two filters for the scenario described by (95) does not always lead to the GLRT or some other detector with theoretically desired properties (e.g., CFAR).

56

If there is no background or we choose to ignore its existence, i.e., we set $S_b = 0$, the statistic (106) reduces to

$$D_{\mathrm{MF}}(x) = \tilde{D}(x)|_{S_b=0} = \frac{x^T P_t x}{x^T P_t^\perp x} \tag{116}$$

which is the well-known matched filter for signals in white noise.

## 7.7  GENERAL LINEARLY CONSTRAINED DETECTORS

As we saw in Section 6 the reduced LMM can be formulated as a special case of the linearly constrained LMM. In general, the two competing hypotheses are

$$x = Sa + w, \quad w \sim \mathcal{N}(0, \sigma_w^2 I) \tag{117}$$

with

$$
\begin{aligned}
H_0 &: \quad Ga = g \quad \text{(Target absent)} \\
H_1 &: \quad Ga \neq g \quad \text{(Target present)}
\end{aligned}
\tag{118}
$$

and the test statistics is given by (108) with the following changes

$$(BS)^2 = (G\hat{a} - g)^T \left[ G(S^T S)^{-1} G^T \right]^{-1} (G\hat{a} - g) \tag{119}$$

$$\mathrm{SINR_o} = \frac{1}{2\sigma_w^2}(Ga - g)^T \left[ G(S^T S)^{-1} G^T \right]^{-1} (Ga - g) \tag{120}$$

where $a$ is the true parameter vector and $\hat{a}$ its LSE estimate under hypothesis $H_1$ [24]. Algorithm (108) follows from the present algorithm using the special values of $G$ and $g$ given in (86).

Since the additivity constraint is important in hyperspectral signal modelling, we propose a new subspace detection algorithm that takes this restriction into consideration. This can be done by imposing the additional additivity constraint $u^T a = 1$ ($u$ is a column vector with unit components) into both hypotheses in (118). This issue is further investigated in [41].

57

## 8. ALGORITHM TAXONOMY AND PRACTICAL CONSIDERATIONS

Table 3 provides a taxonomy of the most fundamental target detection algorithms discussed in this report. We have focused on algorithms derived using well established statistical models and procedures. A number of additional algorithms reported in the literature can be obtained from this table through mathematical approximations or simplifications. For example, low-rank approximations or diagonal loading of the estimated background covariance matrix, lead to algorithms which may provide improved performance under certain practical circumstances. Also, use of clustering algorithms to derive more homogeneous background segments has been shown to improve detection performance in many cases [40].

The implementation of hyperspectral target detection algorithms in a real-world environment involves confrontation with many "practical details" and challenges that result from the violation of the theoretical assumptions used for the derivation of the various algorithms.

The main processing steps for the subspace-based and covariance-based families of target detection algorithms are shown in Figure 34. We notice that we use the correlation matrix to determine the background subspace and the covariance matrix for the matched filter type detectors.
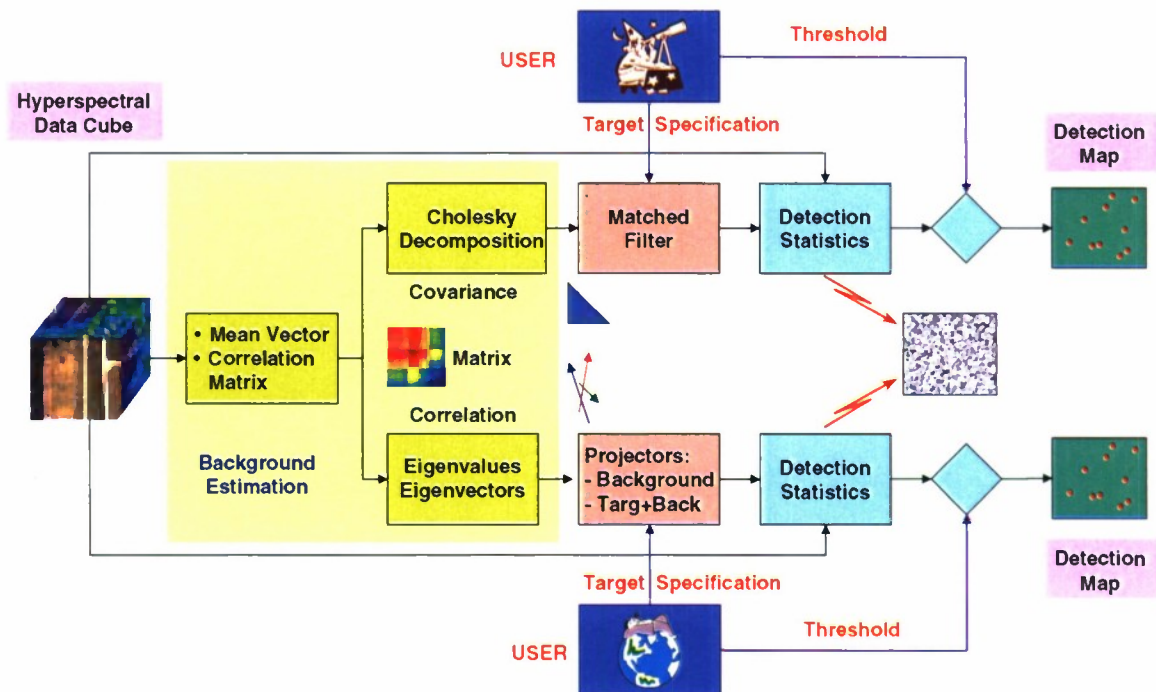


Figure 34. Block diagram illustrating the major processing steps used by covariance and subspace based target detection algorithms.

