A Survey of Spectral Unmixing Algorithms

Nirmal Keshava

Spatial pixel sizes for multispectral and hyperspectral sensors are often large enough that numerous disparate substances can contribute to the spectrum measured from a single pixel. Consequently, the desire to extract from a spectrum the constituent materials in the mixture, as well as the proportions in which they appear, is important to numerous tactical scenarios in which subpixel detail is valuable. With this goal in mind, spectral unmixing algorithms have proliferated in a variety of disciplines that exploit hyperspectral data, often duplicating and renaming previous techniques. This article distills these approaches into a unique set and surveys their characteristics through hierarchical taxonomies that reveal the commonalities and differences between algorithms. A set of criteria organizes algorithms according to the philosophical assumptions they impose on the unmixing problem. Examples demonstrate the performance of key techniques.

Spectral imaging sensors often record scenes in which numerous disparate material substances contribute to the spectrum measured from a single pixel. Given such mixed pixels, we want to identify the individual constituent materials present in the mixture, as well as the proportions in which they appear. Spectral unmixing is the procedure by which the measured spectrum of a mixed pixel is decomposed into a collection of constituent spectra, or *endmembers*, and a set of corresponding fractions, or *abundances*, that indicate the proportion of each endmember present in the pixel. Endmembers normally correspond to familiar macroscopic objects in the scene, such as water, soil, metal, or any natural or man-made material. Unmixing provides a capability that is important in numerous tactical scenarios in which subpixel detail is valuable.

Multispectral imaging sensors such as Landsat provided the first opportunity to derive multichannel spectral information from large scenes on a pixel-by-pixel basis. The most significant products of multispectral data processing have been classification maps that assign a class label to each pixel in a scene. The relatively low numbers of spectral bands in multispectral sensors (usually a dozen or fewer) have proved sufficient to provide classification maps for large scenes with numerous applications to agriculture, forestry, oceanography, and environmental management and protection. As electro-optical remote sensing has evolved, hyperspectral sensors have been developed with hundreds of spectral bands with significantly improved spectral resolution. The ability of spectral unmixing to identify the constituent components of a pixel is a particularly important new application for these sensors.

Pixel-wise classification identifies the material class that a pixel spectrum most closely resembles, but it does not yield any further insight into the other substances—if any—that might also reside within the boundaries of the pixel. Mixed pixels contain a mixture of more than one distinct material substance, and they arise for one of two reasons. First, if the spatial resolution of an imaging sensor is low enough such that adjacent endmembers can jointly occupy a single pixel, then the resulting spectral measurement will be a composite of the individual endmembers.

This occurs in remote sensing platforms flying at a high altitude or performing wide-area surveillance, where low spatial resolution is common. Second, mixed pixels appear when distinct materials are combined into a homogeneous mixture (e.g., sand grains on a beach), which can occur regardless of the spatial resolution of the sensor. Both circumstances arise in military applications; the first occurs when a man-made target of interest is smaller than the size of a pixel, and the second occurs when we attempt to characterize the conditions of naturally occurring background (e.g., dry soil, moist soil, or swamp).

Broadly speaking, spectral unmixing is a special case of the generalized inverse problem that estimates system parameters by using one or more observations of a signal that has interacted with the system before arriving at the sensor. In the case of hyperspectral sensing in the reflective regime, the incident signal is electromagnetic radiation originating from the sun, which is measured by a sensor after the radiation has been reflected upward by natural and man-made materials on the surface of the earth.

The objective of this article is to systematically introduce the reader to the wide variety of techniques being employed for spectral unmixing. To accomplish this, we undertake two important tasks: (1) decompose the end-to-end unmixing process into three algorithmic stages, and (2) organize the algorithms belonging to each stage into hierarchical taxonomies according to a common set of criteria that reveal the fundamental assumptions behind each algorithm.

This article employs the taxonomies as a tool to graphically compare the features of algorithms in a side-by-side fashion. Being inherently hierarchical, the taxonomies demonstrate the commonalities and the differences in a way that facilitates comparisons. For completeness the taxonomies contain many algorithms; only a few entries, however, are necessary to provide the reader with insight on key approaches. A more comprehensive treatment of the technical detail for each algorithm is available in other references [1].

Two additional sections in this article discuss important related subjects. The first is a sidebar that demonstrates how geometric modeling of hyperspectral data has led to a new method for extracting endmembers from a scene, and the second is an appendix that discusses the properties of distance metrics and band selection.

A Taxonomy for Unmixing Algorithms

The various research communities that use hyperspectral data have developed numerous unmixing algorithms to solve particular problems. Earth scientists, for example, have correlated surface geophysical processes with measurements taken from space and, as a result, have approached the task of unmixing from the perspective of meticulous physical models that carefully capture the interactions of light with mixed matter. Quite often, however, in spite of their accuracy, these models are incapable of conveying the statistical variability inherent in remote sensing observations, and the unmixing results, while accurate for the situation described, lack the kind of robustness desired for general implementation. In contrast, engineers and statisticians often eschew the physics in favor of simpler, more tractable descriptions that utilize robust statistical models, which may achieve some measure of optimality. Unfortunately, statistical modeling often fails to reflect the high degree of physical detail that guarantees precision and physically plausible answers for individual pixels.

Unmixing algorithms have been developed by disparate communities of researchers, spanning a wide gamut of approaches. Yet they frequently utilize the same mathematical tools, but with different names and nomenclature. A major objective of developing a taxonomy has been to distill to a minimum set the universe of algorithms that have arisen from disparate communities. In order to perform this distillation, we developed a method for hierarchically organizing algorithms by the philosophies they employ for three important characteristics of processing. Each successive characteristic can be thought of as a sieve of increasing granularity, and the conjunction of the three criteria provides incrementally greater refinement in distinguishing algorithms. As will become apparent, the end-to-end process of unmixing is actually a concatenation of three distinct procedures, each with unique objectives. The same taxonomic approach has been applied to every stage of processing, thus revealing philosophies and methods common across the three algorithm types.

A second objective of a taxonomy is to offer a road map for developers of algorithms, so they can clearly see what has and has not been attempted. In many cases, particularly when performance comparisons are needed, a taxonomy can streamline the tedious process of comparative performance analysis by naturally segregating groups of algorithms requiring similar inputs or common computational modules. Moreover, scientists and engineers tasked with designing a specific application, such as unmixing, in a constrained operational environment will do so with limited information and computational resources. This fact in itself may automatically restrict the set of allowable algorithms from which the developer has to choose. The taxonomies organize information about algorithms in a way that facilitates this kind of analysis.

Stages of Unmixing

Algorithms for spectral unmixing use a variety of different mathematical techniques to estimate endmembers and abundances. Because hyperspectral scenes can possess extremely large volumes of data (e.g., 640 scan lines, 320 samples per line, 200 spectral bands), some unmixing algorithms first reduce the dimension of the data to minimize the corresponding computation. Not surprisingly, the familiar trade-off for representing data in a reduced dimension is a decrease in the accuracy of the intended application product. Nevertheless, we can decompose the complete end-to-end unmixing problem as a sequence of three consecutive stages: dimension reduction, endmember determination, and inversion. Figure 1 illustrates these three stages.

In the dimension-reduction stage we reduce the dimension of the data in the scene. This step is optional and is invoked only by some algorithms to reduce the computational load of subsequent processing. In the endmember-determination stage we estimate the set of distinct spectra (endmembers) that constitute the mixed pixels in the scene. Finally, in the inversion stage we generate abundance planes that allow us to estimate the fractional abundances for each mixed pixel from its spectrum and the endmember spectra.

Numerous approaches exist in the literature for completing the tasks at each stage. In the following sections, we present an algorithm taxonomy for each step that hierarchically organizes the algorithms, and in doing so, we employ an identical top-down structure for each stage. The layers in this structure serve to distinguish the algorithms by the philosophical assumptions they impose on the task they seek to accomplish.

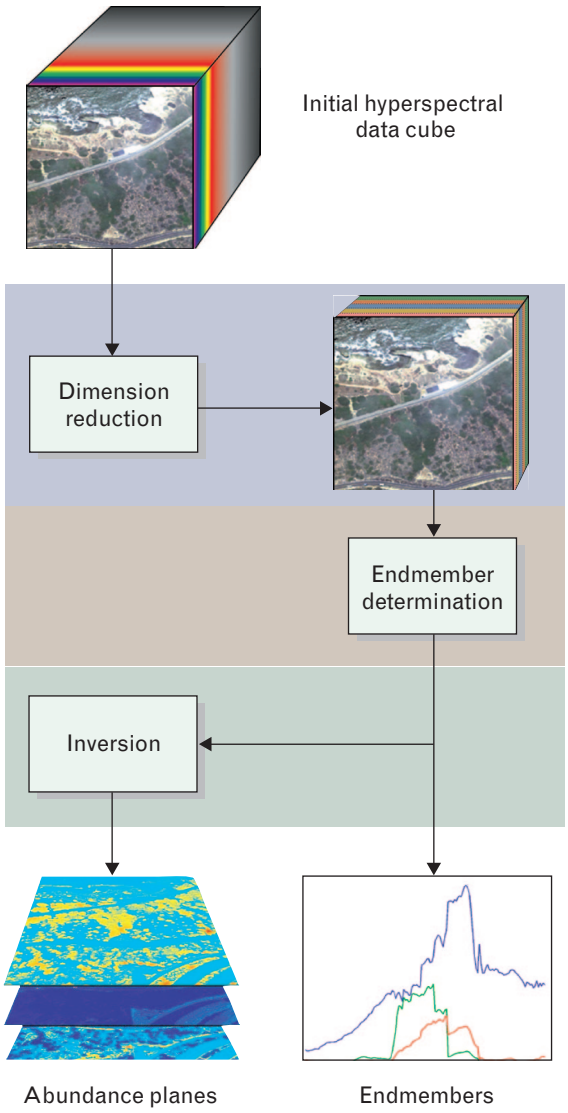

FIGURE 1. The end-to-end stages of spectral unmixing. Dimension reduction reduces the amount of data in order to decrease the computational load in subsequent processing steps. Endmember determination estimates the set of distinct spectra, or endmembers, that comprise the mixed pixels in the scene. The inversion stage produces abundance planes that provide estimates of the fractional abundances for the endmembers in each pixel.

Taxonomy Structure

In botany, starting from the time of the ancient Greeks, plant types were organized into descending categories, creating a top-down organization of known plant species. An important development in the classification of organisms occurred when Carolus Linnaeus (1707–1778) published a taxonomic system that worked in the opposite direction, beginning with individual species and grouping them according to characteristics that emphasized their similarities as well as their differences. This system provided by Linnaeus identified criteria for the classification of living things that serves today as the basis for the system of organization in modern botany.

In our effort to develop a taxonomy for unmixing algorithms, we have uncovered criteria for distinguishing algorithms by integrating both approaches. First, we reframe the unmixing problem as a special case of the generalized inverse problem [2] that has been studied for many years, and we pose basic questions regarding the inherent philosophical perspectives employed by each algorithm. In contrast to the ground-up viewpoint, this top-down assessment focuses on the fundamental assumptions that, knowingly or unknowingly, drive algorithms. Viewed hierarchically, the downward flow of the taxonomy from a single broad class of techniques sharing a common objective occurs along pathways that distinguish the variety of methods with increasing granularity. As an example, algorithms that assume a Gaussian model for the randomness of the data can be distinguished from those which enforce no probabilistic model. Among Gaussian techniques, a more detailed distinction can be drawn between maximum likelihood and maximum *a posteriori* formulations.

