

Hyperspectral Unmixing: Geometrical, Statistical, and Sparse Regression-Based Approaches

José M. Bioucas-Dias^a and Antonio Plaza^b

^a Instituto de Telecomunicações, Instituto Superior Técnico, Technical University of Lisbon, Portugal (e-mail: bioucas@lx.it.pt)

^b Department of Technology of Computers and Communications, University of Extremadura, E-10071 Caceres, Spain (e-mail: aplaza@unex.es)

ABSTRACT

Hyperspectral instruments acquire electromagnetic energy scattered within their ground instantaneous field view in hundreds of spectral channels with high spectral resolution. Very often, however, owing to low spatial resolution of the scanner or to the presence of intimate mixtures (mixing of the materials at a very small scale) in the scene, the spectral vectors (collection of signals acquired at different spectral bands from a given pixel) acquired by the hyperspectral scanners are actually mixtures of the spectral signatures of the materials present in the scene.

Given a set of mixed spectral vectors, spectral mixture analysis (or spectral unmixing) aims at estimating the number of reference materials, also called endmembers, their spectral signatures, and their fractional abundances. Spectral unmixing is, thus, a source separation problem where, under a linear mixing model, the sources are the fractional abundances and the endmember spectral signatures are the columns of the mixing matrix. As such, the independent component analysis (ICA) framework came naturally to mind to unmix spectral data. However, the ICA crux assumption of source statistical independence is not satisfied in spectral applications, since the sources are fractions and, thus, non-negative and sum to one. As a consequence, ICA-based algorithms have severe limitations in the area of spectral unmixing, and this has fostered new unmixing research directions taking into account geometric and statistical characteristics of hyperspectral sources.

This paper presents an overview of the principal research directions in hyperspectral unmixing. The presentation is organized into four main topics: **i)** mixing models, **ii)** signal subspace identification, **iii)** geometrical-based spectral unmixing, **iv)** statistical-based spectral unmixing, and **v)** sparse regression-based unmixing. In each topic, we describe what physical or mathematical problems are involved and summarize state-of-the-art algorithms to address these problems.

1. INTRODUCTION

The development of hyperspectral airborne and spaceborne sensors has improved the capability of ground-based data collection in the fields of agriculture, geography, geology, mineral identification, detection, and classification of targets activities.^{1–5}

Hyperspectral sensors use many contiguous bands of high spectral resolution covering the visible, near-infrared, and shortwave infrared spectral bands ($0.3\mu\text{m} - 2.5\mu\text{m}$).⁶ Letting alone the effects of the atmosphere,

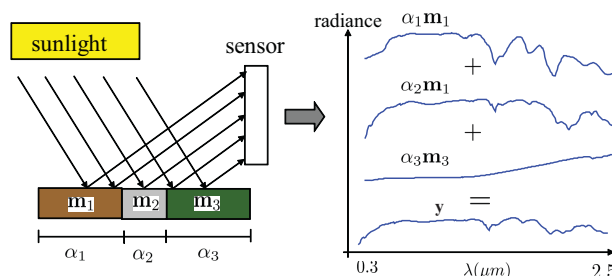


Figure 1. Schematic view of the spectral linear mixing process.

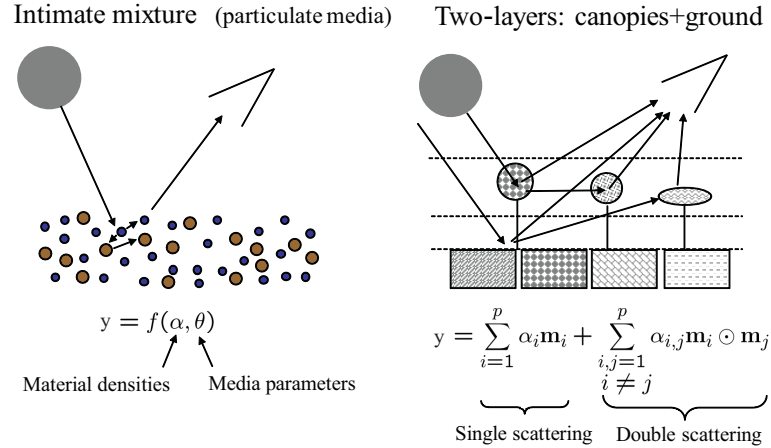


Figure 2. Two nonlinear mixing scenarios. Left hand: intimate mixture; Right hand: multilayered scene.

the signal recorded by a hyperspectral sensor at a given band and from a given pixel is often a mixture of the light scattered by the constituent substances located in the respective pixel spatial coverage.⁷ Figure 1 schematizes the spectral mixing process: at each band, the sensor reads the light scattered by a scene composed of three materials. The acquired spectral vector $\mathbf{y} \in \mathbb{R}^B$ (B stands for the number of bands) is a function (linear in this example) of the material spectral signatures and of the respective fractions they occupy inside the pixel.