The key assumptions used for the detectors using the background subspace model are (a) the spectrum of every pixel can be adequately modelled by the LMM, (b) the modelling error has uncorrelated components with the same variance, (c) the distribution of the modelling error is multivariate normal, (d) the background

59

subspace $B$ is perfectly known, and (e) the target subspace template $S$ is known.

The key assumptions used for the detectors using the covariance matrix of the background are (a) the background is homogeneous and can be modelled by a multivariate normal distribution, (b) the background spectrum interfering with the test pixel spectrum has the same covariance matrix with the background training pixels, (c) the test and training pixels are independent, and (d) the target and background spectra interact in an additive manner (additive instead of replacement model).

The performance evaluation of detection algorithms in practice is challenging due to the limitations imposed by the limited amount of target data. As a result the establishment of accurate ROC curves is quite difficult. Since the number of background pixels is much larger than the number of target pixels, we use the plot illustrated in Figure 11. Such plots give an informative pictorial summary of the performance of each detector. Figures 35 and 36 show an example of such plots for various covariance and subspace based detectors.

We shall compare the various algorithms in terms of their ability to operate in CFAR mode and the enhancement of the separation between targets and background they provide (see Figure 37). The CFAR property depends on the capability to accurately model the detection statistics $D_{(?)}(x)$ of the background pixels for a given algorithm. The enhancement of target visibility will be illustrated visually by plotting the detection statistics for every pixel in the scene.

To investigate the performance of the various detectors, we assess their capacity (a) to enhance the "visibility" of the desired target (background-target separation) and (b) to accurately model the background statistics.

To illustrate various issues regarding the discussed algorithms, we shall use the Forest Radiance I data (see Figure 6) collected with the HYDICE sensor. We have selected the three outlined areas to investigate three different types of background: grass (G), trees (T), and mixed grass-road (GR). The first two scenes are relatively homogeneous, whereas the third scene consists of a non-homogeneous background including different types of grass and roads. Also considered were classes resulting from a supervised classification process performed to isolate spectrally similar (not necessarily spatially adjacent) pixels. Data from classes selected from this analysis were labelled using a number after the class label, for example "grass2".

With regard to targets we have chosen a target with 37 pixels from a certain vehicle. The mean value of the set is used as the target template $s_t$ required for the implementation of each detection algorithm.

The estimate of the background subspace $\hat{B}$ for each scene was obtained using the first $Q$ significant eigenvectors of the estimated correlation matrix of the background pixels (the "known" target pixels were excluded). We have used the first $Q = 10$ eigenvectors for the scenes G, T, and GR, respectively. These eigenvectors capture more than 99 percent of the data cube energy in each case.

The bar charts in Figure 38 provide the range of the detection statistic of the target and the maximum value of the background detection statistics for various backgrounds. The target-background separation or overlap is the quantity used to evaluate target visibility enhancement. For example, it can be seen that the ACE detector performs better than the OSP algorithm for the six data sets shown.

The Q-Q plots in Figure 39 illustrate the comparison between the experimental detection statistics to the theoretically predicted ones for various detection algorithms. We notice that the CEM (matched

filter) and OSP statistics is compared to the theoretically expected normal distribution, whereas the ACE and ASD statistics are compared to the theoretically expected F-distribution. Clearly, there is a deviation between postulated and observed statistics. This deviation varies for different combinations of targets and backgrounds.

The previous results dealt with full-pixel or resolved targets. To evaluate detection performance for subpixel targets, we have simulated subpixel targets using formula (3). Subpixel targets were simulated by adding a randomly chosen target pixel from the target pixel set to each of the background pixels at a constant fraction. This simulation process is illustrated in Figure 40. The results shown in Figure 41, show target-background separability as a function of the target fill factor $a$ for the ACE and OSP detectors. Clearly, target visibility improves with the size of the target. A more detailed comparison of a large set of detection algorithms is provided in [36]. It has been shown that taking into consideration target variability using a subspace model can increase detection performance [37].

TABLE 3. Taxonomy of HSI target detection algorithms.