The top-down examination of the unmixing problem yields a set of three criteria that categorize unmixing algorithms. In order, these criteria are (1) interpretation of data, which indicates how an algorithm interprets mixed-pixel spectra; (2) description of randomness, which indicates how an algorithm incorporates the randomness of the data; and (3) optimization criterion, which indicates what objective function is being optimized by the algorithm.

An algorithm can interpret the data it processes in one of two ways. If an algorithm processes a mixed pixel by using statistical measures (e.g., means or covariances), then the algorithm is statistical. Implicitly, statistical algorithms quantitatively introduce the aggregate behavior of a larger population of data into the processing of an individual pixel, and they do so having no knowledge of the probabilistic nature of the data. In dimension reduction, for example, linear transforms that achieve a basis transformation are often derived from the covariance of the data without

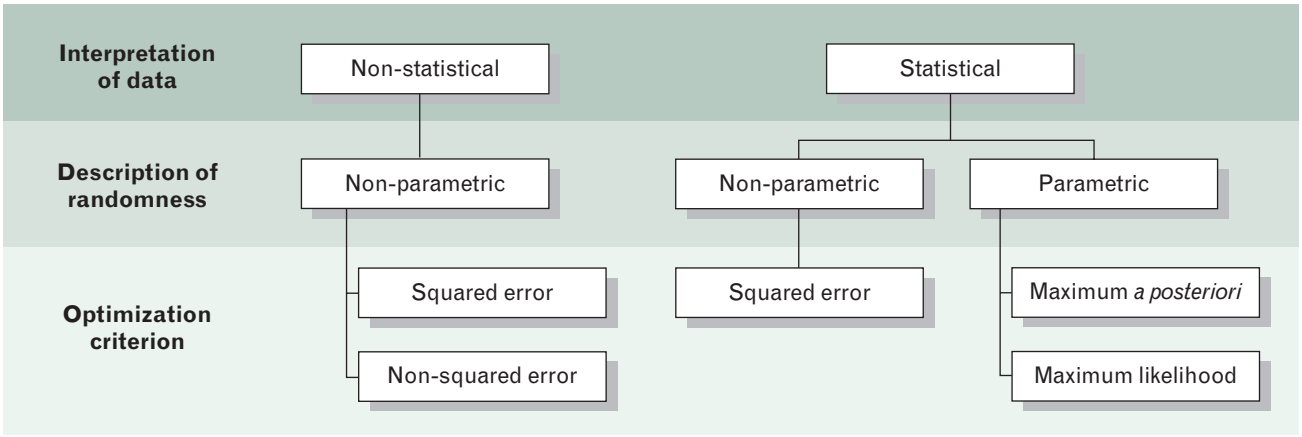

FIGURE 2. Top-down conceptual organizing structure for a taxonomy of unmixing algorithms. Three criteria categorize these algorithms: (1) interpretation of data indicates how an algorithm interprets mixed-pixel spectra, (2) description of randomness indicates how an algorithm incorporates data randomness, and (3) optimization criterion indicates the objective function that is being optimized.

knowledge of the probability distribution that produces the data (e.g., Gaussian or Cauchy). In essence, the dimension of a pixel is reduced by using knowledge gained from a larger population of data. Algorithms that perform their task without statistical measures are non-statistical. This distinction becomes especially important in target detection, among other places, where statistical characterizations of nontarget behavior (background or clutter) can complicate the detection of low-probability targets.

Likewise, the data interpretation upon which an algorithm is based directly reflects how it addresses the fundamental randomness in the data. To represent the aggregate behavior of data, algorithms can use statistical measures, which may also be parameters in analytical expressions that represent all or part of a probability density function. Techniques vested with the assumption that the received data originate from a parameterized probability density function are considered parametric. Examples of parametric methods arise whenever algorithms incorporate Gaussian probability density functions in their derivation. It is important to note, however, that a statistical algorithm is not always parametric, although the converse must be true. Algorithms that do not impose the structure of a particular density function are non-parametric.

Finally, algorithms are deemed optimal if they optimize an objective function. The choice of the objective function is key, and only in certain cases is it dictated naturally by the previous two algorithm classification criteria. Parametric algorithms optimize some combination of densities and fall in the category of either maximum likelihood or maximum *a posteriori* solutions. Non-parametric algorithms can utilize one of a multitude of cost functions, but clearly the most prevalent optimization criterion is minimization of squared error. Not surprisingly, the category of non-squared error encompasses a multitude of optimization metrics that do not minimize squared error.

Figure 2 illustrates the organization of these classification criteria. Independently, each of these criteria captures a fundamental aspect of how the algorithm relates to the data it processes, and because the criteria are not specific to any particular algorithm, the hierarchical partitioning they induce map from one type of algorithm to another. As we will see, complete spectral unmixing is the concatenation of three consecutive and disparate operations, and to consolidate their assessment we applied the same criteria to each taxonomy to consistently organize the algorithms at every stage.

While the three classification criteria in Figure 2 discriminate algorithms for a wide variety of tasks by using a common top-down approach, further differentiation within each class of algorithm is achieved by a set of task-specific features. In contrast to the criteria from Figure 2, which apply to every stage of unmixing, this bottom-up approach identifies boundaries for delineating algorithms on the basis of the operating characteristics that are unique to each stage. For example, the inclusion or omission of noise in signal models provides one axis of organization, whereas the added distinction between Gaussian and non-Gaussian noise yields further granularity. The features that distinguish algorithms are organized into three categories: output characteristics, input characteristics, and noise modeling. The details of these algorithm-specific properties become evident in the appropriate sections. For a particular algorithm, Figure 3 depicts how these features of an algorithm are related to the criteria in Figure 2.

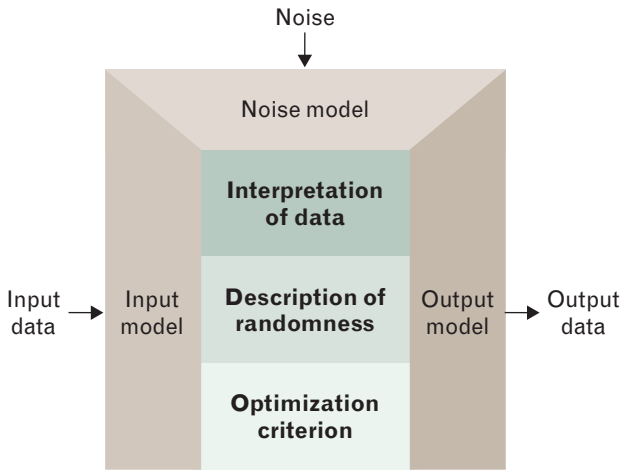

FIGURE 3. Model for algorithm features. Specific algorithm features add further refinement to the classification of algorithms. Models for the inputs, outputs, and noise provide practical processing requirements for how an algorithm instantiates the top-down criteria in Figure 2.

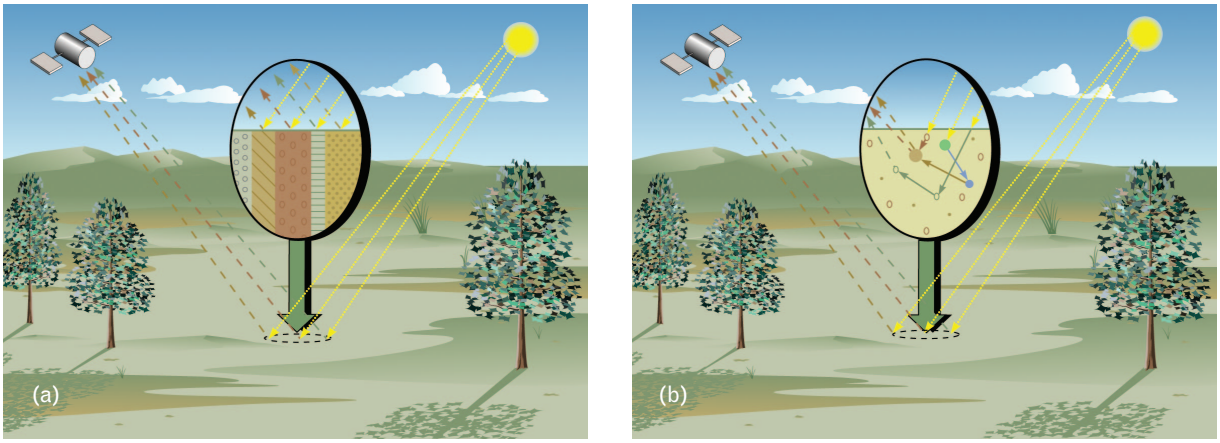

FIGURE 4. Two mixing models. (a) The linear mixing model assumes a well-defined proportional checkerboard mixture of materials, with a single reflection of the illuminating solar radiation. (b) Nonlinear mixing models assume a randomly distributed, homogeneous mixture of materials, with multiple reflections of the illuminating radiation. These models represent the underlying physics at the foundation of hyperspectral phenomenology. Unmixing algorithms use these models to recover endmembers and associated abundances from the mixed-pixel spectrum.

Mixing Models

Any approach for effectively unmixing hyperspectral data must begin with a model describing how constituent material substances in a pixel combine to yield the composite spectrum measured at the sensor. Mixing models attempt to represent the underlying physics that are the foundation of hyperspectral phenomenology, and unmixing algorithms use these models to perform the inverse operation, attempting to recover the endmembers and their associated fractional abundances from the mixed-pixel spectrum. Figure 4 illustrates the two categories of mixing models—the linear mixing model and the nonlinear mixing models.

Linear Mixing Model

The dynamics behind the mixing of two or more substances depends largely upon the kind of mixture within a pixel that scatters the incident solar radiation. Figure 4(a) illustrates the reflecting surface as a checkerboard mixture of endmembers, and the incident radiation bounces only once upon its surface. By this model, if the total surface area is divided proportionally according to the fractional abundances of the constituent substances, then the reflected radiation conveys with the same proportions the characteristics of the associated materials. In this sense, a linear relationship exists between the fractional abundance of the substances comprising the area being imaged and the spectrum of the reflected radiation.

If we have K spectral bands, and we denote the ith endmember spectrum as $\mathbf{s}_i$ and the abundance of the ith endmember as a_i, the observed spectrum $\mathbf{x}$ for any pixel in the scene can be expressed as

$$\mathbf{x} = a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + \cdots + a_M\mathbf{s}_M + \mathbf{w} = \sum_{i=1}^{M} a_i\mathbf{s}_i + \mathbf{w} = \mathbf{S}\mathbf{a} + \mathbf{w},$$

where M is the number of endmembers, $\mathbf{S}$ is the matrix of endmembers, and $\mathbf{w}$ is an error term accounting for additive noise (including sensor noise, endmember variability, and other model inadequacies). This model for pixel synthesis is the linear mixing model (LMM). Block matrix notation is used as an extension for the case in which multiple pixel spectra $\mathbf{x}(n)$, for $n = 1, \ldots, N$, where N is the number of pixels, are organized as columns within a matrix $\mathbf{X}$ having a corresponding abundance matrix $\mathbf{A}$ and noise matrix $\mathbf{W}$, such that $\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{W}$. To be physically realizable, the abundance coefficients should be nonnegative and should sum to one. Satisfying these physically driven abundance constraints, which is discussed in more detail in later sections, is important for various applications.