1.1 Linear and nonlinear mixing models

Hyperspectral unmixing is a technique which decomposes the pixel spectra into a collection of constituent spectra, or spectral signatures, and their corresponding fractional abundances that quantify the proportion of each material, also termed endmember, present in the pixel. Depending on the mixing scales at each pixel and on the geometry of the scene, the observed mixture is either linear or nonlinear.^{3,8} Linear mixing holds when the mixing scale is macroscopic⁹ and the incident light interacts with just one material, as it happens in checkerboard type scenes.^{10,11} This scenario is schematized in Figure 1. Nonlinear mixing holds when the light suffers multiple scattering involving different materials.¹² Figure 2 illustrates two non-linear mixing scenarios: the left hand part represents an intimate mixture, meaning that the materials are in close proximity; the right hand part illustrates a multilayered scene, where there are multiple interactions among the scatterers at the different layers.

- In a linear mixing scenario, the acquired spectral vectors are a linear combination of the endmembers signatures present in the scene weighted by the respective fractional abundances. The exploitation of this model, in spite of its simplicity, has fostered a huge amount of research leading to a plethora of unmixing algorithms developed under the geometrical or the statistical frameworks.
- In a nonlinear mixing scenario, the model for the scattered light is much more complex than its linear counterpart. Radiative transfer theory (RTT)¹³ is a well established mathematical model for the transfer of energy as photons interact with the materials in the scene. The core of the RTT is a differential equation describing radiance read by the sensor. It can be derived via the conservation of energy and the knowledge of the phase function, which represents the probability of light with a given propagation direction be scattered into a specified angle solid around a given scattering direction.

Most of this overview is devoted to the linear mixing model. The reason is that, despite its simplicity, it is an acceptable approximation of the light scattering mechanisms in many real scenarios. Furthermore, the linear mixing model constitutes the basis of many effective unmixing algorithms. This is to be contrasted with the nonlinear mixing model, where the inference of the spectral signatures and of material densities based on the RTT is a complex ill-posed problem, relying on scene parameters often very hard to obtain. A way to sidestep these difficulties is to formulate unmixing as regression problem based, for example, on neural networks or on kernels, in which the model parameters are learnt in a supervised fashion from a collection of examples (see¹⁴ and

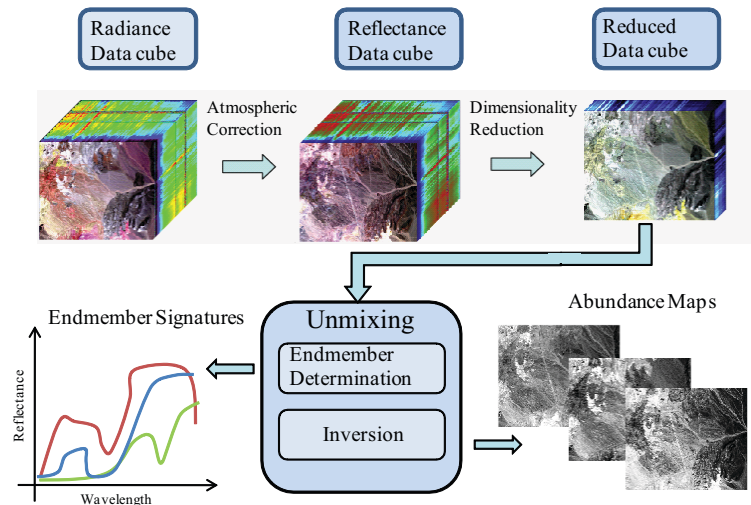


Figure 3. Schematic diagram of the hyperspectral unmixing process.

references therein). Anyway, there are particular situations in which a non-linear model can be converted into a linear one, as it is the case of the two-stream method.¹⁰

1.2 Hyperspectral unmixing processing chain

Figure 3 shows the processing steps usually involved in the hyperspectral unmixing chain: atmospheric correction, dimensionality reduction, endmember determination, and inversion. Often, endmember determination and inversion are implemented simultaneously. Below, we provide a brief characterization of each of these steps:

1. **Atmospheric correction.** The atmosphere attenuates and scatters the light. Its presence affects, therefore, the radiance at the sensor. The atmospheric correction compensates for these effects by converting radiance into reflectance, which is an intrinsic property of the materials. We stress, however, that linear unmixing can be carried out directly on radiance data as far as the the atmosphere affects equally all pixels in the scene.
2. **Data reduction.** Very often the hyperspectral vectors live in a subspace of very low dimension compared with the available number of bands. The identification of this subspace enables a correct dimensionality reduction, yielding gains in algorithm performance and complexity and in data storage. Furthermore, in the linear mixture case, the signal subspace dimension is equal to the number of endmembers, a crucial figure in hyperspectral unmixing.
3. **Endmember determination.** The endmember determination step consists in identifying the endmembers in the scene. There are basically two frameworks: the geometrical, which exploits the fact that linear mixed vectors are in a simplex set or in a positive cone, and the statistical, which models the abundance fractions as random variables and formulates the endmember determination as a statistical inference problem. Meanwhile, a new framework is emerging: sparse regression. In this framework, the unmixing is formulated as a linear sparse regression problem, in a fashion similar to that of compressive sensing.
4. **Inversion.** Given the observed spectral vectors and the identified endmembers, the inversion step consists in solving a constrained least squares problem which minimizes the residual between the observed spectral vectors and the linear space spanned by the inferred spectral signatures; the implicit fractional abundances are, very often, constrained to be nonnegative and to sum one (*i.e.*, they belong to the probability simplex). There are, however, many hyperspectral unmixing approaches in which the endmember determination and inversion steps are implemented simultaneously.

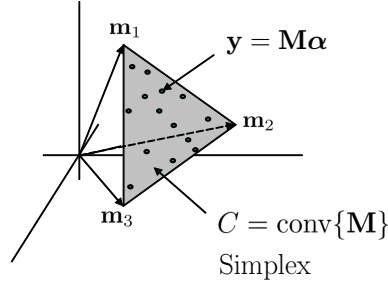


Figure 4. Illustration of the simplex set C for $p=3$ (C is the convex hull of the columns of \mathbf{M} , $C = \text{conv}\{\mathbf{M}\}$). Points in green denote spectral vectors. The vertices of the simplex correspond to the endmembers.

The next sections address linear spectral mixture model (Section 2), subspace identification (Section 3), linear spectral unmixing (Section 4), and experimental results (Section 5). Each section introduces the underlying problem and summarizes state-of-the-art algorithms to address such problem.

2. LINEAR MIXTURE MODEL

In spectral mixture modeling, the basic assumption is that the surface is made of a few number of endmembers of relatively constant spectral signatures, or, at least, constant spectral shapes. If the multiple scattering among distinct endmembers is negligible and the surface is partitioned according to the fractional abundances, then the spectral reflectance of each pixel is well approximated by a linear mixture of endmember reflectances weighted by the correspondent fractional abundances.^{3,7,8,15} Assuming that the sensor radiation pattern is ideal, *i.e.*, constant in the instantaneous field of view and zero outside, the reflectance at channel i from a given pixel is

$$y_i = \sum_{j=1}^p \rho_{ij} \alpha_j + w_i, \quad (1)$$

where ρ_{ij} denotes the reflectance of endmember j at wavenumber λ_i , α_j denotes the fractional abundance of endmember j at the considered pixel, and p is the number of endmembers. Fractional abundances are subject to the sum-to-one and positivity constraints:

$$\sum_{j=1}^p \alpha_j = 1, \quad \alpha_j \geq 0, \quad j = 1, \dots, p, \quad (2)$$

i.e., the fractional abundance vector $\boldsymbol{\alpha} \equiv [\alpha_1, \alpha_2, \dots, \alpha_p]^T$ is in the probability simplex. The notation $(\cdot)^T$ indicates vector transposed.

Let \mathbf{y} be an $B \times 1$ vector, where B is the total number of bands, and $\mathbf{m}_j \equiv [\rho_{1j}, \rho_{2j}, \dots, \rho_{Bj}]^T$ be the so-called signature of the j th endmember. Expression (1) can then be written as

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{w}, \quad (3)$$

where $\mathbf{M} \equiv [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p]$ is the so-called mixing matrix containing the signatures of the endmembers present in the covered area, and $\mathbf{w} \equiv [w_1, \dots, w_B]^T$ models additive noise. Assuming that the columns of \mathbf{M} are affinely independent, *i.e.*, $\mathbf{m}_2 - \mathbf{m}_1, \mathbf{m}_3 - \mathbf{m}_1, \dots, \mathbf{m}_p - \mathbf{m}_1$ are linearly independent, then the set

$$C \equiv \{\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} : \sum_{j=1}^p \alpha_j = 1, \alpha_j \geq 0, j = 1, \dots, p\}$$

is a $(p-1)$ -simplex in \mathbb{R}^B . Figure 4 illustrates the simplex set C for an hypothetical mixing matrix \mathbf{M} containing three endmembers (C is the convex hull of the columns of \mathbf{M}). Points in green denote spectral vectors, whereas the vertices of the simplex correspond to the endmembers. Note that the inference of the mixing matrix \mathbf{M} amounts to identify the vertices of the simplex C . This geometrical point of view, exploited by many unmixing algorithms, will be further developed in Sections 4.1 and 4.2.