| Target Resolution | Signal Model | Assumptions | Detector $y = D(x)$ | Name | Comments |
|---|---|---|---|---|---|
| Full pixel targets | $H_0: x \sim N(\mu_b, \Gamma_b)$ <br> $H_1: x \sim N(\mu_t, \Gamma_t)$ | Known <br> $\mu_b, \mu_t, \Gamma_b, \Gamma_t$ | $(x-\mu_b)^T \Gamma_b^{-1}(x-\mu_b) -$ <br> $(x-\mu_t)^T \Gamma_t^{-1}(x-\mu_t)$ | Bayes or Neyman-Pearson quadratic classifiers | $\Gamma \triangleq \Gamma_t = \Gamma_b \Rightarrow$ <br> $y = (\mu_t - \mu_b)^T \Gamma^{-1} x$ <br> Plug-in detectors: estimate $\mu_t, \mu_b, \Gamma_t, \Gamma_b$ from training data |
| Full pixel targets | | Training data for $\mu_b, \Gamma_b$ only | $(x-\hat{u})^T \hat{\Gamma}^{-1}(x-\hat{u})$ | Mahalanobis distance | Anomaly detection <br> RX algorithm <br> CFAR |
| Subpixel targets | $H_0: x = b$ <br> $H_1: x = Sa + \sigma b$ <br> $S$ known $(L \times P)$ matrix <br> $H_0: x \sim N(0, \Gamma)$ <br> $H_1: x \sim N(Sa, \sigma^2\Gamma)$ | Training data under <br> $H_0: x(1), x(2), \ldots, x(N)$ <br> $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x(n)$ <br> $\hat{\Gamma} = \frac{1}{N}\sum_{n=1}^{N}[x(n)-\hat{u}][x(n)-\hat{u}]^T$ | $\dfrac{x^T\hat{\Gamma}^{-1}S(S^T\hat{\Gamma}^{-1}S)^{-1}S^T\hat{\Gamma}^{-1}x}{\psi_1 + \psi_2 x^T\hat{\Gamma}^{-1}x}$ | <table><tr><td>Name</td><td>$\psi_1$</td><td>$\psi_2$</td><td>$\sigma$</td></tr><tr><td>Kelly</td><td>$N$</td><td>1</td><td>1</td></tr><tr><td>AMF</td><td>1</td><td>0</td><td>1</td></tr><tr><td>ACE</td><td>0</td><td>1</td><td>?</td></tr></table> ? = unknown <br> AMF = Adaptive Matched Filter <br> ACE = Adaptive Cosine/Coherence Estimator | CFAR <br> $SINR_o = (Sa_t)^T\Gamma^{-1}(Sa_t)$ <br> $y = \kappa s^T\hat{\Gamma}^{-1}x$ <br> (Matched filter) <br> Low-rank covariance matrix approximation: <br> $\hat{\Gamma} = \sum_{k=1}^{\varrho}\lambda_k u_k u_k^T \Rightarrow \hat{\Gamma}^{-1} \approx \sum_{k=1}^{\varrho}\frac{1}{\lambda_k}u_k u_k^T$ |
| Subpixel targets | | $P = 1 \Rightarrow S \to s$ | $\dfrac{s^T\hat{\Gamma}^{-1}x}{(s^T\hat{\Gamma}^{-1}s)^{1/2}(\psi_1 + \psi_2 x^T\hat{\Gamma}^{-1}x)^{1/2}}$ | | |
| Subpixel targets | $H_0: x = Ba_b + w$ <br> $H_1: x = Sa_t + Ba_b + w$ | $w \sim N(0, \sigma_w^2 I)$ <br> $Z \triangleq [S\ B]$ <br> $P_A^\perp \triangleq I - A(A^T A)^{-1}A^T$ | $\dfrac{x^T(P_B^\perp - P_Z^\perp)x}{x^T P_Z^\perp x}$ | GLRT detector | CFAR <br> $SINR_o = \dfrac{\|P_B^\perp Sa_t\|}{\sigma_w^2}$ |
| Subpixel targets | | | $sP_B^\perp x$ | Orthogonal subspace projector (OSP) | Non CFAR |

62

*Figure 35. Illustration of covariance-based detectors performance using canonical data sets.*