FIGURE 5. Image acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor during the Alpine Radiance I data collection in 1997. The variety of natural materials and man-made artifacts makes this scene a good candidate for testing unmixing algorithms.

Nonlinear Mixing

Figure 4(b) depicts a more complicated scenario. The arrangement of the constituent material substances is not as orderly as in Figure 4(a) because the substances comprising the medium are not organized proportionally on the surface. This intimate mixture of materials results when each component is randomly distributed in a homogeneous way. As a result, the incident radiation can experience reflections with multiple substances, and the aggregate spectrum of reflected radiation may no longer uphold the linear proportions (either in mass fraction or in volume) of the constituent substance spectrum. Because the LMM is inappropriate to describe this interaction, this scenario, which has many variations, is referred to as nonlinear mixing.

Data

To demonstrate the performance of different unmixing algorithms, we utilize the scene in Figure 5. This image was acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor during the Alpine Radiance I data collection on 22 September 1997. The scene contains several naturally occurring objects (e.g., trees, brush, grass, soil) as well as man-made artifacts (e.g., roads). The image consists of 400 lines of data, 320 samples per line, and 210 spectral bands collected between 400 nm and 2500 nm, with the width of each band varying between 3 nm and 11 nm. The spatial resolution of each pixel is on the order of one meter.

Dimension-Reduction Taxonomy

In this section, we examine the set of algorithms that reduce the dimension of hyperspectral data. We intentionally differentiate between dimension-reduction algorithms and the more widely recognized set of related data-compression algorithms in order to highlight an important distinction. Dimension-reduction algorithms do not reduce the dimension of data with the goal of reconstructing an approximation to the original signal. Instead, the goal of dimension reduction is to arrive at a minimal representation of the signal in a lower-dimensional space that sufficiently retains the requisite information for successful unmixing in the lower dimension. Ideally, dimension-reduction algorithms are designed with consideration of the performance of unmixing procedures performed in the lower dimension.

Figure 6 shows the taxonomy of dimension-reduction algorithms; the algorithms are organized by the top-down criteria appearing in Figure 2. In addition, specific features concerning the outputs, inputs, and noise models used by these algorithms are included according to the model in Figure 3 to specifically distinguish the properties of dimension-reduction algorithms. With regard to outputs, do the algorithms yield axes for dimension reduction that are orthogonal? With regard to inputs, do the algorithms process all of the data for dimension reduction, or just a subset? With regard to noise, do the algorithms incorporate any knowledge of noise? For example, by its position in the taxonomy, principal-component analysis is a statistical, non-parametric algorithm that optimizes a squared-error criterion. While it does not incorporate noise in its signal model, it uses all the data in the scene and yields orthogonal axes for its coordinate transformation.

Interpretation of data		Non-statistical	Statistical	
Description of randomness		Non-parametric		
Optimization criterion		Squared error		Non–squared error Signal-to-noise-ratio
Specifics	**Outputs**	Orthogonal axes		Non-orthogonal axes
	Inputs	Utilizes reduced data set	Utilizes all data in scene	
	Noise	No noise in signal model		Additive observation noise
Algorithm		Optical real-time adaptive spectral identification system (ORASIS)	Principal-component analysis	Maximum noise fraction Noise-adjusted principal components

FIGURE 6. Taxonomy for dimension-reduction algorithms, organized by the same three classification criteria appearing in Figure 2—interpretation of data, description of randomness, and optimization criterion. Specific features concerning the outputs, inputs, and noise models used by these algorithms are also included. For example, by its position in the taxonomy, principal-component analysis is a statistical, non-parametric algorithm that optimizes a squared-error criterion. While it does not incorporate noise in its signal model, it uses all the data in the scene and yields orthogonal axes for its coordinate transformation.

Dimension-Reduction Algorithm

Because the three algorithms in the dimension-reduction taxonomy in Figure 6 do not presume any probability density function for the data, they are all non-parametric. The category of statistical algorithms derives their transformations from statistical information about the data and can be further differentiated by the optimization criterion that is utilized. Principal-component analysis (PCA) [3], a technique based on squared error, identifies orthogonal axes for dimension reduction by performing an eigendecomposition of a covariance estimate of the data,

$$\begin{aligned}\mathbf{\Gamma}_x &= E\left[\left(\mathbf{x}-\boldsymbol{\mu}_x\right)\left(\mathbf{x}-\boldsymbol{\mu}_x\right)^T\right] \\ &= \frac{1}{N}\sum_{n=1}^{N}\left[\mathbf{x}(n)-\boldsymbol{\mu}_x\right]\left[\mathbf{x}(n)-\boldsymbol{\mu}_x\right]^T ,\end{aligned}$$

where $\boldsymbol{\mu}_x$ is the mean vector of the pixel set. The resulting eigendecomposition can be expressed as $\mathbf{\Gamma}_x = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$, where U is a unitary matrix of eigenvectors, and $\mathbf{\Sigma}$ is a diagonal matrix of eigenvalues.

The magnitude of an eigenvalue indicates the energy residing in the data along the component of the data parallel to the associated eigenvector. The larger eigenvalues identify basis components whose average contribution to $\mathbf{x} - \boldsymbol{\mu}_x$ in the squared-error sense is greater than those with smaller eigenvalues. Hence the effective dimensionality of the data can be estimated by counting the number of significantly non-

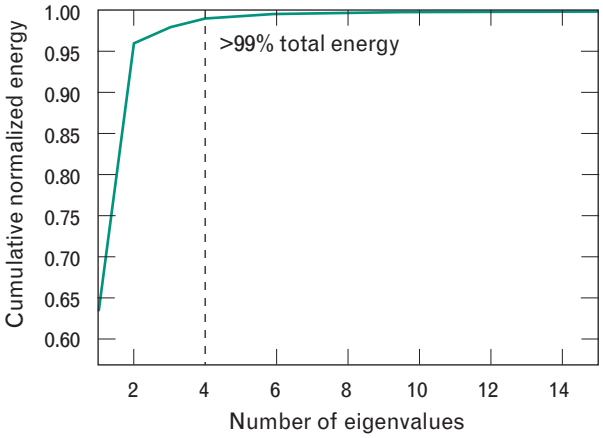

FIGURE 7. Cumulative normalized energy in eigenvalues for the HYDICE sensor scene in Figure 5. Most of the energy of the scene is in the first four eigenvalues. Thus a linear transformation of the scene using the first four eigenvectors transforms the data to a lower dimension, yet retains nearly all the energy of the original higher-dimensional data.

zero eigenvalues. If the data are transformed by premultiplication with $\mathbf{U}^T$, then the resulting linear transformation moves $\mathbf{x}$ to a new system of decorrelated variables oriented along the eigenvectors in $\mathbf{U}$, which can be truncated to retain only those having significant eigenvalues. The result is a lower-dimensional multivariate random vector that conveys most of the energy in the original, higher-dimensional system. Figure 7 depicts the distribution of energy in the eigenvalues for the scene in Figure 5. More than 99% of the total energy is retained by the first four eigenvalues.

Another statistical technique that optimizes signal-to-noise ratio (SNR), is maximum noise fraction (MNF) [4]. This approach, based on the LMM, requires estimating the covariance $\hat{\mathbf{\Gamma}}_w$ for the additive noise in addition to the covariance $\hat{\mathbf{\Gamma}}_x$ of the data. Given a vector $\mathbf{v}$, the ratio of noise energy in $\mathbf{v}$ to the received signal energy is given by

$$\frac{\mathbf{v}^T \hat{\mathbf{\Gamma}}_w \mathbf{v}}{\mathbf{v}^T \hat{\mathbf{\Gamma}}_x \mathbf{v}}.$$

The axes that optimize this ratio are the left-hand eigenvectors of $\hat{\mathbf{\Gamma}}_w \hat{\mathbf{\Gamma}}_x^{-1}$ and, unlike the axes for PCA, are not necessarily orthogonal. They do, however, identify and order the components of the received signal possessing the maximum SNR. The noise-adjusted principal components (NAPC) transform formulates the problem differently [5], but achieves the mathematically equivalent answer as MNF. As in PCA, the ordering of components can estimate one type of effective signal dimensionality, and the set of random variables obtained after the MNF transform can be truncated to retain only those components possessing a minimum SNR.

A non-statistical and non-parametric technique that optimizes squared error for dimension reduction is the optical real-time adaptive spectral identification system (ORASIS) [6], which is a series of hyperspectral processing modules designed to process data from airborne and spaceborne hyperspectral platforms. The dimension reduction is achieved by identifying a subset of representative, or exemplar, pixels that convey the variability in a scene. When a new pixel is collected from the scene by the sensor, its spectrum is compared to each exemplar pixel by using an angle metric (the first pixel in a scene automatically becomes the first exemplar). The appendix entitled "Distance Metrics, Band Selection, and Dimension Reduction" shows how a quantity such as angle is a common measure of spectral difference. If the new pixel is sufficiently different from each of the existing exemplars, it is added to the exemplar set. If it is not sufficiently different, the exemplar set remains unchanged. An orthogonal basis is periodically created from the current set of exemplars by using a modified Gram-Schmidt process, which adds new dimensions until every exemplar can be approximated within a prescribed tolerance. In this reduced dimension, the exemplars are submitted to another module to determine endmembers geometrically. The ORASIS flowchart in Figure 8 illustrates this dimension-reduction procedure.

The performance of the dimension-reduction algorithm is linked to the value of the user-chosen prescreen error-angle threshold, which restricts the admittance of new exemplars. Smaller values of the prescreen threshold admit more exemplars, and hence this parameter may be tuned to admit different levels of scene variability into the exemplar set. The prescreen threshold, however, also defines the required dimension of the projected space to preserve a minimum fidelity in the exemplars. This dimension is critical to the subsequent module that determines endmembers.