63
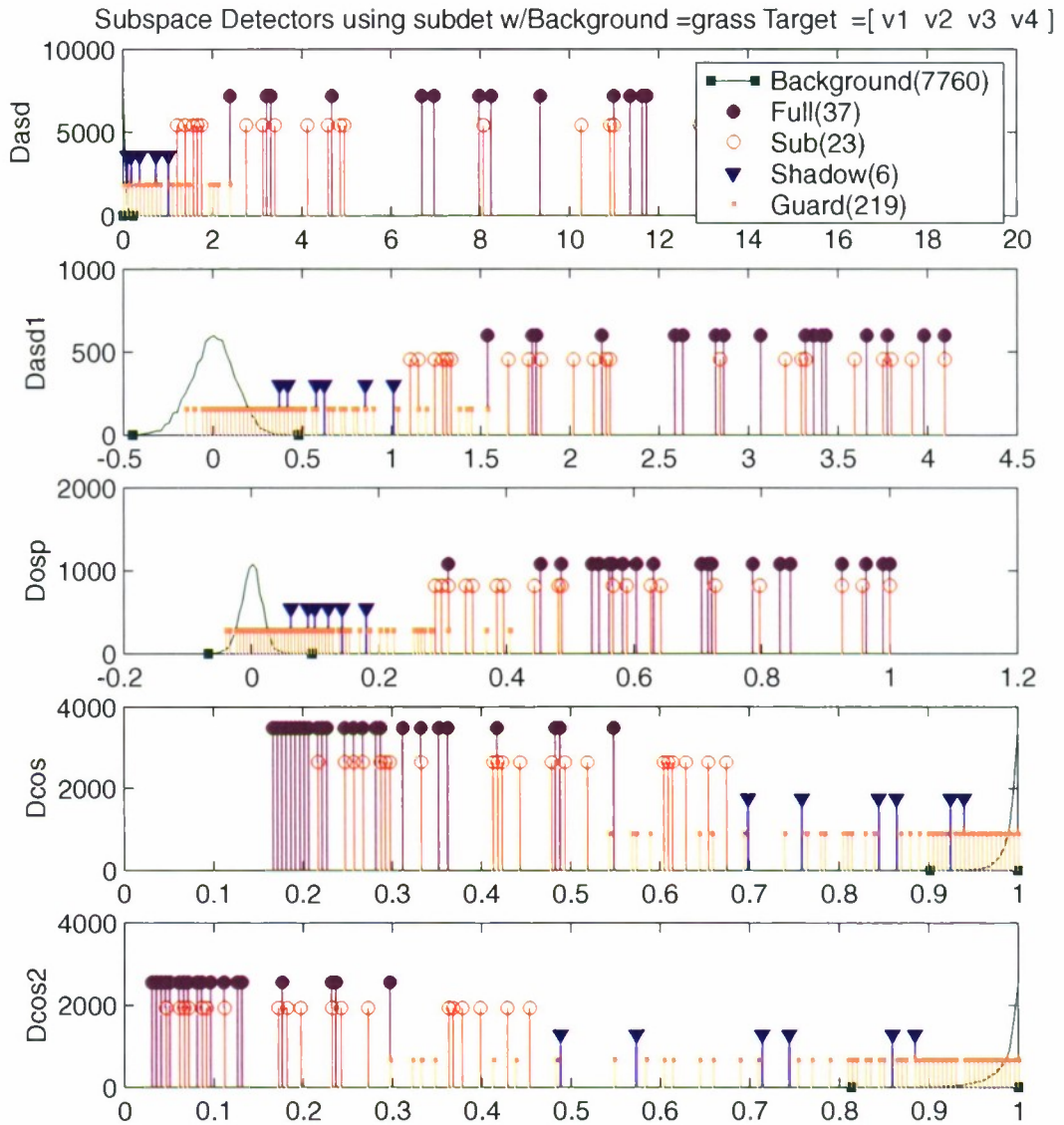
*Figure 36. Illustration of subspace detector performance using canonical data sets.*
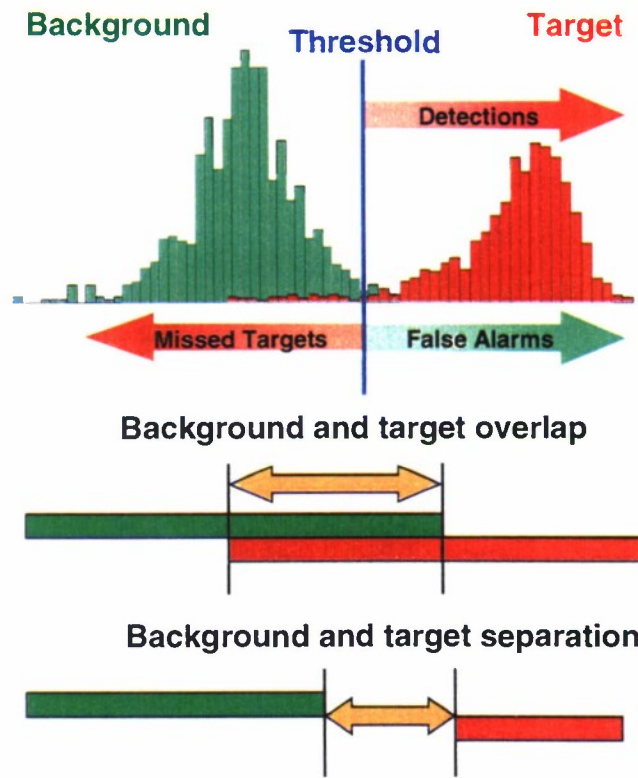
*Figure 37. Illustration of target visibility concept.*
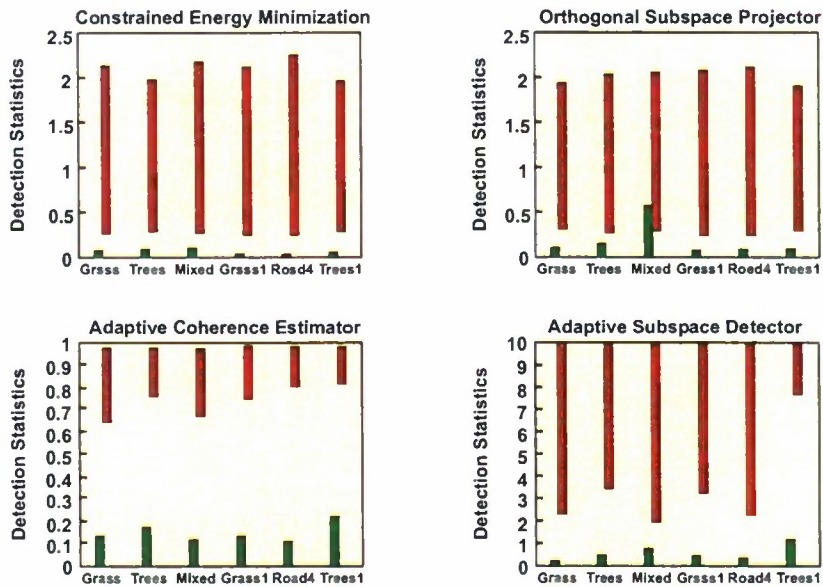