Endmember-Determination Taxonomy

In this section we address endmember determination, which yields the first primary output of unmixing. The objective of endmember-determination algorithms is to estimate the constituent spectra that occupy the columns of $\mathbf{S}$ in the LMM. From a strictly mathematical viewpoint, the determination of $\mathbf{S}$ is comparable to the estimation of a nonorthogonal subset of basis vectors. More physical interpretations, however, begin with the stipulation that the vectors must have non-negative entries in order to be physically realizable. Furthermore, endmembers should retain physical characteristics of the constituent substance such as absorption bands and spectral intervals of high and low reflectance. Consequently, identify-

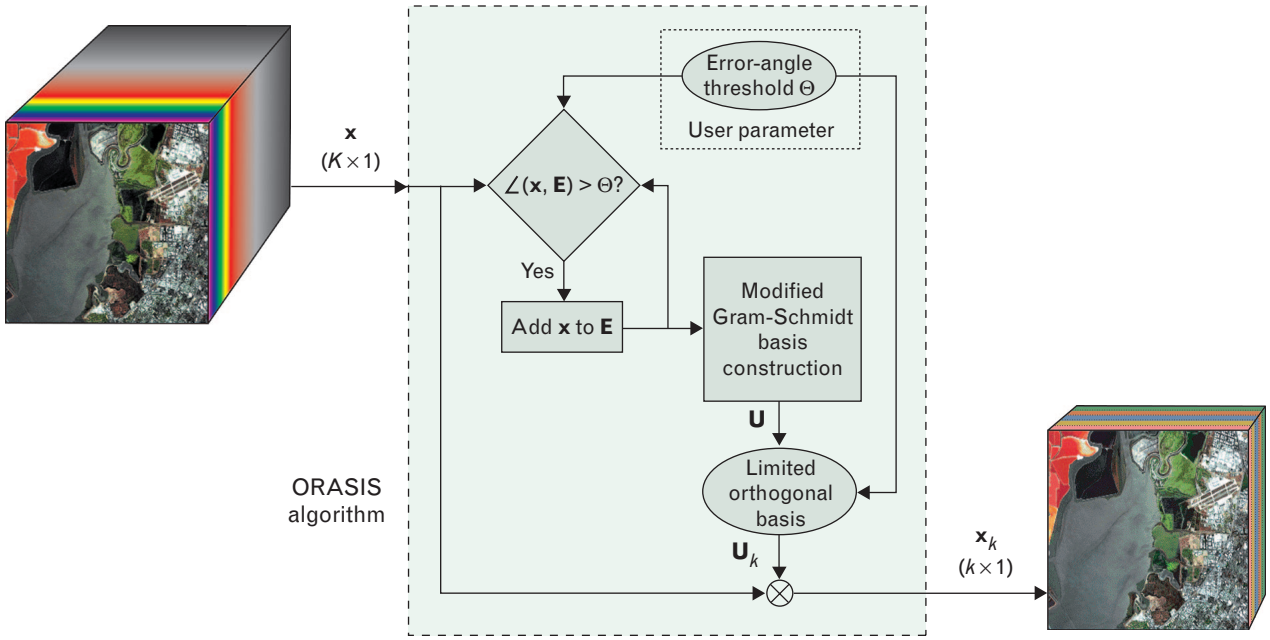

FIGURE 8. Flowchart for the optical real-time adaptive spectral identification system (ORASIS) dimension reduction technique, which transforms a data cube having M spectral channels into a data cube with m dimensions, where $k < K$. ORASIS identifies a set **E** of exemplar pixels that convey the spectral variability in a scene. Scene pixels **x** are individually compared to the current exemplars and added to **E** if any comparison exceeds a user-specified error-angle threshold Θ. The exemplars are periodically orthogonalized into a unitary basis **U** by a modified Gram-Schmidt procedure. A limited subset of the columns in **U**, called $\mathbf{U}_k$, is retained as a reduced-dimension orthogonal basis for the scene, which produces a reduced-dimension representation of **x**, called $\mathbf{x}_k$.

ing endmembers that satisfy both physical and mathematical imperatives is a considerable challenge, making autonomous endmember determination the hardest part of the unmixing problem.

Of the three stages that comprise unmixing, endmember determination is the most closely aligned with the material identification capabilities of unmixing. An accurate assessment of subpixel composition begins with a reliable estimate of what pure substances comprise mixed pixels in the scene. This fact confirms that endmember-determination techniques must not only contend with extracting spectra that are physically meaningful and—we hope—recognizable, but they must also perform in environments having limited and imperfect information. Non-statistical algorithms essentially assume the endmembers are deterministic quantities, whereas statistical approaches view endmembers as either deterministic, with an associated degree of uncertainty, or as fully stochastic, with random variables having probability density functions.

Figure 9 shows a taxonomy of endmember-determination algorithms. As discussed earlier, the algorithms are organized by the three criteria given in Figure 2. In addition, specific features concerning the outputs, inputs, and noise models used by these algorithms are included according to the model in Figure 3 to specifically distinguish the properties of endmember-determination algorithms. With regard to outputs, do the algorithms yield deterministic or stochastic endmembers? With regard to inputs, do the algorithms require all pixel spectra and abundances or just a subset? With regard to noise, do the algorithms incorporate any knowledge of noise?

Endmember-Determination Algorithms

Statistical methods of endmember determination identify endmembers by optimizing objective functions derived from the statistics of the data. Furthermore, the non-parametric algorithms attempt this without optimizing a parametric model, but instead by minimizing an objective function using the statis-

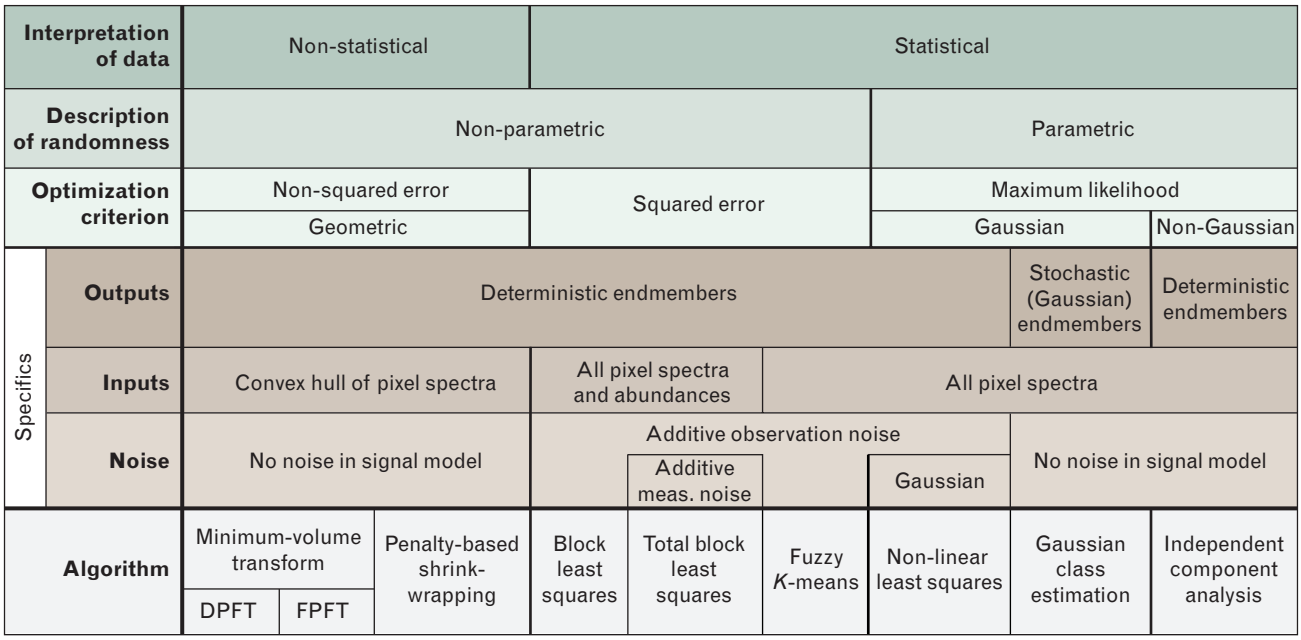

FIGURE 9. Taxonomy of endmember determination algorithms, which are organized by the same three classification criteria appearing in Figure 2—interpretation of data, description of randomness, and optimization criterion. Specific features concerning the outputs, inputs, and noise models used by these algorithms are also included. For example, by its position in the taxonomy, Gaussian class estimation is a statistical, parametric algorithm that optimizes a maximum-likelihood criterion for Gaussian signals. While it does not incorporate noise in its signal model, it uses all the data in the scene, and unlike other endmember determination algorithms, it models endmembers as stochastic.

tics derived from the data. A common objective function for this goal is squared error. Typical examples are clustering algorithms, such as the fuzzy *K*-means partitions algorithm [7], which is a variation on the well-known *K*-means iterative clustering algorithm [8] that progressively minimizes an objective function $J_q(\mathbf{A}, \mathbf{S})$ to simultaneously arrive at optimal estimates of deterministic endmembers and abundances for N pixels. This function is expressed as

$$J_q(\mathbf{A},\mathbf{S}) = \sum_{i=1}^{M}\sum_{n=1}^{N}\left(\mathbf{A}_{in}\right)^q\left(d_{in}\right)^2 .$$

$\mathbf{A}_{in}$ is an estimate of the ith abundance for the nth pixel. The factor d_{in} is the squared error between the nth pixel and the ith centroid, or estimated endmember $\hat{\mathbf{s}}_i$, and is given by

$$\left(d_{in}\right)^2 = \left[\mathbf{x}(n) - \hat{\mathbf{s}}_i\right]^T \mathbf{W}\left[\mathbf{x}(n) - \hat{\mathbf{s}}_i\right].$$

The weighting matrix $\mathbf{W}$ applies different weights to each abundance, and the inverse class covariance may be inserted here. The minimization of $J_q(\mathbf{A}, \mathbf{S})$ is constrained by physical restrictions on the values of abundances, and these requirements are inserted into the estimator for $\mathbf{A}$. Moreover, the shapes of the regions associated with each class vary with q. Suggested values fall in the range $1.5 < q < 3.0$. The optimization of $J_q(\mathbf{A}, \mathbf{S})$ is accomplished iteratively until a minimum value is achieved by using final estimates of $\hat{\mathbf{S}}$ for the endmembers and $\hat{\mathbf{A}}$ for the abundances.

Endmember-determination algorithms that utilize statistical information in conjunction with parametric density functions find estimates by optimizing either a forward or posterior density function. Several algorithms have been formulated to provide maximum-likelihood solutions for endmembers on the basis of different formulations of the LMM as a forward density $p_{\mathbf{x}|\mathbf{S}}(\mathbf{x}|\mathbf{S})$ that is optimized to yield $\hat{\mathbf{S}}$. Each approach differs in technique, and ultimately in their answers, because of the degree of *a priori* knowledge as well as the assumptions they impose on the problem.

The technique of nonlinear least squares [9] mod-

GEOMETRIC ENDMEMBER DETERMINATION

THE SECTION ON dimension reduction demonstrates how principal-component analysis (PCA) can identify the axes that contain the most energy in a scene. The next step in spectral unmixing is to determine the collection of constituent material spectra, or endmembers, that correspond to these principal-component axes. This sidebar gives one technique for determining the endmembers that takes advantage of the results in dimension reduction shown earlier.

Geometric endmember determination techniques exploit the strong parallelism between the linear mixing model (LMM) and the geometric orientation of hyperspectral data in multidimensional spaces. The principal assumption is that pure substances, or endmembers, that appear in mixed pixels must necessarily reside at the extremities of the volume occupied by the pixels in a scene. Thus endmembers can be extracted from the vertices of a simplex that encloses the data in a scene, where a simplex is a convex volume having a number of vertices that is one greater than the dimension of the enclosed data.

The number of endmembers in a scene, along with the endmembers themselves, is unknown, but we can use the PCA results generated earlier to conclude that the number of significant eigenvalues is an approximate upper bound on the number of endmembers in a scene. Thus, with the results in Figure 7 as an example, this assumption permits the multidimensional data to be linearly transformed to a four-dimensional space by using the first four eigenvectors of the scene covariance. In this reduced space, the data are shrinkwrapped by a simplex, and endmembers are estimated from the five vertices of the final simplex enclosing the data.

Because the shrinkwrapping procedure requires the pixels residing only on the outer edge of the volume, the convex hull of the pixels in the reduced four-dimensional space is calculated. Afterward, the minimum-volume transform is used to iteratively minimize the volume of the simplex, starting from an initial structure, by maneuvering each face until no further reduction in volume can be achieved. The five endmember spectra are recovered from the vertices of the final simplex by inverting the linear transformation.

The result of this procedure for the scene appears on the left in Figure A. Shown on the right in Figure A are sample spectra for common natural materials in the same scene. Endmember 3 has a distinct resemblance to the spectrum for trees, whereas endmember 1 shows a resemblance to the spectrum for bare soil (i.e., road). Endmember 4 also shares some spectral characteristics with the spectrum for grass. Endmember 2 is clearly a noise endmember that arises from the need to enclose aberrant pixels as well as shade pixels,

els the additive noise **w** in the LMM as Gaussian, and assumes both **S** and **a** are deterministic unknowns for a single mixed pixel **x**. Moreover, this technique integrates *a priori* knowledge into least-squares formulations for unmixing by initializing an iterative estimator for **S** and **a** with guesses for what endmembers and abundance values should be. Associated covariances are set to values that quantify the confidence of the initial guesses. As a consequence of estimating **S** and **a** concurrently, the forward density can be alternately expressed as $p_{\mathbf{x}|\mathbf{S},\mathbf{a}}(\mathbf{x}\,|\,\mathbf{S},\mathbf{a})$. By virtue of Gaussianity, the likelihood function is quadratic in form, but the matrix product **Sa** recasts the simpler traditional maximum-likelihood optimization into a nonlinear least-squares problem [3].

The Gaussian class estimation method [10] also assumes Gaussianity, but does so without the presumption of additive noise. Instead, the method presumes

which typically have very low reflectances. Endmember 5 corresponds clearly to another distinct material present in the scene.

While three endmembers show a resemblance in shape to three materials in the scene, their absolute reflectance values are significantly different. In fact, some of the geometrically determined endmembers have reflectance values that are greater than one and less than zero, which is a physical impossibility! There are several reasons why these values occur. First, endmember reflectance values exceed one because the simplex encloses the data, which means the position of a vertex near a cluster of pixels of the same type must be positioned to enclose every pixel. If some clusters have high values and also exhibit significant variability, the vertex may have to be positioned well beyond a reflectance value of one to provide complete enclosure. The magnitude of the spectrum might be well beyond the average value of the material, but the general shape will be maintained. A similar argument can be made for endmember values that are less than zero.

In contrast to statistical techniques for endmember detection, geometric techniques are capable of revealing low-probability targets. A drawback, however, is that aberrant pixels that arise due to sensor artifacts may also be identified as endmembers. As a consequence, geometric techniques are especially suited to identifying low-probability targets, but they perform best with data that are free of artifacts that can create spurious results.

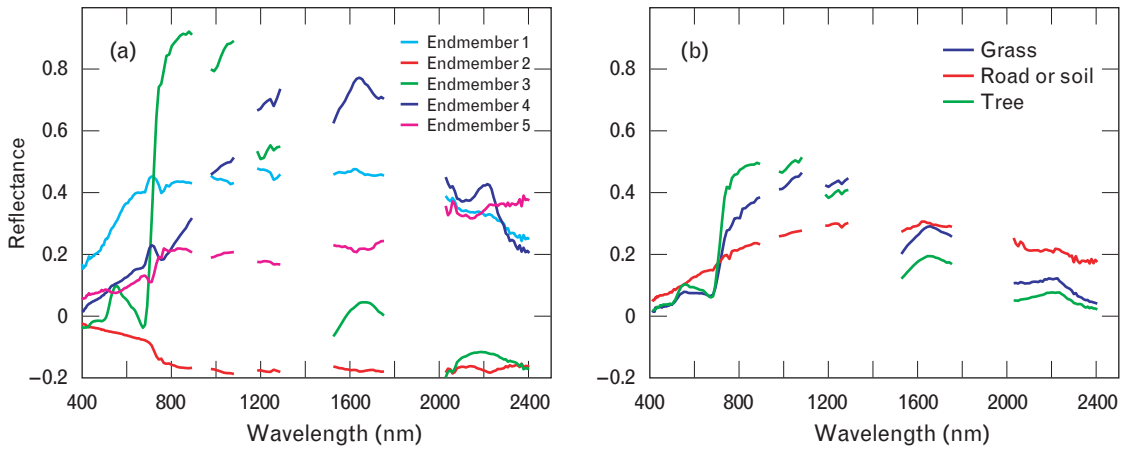

FIGURE A. (a) Five endmembers extracted from four-dimensional reduced data by using geometric endmember determination techniques; (b) sample spectra extracted from different materials and objects in the scene.

that the endmembers are Gaussian classes, and this technique is a deliberate attempt to fuse the geometric interpretation of linear mixing with Gaussian mixture modeling and maximum-likelihood estimation techniques. The stochastic mixing model (SMM) introduces the concept of hard endmember classes as the stochastic extension of deterministic endmembers, and it assumes that all data in a scene are Gaussian and arise from a linear combination of at least one hard class. Geometrically, each stochastic endmember describes one cluster of pixels on the perimeter of the data cloud that is parameterized by a mean, covariance, and prior probability $N(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \pi)$ so that every mixed pixel belongs to a combination of hard classes. If each combination of hard classes is a mixed class with its own Gaussian statistics (i.e., mean, covariance, and prior probability), then optimal unmixing determines the combination of classes having the

Interpretation of data		Non-statistical							Statistical			
Description of randomness		Non-parametric										
Optimization criterion		Non-squared error				Squared error						
Specifics	**Outputs**	Continuous-valued abundances										
	Full additivity	No	Yes			No	Yes		No			Yes
	Non-negativity	No	Yes			No		Yes		No		Yes
	Purity	No	Yes			No						Yes
	Inputs	Deterministic endmembers								Endmembers not required		
	Noise	Additive obs. noise	No noise in signal model						Additive observation noise			
Algorithm		Filter vector algorithm	I-divergence	Confidence estimates	ULS	ULS with full additivity	NNLS	Projection on convex sets	MVUE	Ground-truth-based estimators		Fuzzy *K*-means
										Classical	Inverse	

FIGURE 10. Taxonomy of inversion algorithms (continues on facing page), which are organized by the three classification criteria appearing in Figure 2—interpretation of data, description of randomness, and optimization criterion. Specific features concerning the outputs, inputs, and noise models used by these algorithms are also included. For example, by its position in the taxonomy, unconstrained least squares (ULS) with full additivity is a non-statistical, non-parametric algorithm that optimizes a squared-error criterion. It does not incorporate noise in its signal model, and it does assume endmembers are deterministic. Its estimates for abundances are continuous-valued and uphold the full additivity condition, but do not enforce the purity and non-negativity conditions.

greatest likelihood of realizing that pixel. After an initial partition of the data, the reclassified pixels are used to generate new class statistics, and the procedure repeats until the SMM class parameters cease to change.

An area of active research for endmember determination models the volume in a high-dimensional space occupied by hyperspectral data collected from a scene. Geometric approaches to endmember determination are non-parametric and non-statistical, and instead exploit the strong parallelism between the LMM and the theory of convex sets [11, 12]. These approaches rely on the assumption that pixel spectra from a scene reside within a high-dimensional volume, and by virtue of the LMM, endmembers must reside at the extremities of this volume in order for mixed pixels to arise. The objective then is to estimate the position of the vertices that demarcate an enclosed surface having minimum volume, while still enclosing every pixel.

Typically, the determination consists of two steps. The first involves data reduction, which like dimension reduction essentially minimizes computation by discarding information (a dimension-reduction step, such as PCA, commonly precedes data reduction). In this case, however, data reduction discards pixels, not bands. Only the perimeter of the volume occupied by the scene data is necessary to establish the location of the endmembers, and consequently the pixels residing within the convex hull [13] of the data are discarded.

The second step, known as shrinkwrapping, employs a class of procedures known as minimum-volume transforms (MVT) to iteratively fit a multifaceted simplex of minimum volume around the convex hull, where a simplex is a geometric surface in an

<table>
<tr><td colspan="8">Statistical</td></tr>
<tr><td colspan="8">Parametric</td></tr>
<tr><td colspan="5">Maximum likelihood</td><td colspan="3">Maximum a posteriori</td></tr>
<tr><td colspan="4">Continuous-valued abundances</td><td>Discretized abundances</td><td colspan="3">Continuous-valued abundances</td></tr>
<tr><td>No</td><td colspan="4">Yes</td><td>No</td><td colspan="2">Yes</td></tr>
<tr><td>No</td><td>Yes</td><td>No</td><td colspan="5">Yes</td></tr>
<tr><td>No</td><td>Yes</td><td>No</td><td colspan="5">Yes</td></tr>
<tr><td>Deterministic endmembers</td><td>Stochastic endmembers</td><td colspan="2">Endmembers not required</td><td colspan="4">Deterministic endmembers</td></tr>
<tr><td>Additive Gaussian observation noise</td><td>Additive Gaussian endmember noise</td><td>Additive Gaussian observation noise</td><td>No noise in signal model</td><td colspan="4">Additive Gaussian observation noise</td></tr>
<tr><td rowspan="2">Gaussian minimum-variance unbiased estimator</td><td rowspan="2">Weighted Mahalanobis distance</td><td rowspan="2">Nonlinear least squares</td><td rowspan="2">Gaussian class estimation</td><td colspan="3">Gaussian prior</td><td>Entropy prior</td></tr>
<tr><td>Fuzzy membership</td><td>Log-odds</td><td>Maximum entropy</td></tr>
</table>

n-dimensional hyperspace defined by exactly $n + 1$ vertices. The dark-point-fixed transform (DPFT) is a variant of the MVT that assumes knowledge of the dark point of the sensor. In contrast, the fixed-point-free transform (FPFT) assumes the dark point is unknown. Shrinkwrapping begins with a simplex of arbitrarily large size whose faces are consecutively adjusted to minimize its volume.

Geometric methods for endmember detection are decidedly non-statistical. They presume the presence of deterministic endmembers at the vertices of the simplex and, consequently, are sensitive to outliers or bad pixels arising from faulty elements on the sensor focal-plane array. However, geometric techniques can expose rare objects that would otherwise go unnoticed by statistical approaches. The sidebar entitled "Geometric Endmember Determination" describes the processing module that determines endmembers geometrically.

Inversion Taxonomy

In this section we address inversion, which yields the second primary output of unmixing. The objective of these algorithms is to determine the fractional presence of each endmember in the received pixel spectrum or, in terms of the LMM, find the vector **a** whose entries weight the columns of **S** to yield **x**. Any meaningful estimate of **a**, however, must comply with constraints that make it physically realizable. In fact, the single most challenging aspect of inversion is determining how to reconcile mathematical and statistical techniques with the underlying physical restrictions. With this in mind, any estimate of **a** should obey the following constraints for non-negativity, purity, and full or partial additivity. With regard to non-negativity, the abundances should be non-negative to be meaningful in a physical sense ($a_i \geq 0$, $i = 1, \ldots, M$). With regard to purity, a fractional abundance coefficient cannot exceed 1 ($a_i \leq 1$, $i = 1, \ldots, M$). With regard to additivity, full additivity requires the abundances for a mixed pixel to sum to one, with the implicit assumption that all the endmembers comprising the pixel spectrum in **x** are present in the columns of **S**,

$$\sum_{i=1}^{M} a_i = 1.$$

Partial additivity is a generalization that requires only the sum of abundances to be less than or equal to one, and it applies when the set of endmembers in the scene might be incomplete, which occurs when

$$\sum_{i=1}^{M} a_i < 1.$$

Several endmember-determination algorithms displayed in Figure 9 also simultaneously estimate the abundance vector **a** in the LMM. More often than not in unmixing, both the endmembers and abundances are unknown in a scene, and sometimes both quantities are sought simultaneously. In many situations, however, the endmembers for a scene are known, and the only remaining task is to recover the abundances. Both these families of algorithms are included in the inversion taxonomy in Figure 10. As discussed earlier, the algorithms are organized by the broad categories appearing in Figure 2. In addition, specific features concerning the outputs, inputs, and noise models used by these algorithms are included according to the model in Figure 3 to specifically distinguish the properties of inversion algorithms. With regard to outputs, do the algorithms yield abundance values that are continuous-valued or defined only on a discrete lattice of values? Does the estimation also uphold the full additivity, non-negativity, and purity constraints? With regard to inputs, do the algorithms utilize stochastic or deterministic endmembers? With regard to noise, do the algorithms employ a noise model?