*Figure 38. Illustration of target visibility enhancement for different target detection algorithms.*
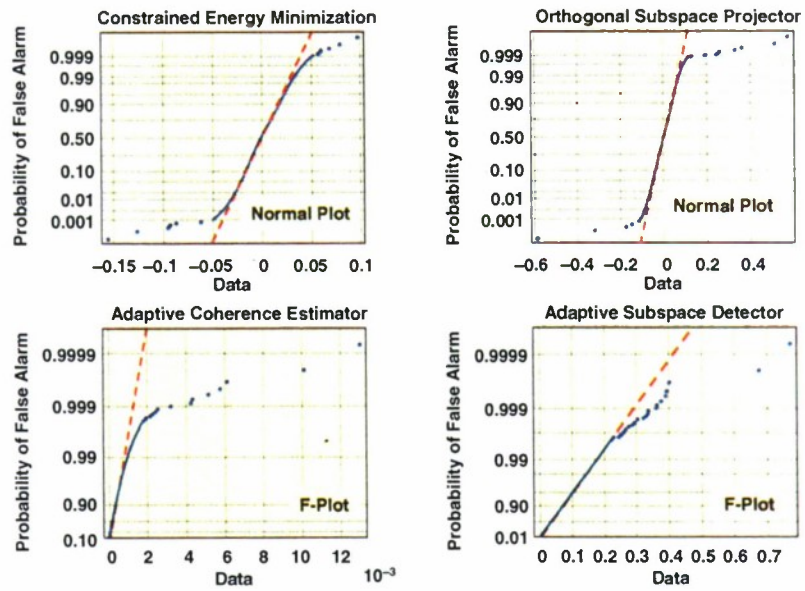
*Figure 39. Evaluation of the matched filter CFAR property for two different backgrounds.*
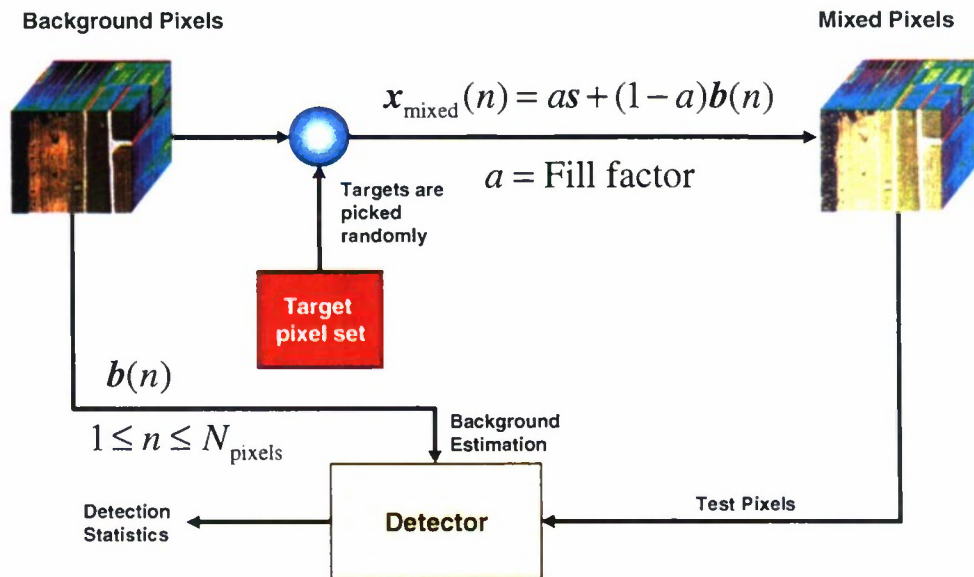


*Figure 40. Block diagram for the generation of subpixel targets for target detection studies.*

66
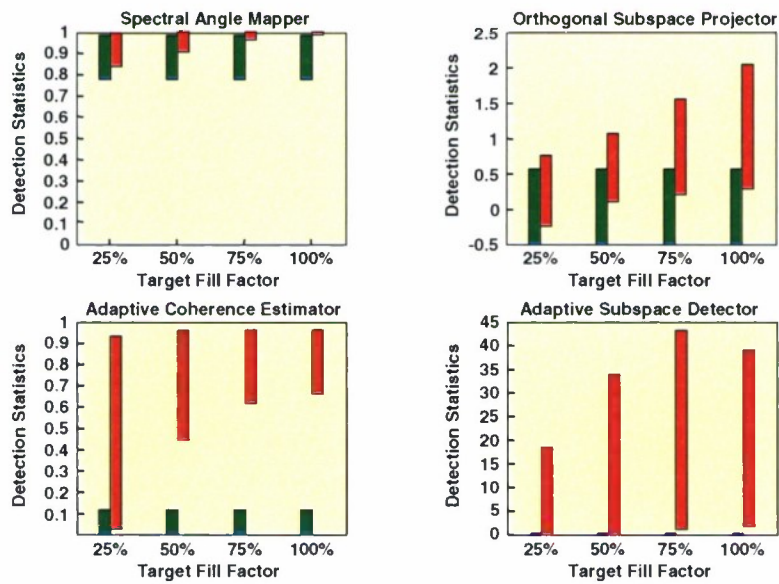
*Figure 41. Target-background separability as a function of the target fill factor.*

# 9. CONCLUSIONS

This report provided a tutorial review of the state of the art of target detection algorithms for hyperspectral imaging applications. The main obstacles in the development of effective detection algorithms are the inherent variability target and background spectra. The use of adaptive algorithms deals quite effectively with the problem of unknown background; however, the lack of sufficient target data makes the development and estimation of target variability models challenging. Hyperspectral target detection is a multidisciplinary problem that draws upon different scientific and engineering areas. However, there is a significant signal processing component where expertise from the areas of array processing, radar, and detection theory will provide crucial contributions to future progress.