Inversion Algorithms

Inversion algorithms are dominated by approaches that invoke some aspect of the method of least squares. Many statistical and parametric algorithms indirectly fall under the umbrella of least-squares analysis either by the squared-error-minimizing properties of the singular-value decomposition and eigendecomposition, or the Euclidean geometry of Gaussian analysis. The predominance of inversion methods based on least squares discussed here demonstrates how much of the current development of hyperspectral algorithms is based, correctly or not, on a single notion of distance.

The family of non-statistical, non-parametric, squared-error inversion algorithms begins with the unconstrained least squares (ULS) solution [3] for **a**,

$$\hat{\mathbf{a}}^U = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{x}.$$

Under the assumption of the LMM and no additive noise, this unconstrained estimate for **a** minimizes $|\mathbf{x} - \mathbf{S}\hat{\mathbf{a}}^U|^2$. This estimate exists when there are more bands than columns (a reasonable assumption for hyperspectral sensing), and when **S** has full column rank (i.e., endmembers are linearly independent). This estimate enforces none of the physical constraints on **a**.

The simplest variations on $\hat{\mathbf{a}}^U$ incorporate the key physical constraints on **a** by constraining the set of allowable solutions for $\hat{\mathbf{a}}$. Full additivity requires the abundances in **a** to sum to one [14], and this requirement restricts the solution to lie on the hyperplane given by

$$\sum_{i=1}^{M} a_i = 1.$$

The general solution for a least-squares estimate having linear constraints is given by

$$\hat{\mathbf{a}}^F = \hat{\mathbf{a}}^U - (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{Z}^T\left[\mathbf{Z}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{Z}^T\right]^{-1}(\mathbf{Z}\hat{\mathbf{a}}^U - \mathbf{b}),$$

where **Z** is a $1 \times M$ row vector having all ones and $\mathbf{b} = 1$. Closer examination of $\hat{\mathbf{a}}^F$ reveals that this solution consists of the unconstrained least-squares solution $\hat{\mathbf{a}}^U$ with an additive correction term that depends on the matrix of endmembers **S** and the error incurred by $\hat{\mathbf{a}}^U$ in satisfying the full additivity constraint.

The second constraint, non-negativity, is not as easily addressed in closed form as full additivity. Minimizing $|\mathbf{x} - \mathbf{S}\mathbf{a}|^2$ while maintaining $a_i \geq 0$, for $i = 1, \ldots, M$, falls in the domain of quadratic programming with linear inequalities as constraints. An alternative that has been employed in practice is the non-negative least-squares algorithm (NNLS) [15]. The approach here is to iteratively estimate **a** and, at every iteration, find a least-squares solution for just those coefficients of **a** which are negative by using only the associated columns of **a**. By selectively refining and moving those entries of **a** which still violate

the non-negativity condition, and integrating the newfound coefficients with the existing non-negative values, the algorithm acquires a new estimate of $\mathbf{a}$. The procedure is repeated until the algorithm converges to a final non-negative estimate $\hat{\mathbf{a}}^{NN}$, which has a subset of entries that are exactly zero. In comparison, whereas $\hat{\mathbf{a}}^{U}$ is an unconstrained solution for $\mathbf{a}$ that best minimizes $|\mathbf{x} - \mathbf{S}\mathbf{a}|^2$, $\hat{\mathbf{a}}^{NN}$ is a constrained solution that minimizes a similar but different criterion, $|\mathbf{x} - \mathbf{S}_{pos}\mathbf{a}|^2$, where $\mathbf{S}_{pos} = \mathbf{S}$ for those columns of $\mathbf{S}_{pos}$ where the associated entry of $\hat{\mathbf{a}}^{NN} > 0$, and the zero vector appears in columns where $\hat{\mathbf{a}}^{NN} = 0$.

In the previous examples of least-squares inversion algorithms, the common objective has been to estimate abundances that minimize the squared error between the actual spectrum and the approximated spectrum. The statistical and non-parametric analog of least-squares estimation minimizes the variance of the estimator. Given that the additive noise vector $\mathbf{w}$ in the LMM is a zero-mean random process and has a covariance $\mathbf{\Gamma}_w$, the minimum-variance estimate of the abundances $\hat{\mathbf{a}}^{V}$ is

$$\hat{\mathbf{a}}^{V} = (\mathbf{S}^T \mathbf{\Gamma}_w^{-1} \mathbf{S})^{-1} \mathbf{S}^T \mathbf{\Gamma}_w^{-1} \mathbf{x}.$$

This estimator of $\mathbf{a}$ is not only the best linear unbiased estimator (BLUE), it is also the minimum-variance unbiased estimator (MVUE).

Clustering algorithms that are used for endmember determination frequently also yield abundance estimates. Figure 11 depicts the results of using the fuzzy *K*-means partitions algorithm under the assumption of four endmembers and $q = 2.5$. By virtue of the rules for assigning abundance values to each pixel at each iteration, the abundance estimates necessarily fulfill the full additivity, purity, and non-negativity constraints. Endmember 1 resembles the spectrum of bare soil and road, whereas endmember 2 is most closely associated with grass and low brush. Endmembers 3 and 4 correspond to two different varieties of trees in the scene.

The family of statistical and parametric inversion algorithms consists of abundance estimators that derive estimates for $\mathbf{a}$ by optimizing either forward or posterior densities. Many of the maximum-likelihood approaches discussed earlier for endmember determination also estimate abundances simultaneously. Another maximum-likelihood formulation is the weighted Mahalanobis distance [16], which restructures the LMM by eliminating the uncertainty introduced by additive Gaussian noise, and considering the endmembers as stochastic, each having a multivariate Gaussian mean and covariance. The LMM can then be rewritten as $\mathbf{x} = (\mathbf{S} + \mathbf{E})\mathbf{a}$, where $\mathbf{E}$ is a random Gaussian matrix. Consequently, the covariance for $p_{\mathbf{x}|\mathbf{a}}(\mathbf{x}|\mathbf{a})$ is a sum of the covariances for each endmember, weighted by $\mathbf{a}$. The dependence of the covariance on $\mathbf{a}$ makes the optimal estimation of $\mathbf{a}$ considerably more difficult, and an iterative solution achieves the result.

Nonlinear Unmixing

While unmixing algorithms based on the LMM have emerged and proliferated, nonlinear unmixing algorithms have yet to demonstrate that the physical mechanisms present in a scene can be capably modeled to perform unmixing reliably. A few distinctive techniques, however, have emerged. They include the two-stream method [17], which incorporates multiple scattering into the expression for the bidirectional reflectance function for a surface consisting of particles of arbitrary shape in close proximity to one another. Inversion is accomplished by means of a single parameter known as the single-scattering albedo, which is the probability that a photon survives interactions such as Fresnel reflection, absorption, scattering, and diffraction due to the presence of a single grain of a substance.

A more macroscopic viewpoint is achieved from the perspective of radiosity [18]. Formulated for specific geometries of soil and vegetation, assessments of fractional leaf cover are estimated by balancing equations representing the transmitted and reflected radiation. Solutions have been formulated for simple canopy geometries in which multiple reflections occur between soil, leaf layers, layered canopies, and rough soil.

Not surprisingly, in contrast to the LMM, which trades physical specificity for analytical tractability, physical models do not easily lend themselves to simple mathematical solutions. Despite the complexity of these solutions, researchers continue to seek efficient ways to perform unmixing. As progress is

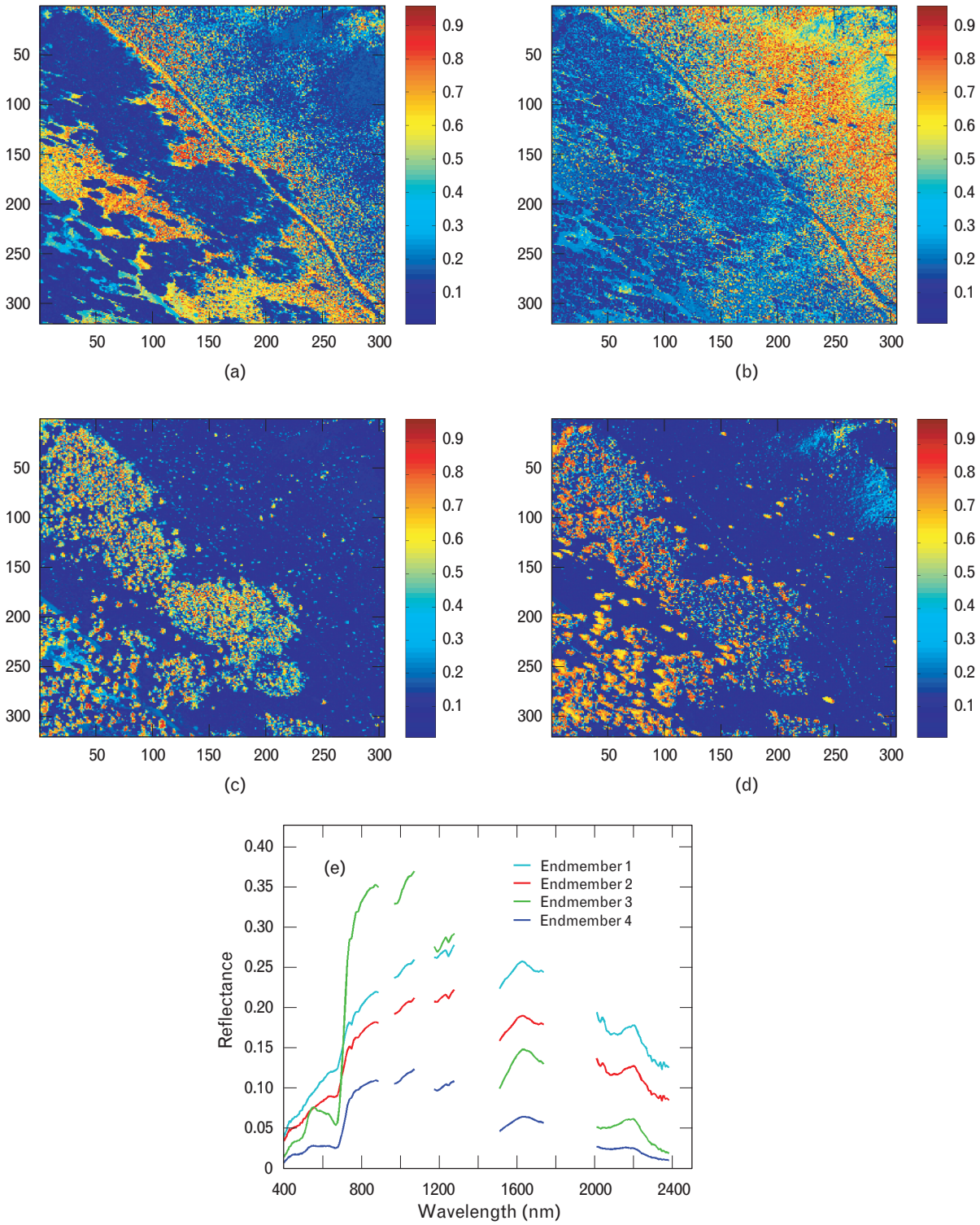

FIGURE 11. Abundance maps for four endmembers derived by using fuzzy *K*-means clustering with $q = 2.5$. (a) Abundance map for endmember 1; (b) abundance map for endmember 2; (c) abundance map for endmember 3; (d) abundance map for endmember 4; (e) endmember spectra. On the basis of knowledge of known materials in the scene, endmember 1 resembles the spectrum of bare soil and road, whereas endmember 2 is associated with grass and low brush, and endmembers 3 and 4 correspond to two distinct varieties of trees.

made, it is certain that a taxonomy of physically based unmixing algorithms will naturally emerge around core principles, in the same way that the organizing criteria for the linear spectral unmixing taxonomies discussed here became evident.

Conclusion

We have demonstrated an approach to hierarchically organizing the set of algorithms in the literature that comprise spectral unmixing. To organize the algorithms in this way, we discussed the progression of sensors that have to led to the development of hyperspectral sensing, and in parallel, we touched upon the evolution of passive remote sensing that has culminated in the ability to decompose mixed pixels into their constituent substances.

The taxonomies we have introduced have been constructed from the viewpoint that unmixing is an inverse problem that attempts to estimate important physical parameters from an electromagnetic signal that has interacted with the material of interest. Absolute knowledge of every variable is impractical, and this fact is manifested in varying degrees by the assumptions imposed by algorithms on the problems they attempt to solve.

Most importantly, the taxonomies reflect the wide disparity that exists in the approaches undertaken to solve the same problem. Hyperspectral sensing exemplifies a subject area that has drawn together an eclectic collection of participants, but increasingly this is the nature of many endeavors on the cutting edge of science and technology. Organizing and standardizing these approaches will be an increasingly important task for similar endeavors, and it is anticipated that future taxonomies developed for hyperspectral unmixing will reflect the increasing sophistication of a field that is rapidly maturing at the intersection of many different disciplines.

Acknowledgments

I would like to thank Peter W. Boettcher for implementing the geometric endmember-determination algorithm. I would also like to thank Carolyn Upham and Craig Richard for help in finding and processing data, and Glenn Cook for his dedication and close collaboration on the design and drawing of the taxonomies. Finally, I would like to thank Dimitris Manolakis, Gary Shaw, and John Kerekes for their technical contributions toward this effort.

REFERENCES

1. N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An Algorithm Taxonomy for Hyperspectral Unmixing," *SPIE* **4049,** 2000, pp. 42–63.
2. A. Tarantola and B. Valette, "Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion," *Rev. Geophys. Space Phys.* **20** (2), 1982, pp. 219–232.
3. G. Strang, *Linear Algebra and Its Applications*, 3rd ed. (Harcourt Brace Jovanovich, San Diego, 1988).
4. A.A. Green, M. Berman, P. Switzer, and M.D. Craig, "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal," *IEEE Trans. Geosci. Remote Sens.* **26** (1), 1988, pp. 65–74.
5. J.B. Lee, A.S. Woodyatt, and M. Berman, "Enhancement of High Spectral Resolution Remote-Sensing Data by a Noise-Adjusted Principal Components Transform," *IEEE Trans. Geosci. Remote Sens.* **28** (3), 1990, pp. 295–304.
6. J. Bowles, J. Antoniades, M. Baumback, J. Grossman, D. Haas, P. Palmadesso, and J. Stracka, "Real Time Analysis of Hyperspectral Data Sets Using NRL's ORASIS Algorithm," *SPIE* **3118,** 1997, pp. 38–45.
7. G.M. Foody. and D.P. Cox, "Sub-Pixel Land Cover Composition Estimation Using a Linear Mixture Model and Fuzzy Membership Model and Fuzzy Membership Functions," *Int. J. Remote Sens.* **15** (3), 1994, pp. 619–631.
8. J.C. Bezdek, R. Ehrlich, and W. Full, "FCM: The Fuzzy C-Means Clustering Algorithm," *Comput. & Geosci.* **10** (2–3), 1984, pp. 191–203.
9. S. Tompkins, J.F. Mustard, C.M. Pieters, and D.W. Forsyth, "Optimization of Endmembers for Spectral Mixture Analysis," *Remote Sens. Environ.* **59** (3), 1997, pp. 472–489.
10. A.D. Stocker and A.P. Schaum, "Application of Stochastic Mixing Models to Hyperspectral Detection Problems," *SPIE* **3071,** 1997, pp. 47–60.
11. J.W. Boardman, "Geometric Mixture Analysis of Imaging Spectrometry Data," *Proc. IGARSS 1994* **4,** *Pasadena, Calif., 8–12 Aug. 1994,* pp. 2369–2371.
12. M.D. Craig, "Minimum-Volume Transforms for Remotely Sensed Data," *IEEE Trans. Geosci. Remote Sens.* **32** (3), 1994, pp. 542–552.
13. F.P. Preparata and M.I. Shamos, *Computational Geometry: An Introduction* (Springer-Verlag, New York, 1985).
14. S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall Signal Processing Series, A.V. Oppenheim, ed. (Prentice Hall, Englewood Cliffs, N.J., 1993).
15. C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems,* Prentice Hall Series in Automatic Computation (Prentice Hall, Englewood Cliffs, N.J., 1974).
16. M.S. klein Gebbinck and T.E. Schouten, "Decomposition of Mixed Pixels," *SPIE* **2579,** 1995, pp. 104–114.
17. B. Hapke, "Bidirectional Reflectance Spectroscopy 1. Theory," *J. Geophys. Res.* **86** (B4), 1981, pp. 3039–3054.
18. C.C. Borel and S.A.W. Gerstl, "Nonlinear Spectral Mixing Models for Vegetative and Soil Surfaces," *Remote Sens. Environ.* **47** (3), 1994, pp. 403–416.

APPENDIX: DISTANCE METRICS, BAND SELECTION, AND DIMENSION REDUCTION

The subject of dimension reduction has become increasingly prominent for sensor systems that generate high volumes of data. With hundreds of spectral measurements for each spatial pixel, the processing of hyperspectral data from a single scene can be daunting. Moreover, the associated subsystems for a sensor must bear the burden of storage, transmission, and visualization of the data. In light of this, minimizing the quantity of data necessary for applications to succeed has become a top priority.

Dimension reduction differs from compression in that the goal for the former is to identify and preserve features in a lower-dimensional space that still allow algorithms to achieve acceptable performance. Compression, on the other hand, minimizes the quantity of data representing a signal to meet some bandwidth constraint (e.g., for storage or transmission) with the eventual goal of reconstructing the signal in its original dimension. In the case of images and video, the quality of the reconstruction is measured through its visual fidelity. However, if the goal is for algorithms to subsequently detect, classify, and identify small and possibly rare objects in the images, lower-dimensional representations must retain such information. In this sense, dimension-reduction algorithms include in their formulation constraints from the eventual application that will utilize the data, whereas compression schemes typically find compact representations of data independent of any subsequent analysis.

At the core of hyperspectral dimension-reduction and compression algorithms are distance metrics that quantify the difference between two K-dimensional signals [1]. A distance metric $d(\cdot,\cdot)$ yields a scalar measure of distance and must satisfy the following three basic properties:

$$d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x})$$
$$d(\mathbf{x},\mathbf{x}) = 0$$
$$d(\mathbf{x},\mathbf{z}) \le d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}).$$

Thus, given two K-dimensional signals $\mathbf{x}$ and $\mathbf{y}$, the distance $d(\mathbf{x}, \mathbf{y})$ between them can be measured several different ways, depending on the feature that the metric exploits. In the same way, a dimension reduction technique formulated around a specific distance metric identifies a reduced dimension that optimizes $d(\cdot,\cdot)$ over a class of signals.

For instance, principal-component analysis (PCA) is a statistical technique that exploits the Euclidean minimum distance, or EMD (also known as the 2-norm), as its distance metric:

$$d(\mathbf{x},\mathbf{y}) = \left\|\mathbf{x}-\mathbf{y}\right\|_2.$$

Figure 1(a) illustrates EMD for two vectors $\mathbf{x}$ and $\mathbf{y}$.

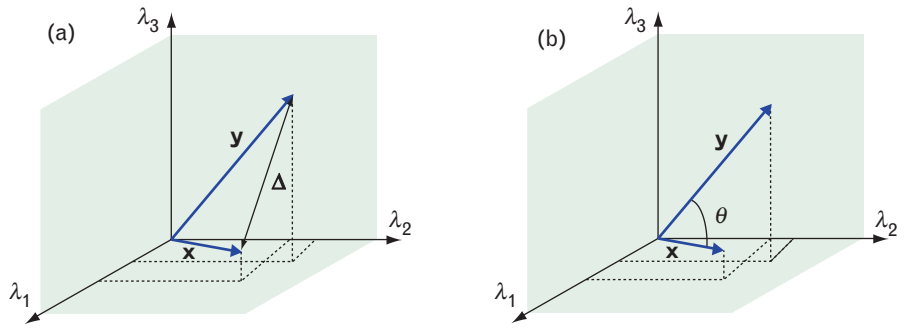

FIGURE 1. (a) The Euclidean minimum distance (EMD) is the scalar distance Δ between **x** and **y**; (b) the spectral angle mapper (SAM) gives the angle θ between **x** and **y**.

Given a zero-mean, K-dimensional random vector $\mathbf{x}$ having covariance $\mathbf{\Gamma}$, PCA essentially identifies the vector components that minimize the average distance between $\mathbf{x}$ and a rank k (where $k = 1, \ldots, K$) approximation of $\hat{\mathbf{x}}$.

$$E\left[d^2(\mathbf{x}, \hat{\mathbf{x}})\right] = E\left[\left\|\mathbf{x} - \hat{\mathbf{x}}\right\|_2^2\right].$$

The vector components that PCA identifies for dimension reduction are exactly the eigenvectors of Γ, and by the ordering provided by their associated eigenvalues, they can be used in linear transforms to achieve lower-dimensional representations that minimize the average representation error between $\mathbf{x}$ and $\hat{\mathbf{x}}$, as measured by EMD.

There is, however, no certainty that the reduced-dimension representations from PCA retain all the information necessary for algorithms to succeed. In fact, targets of interest that appear with a low probability, and whose spectrum is significantly different from the majority of the scene, can often be corrupted by statistical dimension-reduction techniques, causing subsequent classification or detection algorithms to fail. In short, although PCA (and implicitly EMD) may preserve the energy in signals, the conservation of energy may not be the driving factor that maintains algorithm performance.

EMD is one common metric in hyperspectral processing, but the most frequently used distance metric for comparing two spectra is the spectral angle mapper (SAM), which measures the angle between two hyperspectral signals. Given two K-dimensional spectra $\mathbf{x}$ and $\mathbf{y}$, the angle θ between them is given by

$$\theta(\mathbf{x}, \mathbf{y}) = \operatorname{acos}\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}, \quad 0 \le \theta \le \frac{\pi}{2}.$$

Figure 1(b) illustrates SAM for two vectors $\mathbf{x}$ and $\mathbf{y}$.