# APPENDIX A
## USEFUL PROJECTION MATRIX IDENTITIES

Using the following matrix inversion formula for partitioned matrices [24, 42]

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} E^{-1} & -A^{-1}BH^{-1} \\ -H^{-1}CA^{-1} & H^{-1} \end{bmatrix} \tag{1}$$

where

$$E = A - BD^{-1}C \tag{2}$$

we can easily obtain the recursion

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -D^{-1}C \end{bmatrix}$$
$$\cdot E^{-1} \begin{bmatrix} I & -BD^{-1} \end{bmatrix} \tag{3}$$

Using the partitioning

$$S = \begin{bmatrix} S_1 & S_2 \end{bmatrix} \tag{4}$$

we have

$$S^T S = \begin{bmatrix} S_1^T S_1 & S_1^T S_2 \\ S_2^T S_1 & S_2^T S_2 \end{bmatrix} \tag{5}$$

and

$$E = S_1^T S_1 - S_1^T S_2 (S_2^T S_2)^{-1} S_2^T S_1 = S_1^T P_2^\perp S_1 \tag{6}$$

Substitution into (3) gives

$$(S^T S)^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & (S_2^T S_2)^{-1} \end{bmatrix} +$$
$$\begin{bmatrix} I \\ -(S_2^T S_2)^{-1} S_2^T S_1 \end{bmatrix}$$
$$\cdot (S_1^T P_2^\perp S_1)^{-1} \begin{bmatrix} I & -S_1^T S_2 (S_2^T S_2)^{-1} \end{bmatrix}$$

Using this recursion and the definition of the projection matrix, we have

$$P = S(S^T S)^{-1} S^T = P_2^\perp + P_2^\perp S_1 (S_1^T P_2^\perp S_1)^{-1} S_1^T P_2^\perp$$

71

or

$$P = P_2 + P_{P_2^\perp S_1} \tag{7}$$

which provides a decomposition of the observation space into two orthogonal subspaces.

If we define $Z \triangleq P_2^\perp S_1$, we can easily see that

$$
\begin{aligned}
P_Z = Z(Z^T Z)^{-1} Z^T &= P_2^\perp S_1 (S_1^T P_2^\perp S_1)^{-1} S_1^T P_2^\perp \\
&= P_2^\perp P_Z = P_Z P_2^\perp = P_2^\perp P_Z P_2^\perp
\end{aligned} \tag{8}
$$

because $P_2^\perp P_2^\perp = P_2^\perp$. Using the previous two identities, we can also show that

$$P_2^\perp - P^\perp = P_2^\perp P_Z P_2^\perp \tag{9}$$

and

$$
\begin{aligned}
P^\perp &= P_2^\perp - P_2^\perp P_Z P_2^\perp \\
&= P_2^\perp (I - P_Z) P_2^\perp = P_2^\perp P_Z^\perp P_2^\perp
\end{aligned} \tag{10}
$$

which are used for the implementation of the subspace detector.

72

# GLOSSARY

| | |
|---|---|
| ACE | Adaptive Cosine/Coherence Estimator |
| AMF | Adaptive Matched Filter |
| CFAR | Constant False Alarm Rate |
| ECD | Elliptically Contoured Distribution |
| GLRT | Generalized Likelihood Ratio Test |
| LMM | Linear Mixing Model |
| MLE | Maximum Likelihood Estimator |
| OSP | Orthogonal Subspace Projector |
| Q-Q | Quantile-Quantile |
| ROC | Receiver Operating Characteristic |
| SINR | Signal to Interference plus Noise Ratio |
| SSE | Sum of Squared Errors |

# REFERENCES

[1] J. B. Adams, M. O. Smith, and A. R. Gillespie. Remote geochemical analysis: Elemental and mineralogical composition. In C. M. Pieters and P. A. J. Englert, editors, *Imaging spectroscopy: Interpretation based on spectral mixture analysis*, pages 145–166, Cambridge, 1993. Cambridge University Press.

[2] R. J. Adler, R. E. Feldman, and M. S. Taqqu. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhauser, Boston, 1998.

[3] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, NY, 1984.

[4] Y. Bresler. Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing. pages 172–175. Fourth Annual ASSP Workshop on Spectral Estimation and Modeling, 1988.

[5] B. D. Carlson. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE AES*, 24(4):397–401, 1988.

[6] C. Chang, X. Zhao, M. Althouse, and J. Pan. Least squares subspace projection approach to mixed pixel classification for hyperspectral images. *IEEE Trans. on Geoscience and Remote Sensing*, 36(3):898–912, May 1998.

[7] W. Chen and I. S. Reed. A new CFAR detection test for radar. *Digital Signal Processing*, 1(4):198–214, 1991.

[8] O. P. Cheremisin. Efficiency of an adaptive algorithm with regularization of the sample covariance matrix. *Radio Engineering and Electronic Physics*, (10):69–77, 1982.

[9] E. Conte, M. Lops, and G. Ricci. Asymptotically optimum radar detection in compound-Gaussian clutter. *IEEE Trans. on Aerospace and Electronic Systems*, 31(2):617–625, 1995.

[10] K. T. Fang, S. Kotz, and K. W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, New York, 1990.

[11] W. H. Farrand and J. C. Harsanyi. Discrimination of poorly exposed lithologies in imaging spectrometer data. *Journal of Geophysical Research*, 100(E1):1565–1578, January 1995.

[12] W. H. Farrand and J. C. Harsanyi. Mapping the distribution of mine tailings in the Coeur D' Alene River Valley, Idaho, through the use of a constrained energy minimization technique. *Remote Sens. Environ.*, 59:64–76, 1997.

[13] R. A. Fisher. The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[14] K. Fukunaga. *Introduction to Statistical Patern Recognition*. Academic Press, san Diego, 1990.

[15] C. H. Gierull. Statistical analysis of the eigenvector projection method for adaptive spatial filtering of interference. *IEE Proc. on Radar, Sonar Navigation*, 144:57–63, 1997.

[16] P. Gill, W. Murray, and M. Wright. *Practical Optimization*. Academic Press, London, UK, 1981.

[17] R. Gnanadesikan. *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York, second edition, 1997.

[18] J. Goldstein and I. Reed. Theory of partially adaptive radar. *IEEE Trans. AES*, 33(4):1309–1325, Oct. 1997.

[19] J. C. Harsanyi. *Detection and Classification of Subpixel Spectral Signatures in Hyperspectral Image Sequences*. PhD thesis, University of Maryland, 1993.

[20] J. C. Harsanyi and C. I. Chang. Detection of low probability subpixel targets in hyperspectral image sequences with unknown backgrounds. *IEEE Trans. Geoscience and Remote Sensing*, 32(4):779–785, July 1994.

[21] M. Healy. Multivariate normal plotting. *Appl. Stat.*, 17:157–161, 1968.

[22] R. R. Hocking. *Methods and Applications of Linear Models*. Wiley, New York, 1996.

[23] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey, 1998.

[24] S. M. Kay. *Fundamentals of Statistical Signal Processing*. Prentice Hall, New Jersey, 1998.

[25] E. J. Kelly. An adaptive detection algorithm. *IEEE Trans. on Aerospace and Electronic Systems*, 22(1):115–127, March 1986.

[26] E. J. Kelly. Adaptive detection in non-stationary interference, part III. Technical Report 761, MIT Lincoln Laboratory, Lexington, MA, 1987.

[27] E. J. Kelly and K. M. Forsythe. Adaptive detection and parameter estimation for multidimensional signal models. Technical Report 848, MIT Lincoln Laboratory, Lexington, MA, April 1989.

[28] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw. An algorithm taxonomy for hyperspectral unmixing. In *Algorithms for Multispectral and Hyperspectral Imagery*, volume 4049, pages 42–63, Orlando, FL, 2000. SPIE.

[29] D. L. Kessel and K. Fukunaga. A test for multivariate normality with unspecified parameters. Technical report, Purdue University School of Electrical Engineering, 1972.

[30] S. Kogon and D. Williams. Characteristic function based estimation of stable distribution parameters. In R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, *A Pratical Guide to Heavy Tails: Statistical Techniques and Applications*, pages 311–335. Birkhauser, Boston, 1998.

[31] S. Kraut and L. Scharf. The CFAR adaptive subspace detector is a scale-invariant GLRT. *IEEE Trans. Signal Processing*, 47:2538–2541, Sept. 1999.

[32] S. Kraut, L.L. Scharf, and L. T. McWhorter. Adaptive subspace detectors. *IEEE Transactions on Signal Processing*, 49(1):1–16, January 2001.

[33] D. Manolakis. Overview of algorithms for hyperspectral target detection: Theory and practice. In S. Shen and M. Descour, editors, *SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, Orlando, FL, April 2002. SPIE.

[34] D. Manolakis and G. Shaw. Directionally constrained or constrained energy minimization adaptive matched filter: Theory and practice. In M. Descour and S. Shen, editors, *Imaging Spectrometry*, San Diego, CA, August 2001. SPIE.

[35] D. Manolakis, G. Shaw, and N. Keshava. Comparative analysis of hyperspectral adaptive matched filter detectors. In S. S. Chen and M. R. Descour, editors, *Algorithms for Multispectral and Hyperspectral Imagery VI*, volume 4049, pages 2–17, Orlando, FL, April 2000. SPIE.

[36] D. Manolakis, C. Siracusa, D. Marden, and G. Shaw. Hyperspectral adaptive matched filter detectors: Practical performance comparison. In *Algorithms for Multispectral and Hyperspectral Imagery*, Orlando, FL, April 2001. SPIE.

[37] D. Manolakis, C. Siracusa, and G. Shaw. Adaptive matched subspace detectors for hyperspectral imaging applications. Salt Lake City, Utah, May 2001. IEEE International Conference on ASSP.

[38] D. Manolakis, C. Siracusa, and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Trans. on Geoscience and Remote Sensing*, 39(7):1392–1409, July 2001.

[39] D. G. Manolakis, V. K. Ingle, and S. M. Kogon. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. McGraw-Hill Companies, Inc., Boston, 2000.

[40] D. Marden and D. Manolakis. Algorithms for hyperspectral imaging data exploitation using non-gaussian elliptically contoured distributions. In S. S. Shen, editor, *Imaging Spectroscopy VIII*, Seattle, WA, 2002. SPIE.

[41] D. Marden and D. Manolakis. Constrained detectors for hyperspectral remote sensing applications: Theory versus practice. In S. S. Shen, editor, *Imaging Spectroscopy VIII*, Seattle, WA, 2002. SPIE.

[42] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.

[43] M. Rangaswamy, D. Weiner, and A. Ozturk. Non-gaussian random vector identification using spherically invariant random processes. *IEEE Transactions on Aerospace and Electronic Systems*, 29(1):111–124, January 1993.

[44] M. Rangaswamy, D. Weiner, and A. Ozturk. Computer generation of correlated non-gaussian radar clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):106–116, January 1995.

[45] I. S. Reed and X. Yu. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 38(10):1760–1770, October 1990.