While SAM and EMD offer complementary measures derived from simple Euclidean geometry, SAM possesses mathematical properties that are useful for overcoming a particular form of signature variability. Atmospheric compensation algorithms convert radiance measurements from the sensor into estimates of the intrinsic reflectance value belonging to materials under observation. However, depending on its angular orientation with respect to the sun and the sensor, the measured radiance from a material can vary, and under ideal circumstances, the variation is in the form of a multiplicative constant. Since the angular orientation of objects observed by a sensor is unknown, the recovered estimate of the reflectance spectrum is also inexact, but known within an unknown near-multiplicative constant. Figure 2 depicts the spectra for several pixels derived from the same target, and while each pixel spectrum possesses similar artifacts, a near-multiplicative scaling is discernible between spectra.

By the definition of SAM, the angle created by two spectra is invariant to scaling by multiplicative constants a and b; namely, $\theta(a\mathbf{x}, b\mathbf{y}) = \theta(\mathbf{x}, \mathbf{y})$. Thus if $\mathbf{x}$ is a library spectrum for a target of interest that is compared to a test pixel $\mathbf{y}$, the invariance of SAM to multiplicative variability in $\mathbf{y}$ becomes a useful benefit.

SAM possesses another intriguing property. Again, let $\mathbf{x}$ and $\mathbf{y}$ be two K-dimensional spectra, but now represent each spectrum as the concatenation of two smaller vectors: $\mathbf{x} = [\,\mathbf{x}_A\ \mathbf{x}_B\,]$ and $\mathbf{y} = [\,\mathbf{y}_A\ \mathbf{y}_B]$. Then $\mathbf{x}_A$ and $\mathbf{y}_A$ are K_1-dimensional, and $\mathbf{x}_B$ and $\mathbf{y}_B$ are K_2-dimensional, such that $K_1 + K_2 = K$. A simple reorganization of $\cos\theta(\mathbf{x}, \mathbf{y})$ can be expressed as

$$\begin{aligned}\cos\theta(\mathbf{x}, \mathbf{y}) &= \left\{\frac{\langle \mathbf{x}_A, \mathbf{y}_A \rangle}{\|\mathbf{x}_A\|\|\mathbf{y}_A\|}\right\} \cdot \\ &\quad \left\{\frac{1 + \dfrac{\langle \mathbf{x}_B, \mathbf{y}_B \rangle}{\langle \mathbf{x}_A, \mathbf{y}_A \rangle}}{\sqrt{1 + \dfrac{\|\mathbf{x}_B\|^2}{\|\mathbf{x}_A\|^2}}\sqrt{1 + \dfrac{\|\mathbf{y}_B\|^2}{\|\mathbf{y}_A\|^2}}}\right\} \\ &= \cos\theta(\mathbf{x}_A, \mathbf{y}_A) \cdot \beta(\mathbf{x}_A, \mathbf{y}_A, \mathbf{x}_B, \mathbf{y}_B).\end{aligned}$$

In essence, this decomposition indicates that $\cos\theta(\mathbf{x}, \mathbf{y})$ can be interpreted as $\cos\theta(\mathbf{x}_A, \mathbf{y}_A)$, scaled by a function β of the band values in $\mathbf{x}_B$ and $\mathbf{y}_B$. More importantly, this decomposition provides the framework for a simple iterative technique, called band add-on (BAO), to identify a subset of bands $\mathbf{B}$ in $\mathbf{x}$ and $\mathbf{y}$ that increase their angular separation. Starting with an initial pair of bands to serve as $\mathbf{x}_A$ and $\mathbf{y}_A$ (e.g., the most or least orthogonal pair of bands in $\mathbf{x}$ and $\mathbf{y}$),

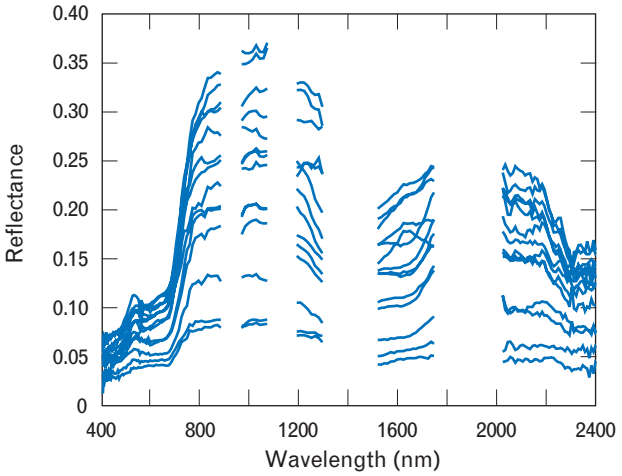

FIGURE 2. Reflectance spectra for several pixels derived from the same target vehicle. A multiplicative scaling effect is clearly evident in the spectra, but the shape of the spectra itself is nearly invariant.

the value of β can be scored for the remaining bands (individually or as groups) to determine how they scale $\cos\theta(\mathbf{x}_A, \mathbf{y}_A)$. The band or bands that yield the smallest $\beta < 1$ can then be appended to the starting pair of bands, and the process is repeated until no bands further increase the angle.

The procedure can be illustrated with two spectra, **x** and **y**, as shown in Figure 3(a). Here the red and blue curves represent the reflectance spectra for two different materials. From the original 210 spectral bands, several bands have been discarded (gaps in the spectra appear at those locations) because water-vapor absorption significantly degrades the measurements in those spectral intervals, leaving 144 usable bands. The angle between **x** and **y** using all 144 bands is 28.1°.

Band selection was performed on this pair of spectra by using two variants of the aforementioned techniques. In the first variation, BAO was initialized with the most orthogonal pair of bands (BAO-MO), and four bands were incrementally added. Figure 3(a) illustrates these six selections (629, 2099, 613, 417, 621, and 420 nm) as dashed magenta stripes. In the second variation, BAO was initialized with the least orthogonal pair of bands (BAO-LO), and five bands were subsequently added. Figure 3(a) illustrates these seven selections (466, 577, 2099, 629, 613, 417, and 420 nm) as dashed green stripes. Figure 3(b) shows the increase in spectral angle as bands are incrementally added for each method, and compares the results to the angle of 28.1° achieved with all 144 bands. Table 1 summarizes the results.

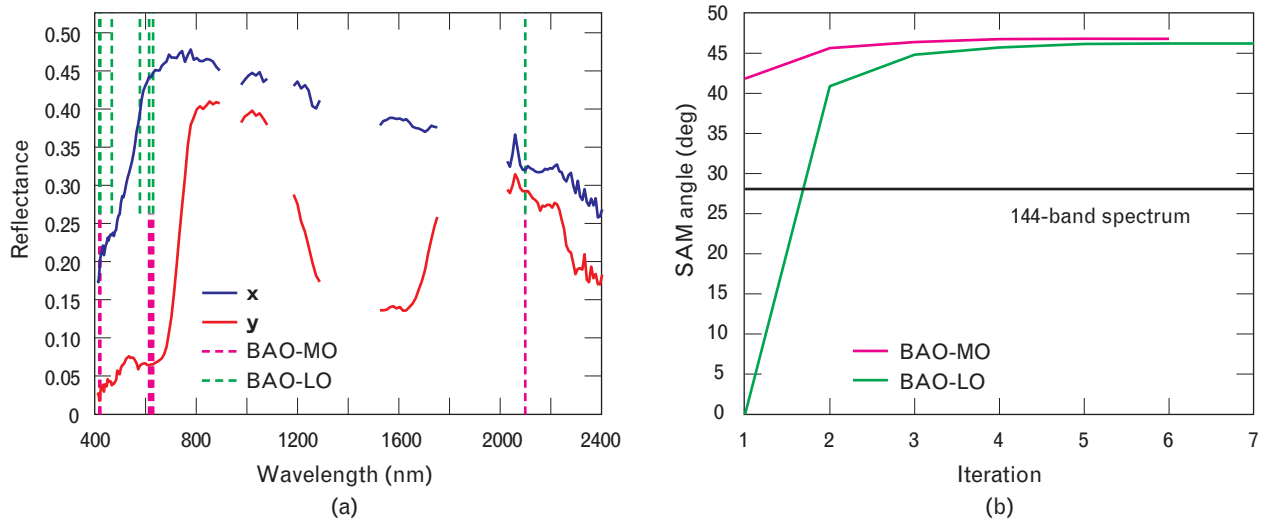

FIGURE 3. (a) Band selection results using two variations—least orthogonal (LO) and most orthogonal (MO)—of the band add-on (BAO) algorithm. The dashed magenta stripes indicate the wavelengths selected by BAO-MO, and the dashed green stripes indicate the wavelengths selected by BAO-LO. (b) This plot shows the SAM angle between two spectra for BAO-MO and BAO-LO as spectral bands are incrementally added. When all 144 usable bands are employed, the angle between the two spectra is 28.1°.

Table 1. Summary of Results for Two-Band Selection Algorithms That Maximize the Angle between Two Spectra

Method	*Starting bands (nm)*	*Additional bands (nm)*	*Total number of bands*	*Final angle (degrees)*
Least orthogonal (LO)	466, 577	2099, 629, 613, 417, 420	7	46.2
Most orthogonal (MO)	629, 2099	613, 417, 621, 420	6	46.8
Entire spectrum			144	28.1

Both variants of BAO exceeded the spectral angle produced when all 144 bands were used. Perhaps more compelling is that, despite the use of different initial conditions, five bands were commonly selected by both techniques, indicating a strong predilection for certain parts of the spectrum to increase angular contrast. While BAO is an inherently suboptimal technique for finding the largest angle, exhaustive evaluations of all possible subsets of bands have shown that BAO-MO frequently arrives at an angle near or equal to the maximum.

Band selection for SAM is an example of feature selection for a non-convex cost function, which is a common problem in pattern recognition. BAO has been extended to increase the angle between two classes of spectra, where each class consists of a handful of spectra, but not enough to estimate a covariance. This simple model for class variability strongly parallels the hyperspectral material identification problem, in which each material class is characterized by a handful of instrument measurements, and a distance metric compares an unknown pixel with a template from each class. Experiments have shown that band selection has significantly increased the angular contrast between classes and, hence, the capability of distinguishing spectrally similar targets, a common occurrence in camouflage and concealment scenarios [2,3]. Dimension reduction is a concomitant benefit. The strength of hyperspectral processing is in the physical measurements it collects. Exploiting metrics with a physical underpinning is a key part of taking full advantage of that strength.

REFERENCES

1. N. Keshava, "Distance Metrics and Band Selection in Hyperspectral Processing with Applications to Material Identification and Spectral Libraries," Project Report HTAP-12, Lincoln Laboratory (18 Dec. 2002).
2. N. Keshava and P. Boettcher, "On the Relationships between Physical Phenomena, Distance Metrics, and Best Bands Selection in Hyperspectral Processing," *SPIE* **4381,** 2001, pp. 55–67.
3. N. Keshava, "Angle-Based Band Selection for Material Identification in Hyperspectral Processing," *SPIE* **5093,** 2003.

NIRMAL KESHAVA
is a staff member in the Laser and Sensor Applications group. The focus of his research is in mathematical and physical models for hyperspectral signal processing. He has been at Lincoln Laboratory since 1997. Previously, he worked for Bell Communications Research in the Operations Technology group, where he developed requirements for network elements that perform alarm control and performance monitoring. He has a B.S. degree from UCLA and an M.S. degree from Purdue University, both in electrical engineering. He received a Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, where his doctoral thesis examined ways to represent Gaussian random processes by wavelet packet structures using stochastic distance measures. He is a member of the IEEE and the Federation of American Scientists.