[46] C. D. Richmond. Derived PDF of maximum likelihood signal estimator which employs an estimated noise covariance. *IEEE Trans. on Signal Processing*, 44(2):305–315, February 1996.

[47] C. D. Richmond. Performance of the adaptive sidelobe blanker detection algorithm in homogeneous environments. *IEEE Transactions on Signal Processing*, 48(5):1235–1247, May 2000.

[48] L. J. Rickard, R. Basedow, E. Zalewski, P. Silvergate, and M. Landers. HYDICE: An airborne system for hyperspectral imaging. *Proceedings of SPIE, Imaging Spectrometry of the Terrestrial Environment*, 1937, 1993.

[49] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg. A CFAR adaptive matched filter detector. *IEEE Trans. on Aerospace and Electronic Systems*, 28(1):208–218, January 1992.

[50] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York, 1994.

[51] L. L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Trans. on Signal Processing*, 42(8):2146–2157, August 1994.

[52] L. L. Scharf and L. T. McWhorter. Adaptive matched subspace detectors and adaptive coherence. In *Proc. 30th Asilomar Conf. On Signals and Systems*, pages 1114–117, 1996.

[53] A. Schaum and A. Stocker. Spectrally-selective target detection. *Proceedings of ISSSR*, 1997.

[54] R. A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, San Diego, 1997.

[55] R. Sitgreaves. On the distribution of two random matrices used in classification. *Annals of Math. Statistics*, 23(2):263–270, June 1952.

[56] M. O. Smith, P. E. Johnson, and J. B. Adams. Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis. *Journal of Geophysical Research*, 90:C797–C804, February 1985.

[57] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker. Anomaly detection from fyperspectral imagery. *Signal Processing Magazine*, 2002.

[58] D.W. Stein and S.G. Beaven. The fusion of quadratic detection statistics applied to hyperspectral imagery. *IRIA-IRIS proceedings of the 2000 Meeting of the MSS Specialty Group on Camouflage, Concealment, and Deception*, pages 271–280, March 2000.

[59] M. Steiner and K. Gerlach. Fast converging adaptive processor or a structured covariance matrix. *IEEE AES*, 36(4):1115–1125, Oct. 2000.

[60] A.D. Stocker and A. Schaum. Application of stochastic mixing models to hyperspectral detection problems. *SPIE Proceedings*, 3071, April 1997.

[61] B. Thai and G. Healey. Invariant subpixel target identification in hyperspectral imagery. In S. S. Shen and M. R. Descour, editors, *Algorithms for Multispectral and Hyperspectral Imagery IV*, volume 3717, pages 14–24, Orlando, Florida, 1999. SPIE.

[62] J. Scott Tyo, J. Robertson, J Wollenbecker, and R. C. Olsen. Statistics of target spectra in HSI scenes. *SPIE Imaging Spectrometry VI*, August 2000.

[63] X. Yu and I.S. Reed. Comparative performance analysis of adaptive multiband detectors. *IEEE Transactions on Signal Processing*, 41:2639–2656, August 1993.

# REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| 1. AGENCY USE ONLY (*Leave blank*) | 2. REPORT DATE<br>7 February 2002 | 3. REPORT TYPE AND DATES COVERED<br>Project Report |
|---|---|---|

**4. TITLE AND SUBTITLE**

Detection Algorithms for Hyperspectral Imaging Applications

**5. FUNDING NUMBERS**

C — F19628-00-C-0002

**6. AUTHOR(S)**

D. Manolakis

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Lincoln Laboratory, MIT
244 Wood Street
Lexington, MA 02420-9108

**8. PERFORMING ORGANIZATION REPORT NUMBER**

PR-HTAP-8

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

DUSD (S&T)
Rosslyn Plaza North, Suite 9030
1777 N. Kent St.
Rosslyn, VA 22209

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

ESC-TR-2001-044

**11. SUPPLEMENTARY NOTES**

None

AD #
A399744

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution is unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (*Maximum 200 words*)**

Detection and identification of military and civilian targets from airborne platforms using hyperspectral sensors is of great interest. Relative to multispectral sensing, hyperspectral sensing can increase the detectability of pixel and subpixel size targets by exploiting finer detail in the spectral signatures of targets and natural backgrounds. A multitude of adaptive detection algorithms for resolved and subpixel targets, with known or unknown spectral characterization, in a background with known or unknown statistics, theoretically justified or ad hoc, with low or high computational complexity, have appeared in the literature or have found their way into software packages and end-user systems. The purpose of this report is twofold. First, we present a unified mathematical treatment of most adaptive matched filter detectors suing common notation, and we state clearly the underlying theoretical assumptions. Whenever possible, we express existing ad hoc algorithms as computationally simpler versions of optimal methods. Second, we present a comparative performance analysis of the basic algorithms using theoretically obtained performance characteristics. We focus on algorithms characterized by theoretically desirable properties, practically desired features, or implementation simplicity. Sufficient detail is provided for others to verify and expand this evaluation and framework. A primary goal is to identify best-of-class algorithms for detailed performance evaluation. Finally, we provide a taxonomy of the key algorithms and introduce a preliminary experimental framework for evaluating their performance.

**14. SUBJECT TERMS**

**15. NUMBER OF PAGES**
90

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT<br>Unclassified | 18. SECURITY CLASSIFICATION OF THIS PAGE<br>Same as Report | 19. SECURITY CLASSIFICATION OF ABSTRACT<br>Same as Report | 20. LIMITATION OF ABSTRACT<br>Same as Report |
|---|---|---|---|