3. SIGNAL SUBSPACE IDENTIFICATION

The number of endmembers present in a given scene is, very often, much smaller than the number of bands B . Therefore, spectral vectors lie in a low dimensional linear subspace. The identification of this subspace enables the representation of spectral vectors in a low dimensional subspace, thus yielding gains in computational time and complexity, in data storage, and in *signal-to-noise-ratio* (SNR). Furthermore, several unmixing algorithms only work on the signal subspace requiring, therefore, a signal subspace identification as a first processing step.

Unsupervised subspace identification has been approached in many ways. Band selection or band extraction, as the name suggests, exploits the high correlation existing between adjacent bands to select a few spectral components among those with higher SNR.^{16,17} Projection techniques seek for the best subspace to project data by minimizing an objective function. For example, *principal component analysis* (PCA)¹⁸ computes the *Karhunen-Loève* transform, which is the best data representation in the least squares sense; *singular value decomposition* (SVD)¹⁹ provides the projection that best represents data in the maximum power sense; if the noise is additive and white and the subspace dimension is known, say p , then a *maximum likelihood* (ML) basis for the estimated signal subspace is given by the first p eigenvectors of the empirical correlation matrix;¹⁹ *maximum noise fraction* (MNF)²⁰ and *noise adjusted principal components* (NAPC)²¹ seek the projection that optimizes the ratio of noise power to signal power. NAPC is mathematically equivalent to MNF²¹ and can be interpreted as a sequence of two principal component transforms: the first applies to the noise and the second applies to the transformed data set.

On the other hand, topological methods are local approaches that infer the manifold, usually of low dimension, where data set live.²² For example, curvilinear component analysis,²³ curvilinear distance analysis,²⁴ and manifold learning^{25–28} are non-linear projections based on the preservation of the local topology. Independent component analysis,^{29,30} projection pursuit,^{31,32} and wavelet decomposition^{33,34} have also been considered.

The identification of the signal subspace is a model order inference problem to which information theoretic criteria like the *minimum description length* (MDL)^{37,38} or the *Akaike information criterion* (AIC)⁴¹ comes to mind. These criteria have in fact been used in hyperspectral applications⁴² adopting the approach introduced by Wax and Kailath in.³⁹ In turn, Harsanyi, Farrand, and Chang⁴⁰ developed a Neyman-Pearson detection theory-based thresholding method (HFC) to determine the number of spectral endmembers in hyperspectral data, referred to in⁴² as *virtual dimensionality* (VD). The HFC method is based on a detector built on the eigenvalues of the sample correlation and covariance matrices. A modified version, termed noise-whitened HFC (NWHFC), includes a noise-whitening step.⁴²

Finally, the HySime (hyperspectral signal identification by minimum error)⁴³ is a recently introduced minimum mean squared error based approach to infer the signal subspace in hyperspectral imagery. The method is eigendecomposition based, unsupervised, and fully-automatic (*i.e.*, it does not depend on any tuning parameters). It first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense.

3.1 Projection on the signal subspace

Assume that the signal subspace \mathcal{S} has been identified by using one of the above referred to methods and let $\{\mathbf{f}_1, \dots, \mathbf{f}_p\}$ an orthonormal basis for \mathcal{S} , where $\mathbf{f}_i \in \mathbb{R}^B$, for $i = 1, \dots, p$. The coordinates of the orthogonal projection of a spectral vector $\mathbf{y} \in \mathbb{R}^B$ onto \mathcal{S} , with respect to the basis $\mathbf{F} \equiv [\mathbf{f}_1, \dots, \mathbf{f}_p]$, are given by $\mathbf{y}_S = \mathbf{F}^T \mathbf{y} \in \mathbb{R}^p$. Replacing \mathbf{y} with the observation model (3), we have

$$\mathbf{y}_S = \mathbf{F}^T \mathbf{M} \boldsymbol{\alpha} + \mathbf{F}^T \mathbf{w}.$$

As already referred to before, the projection of the observed spectral vectors onto the signal subspace yields very often large computational, storage, and SNR gains. The first two are a direct consequence of $p \ll B$ in most applications; to briefly explain the latter, let us assume that the noise is zero-mean and has covariance $\sigma^2 \mathbf{I}$ (\mathbf{I} stands for the identity matrix). The mean power of the projected noise term $\mathbf{F}^T \mathbf{w}$ is then $\mathbb{E} \|\mathbf{F}^T \mathbf{w}\|^2 = \sigma^2 p$ ($\mathbb{E}(\cdot)$ denotes mean value). The relative attenuation of the noise power implied by the projection is then p/B .

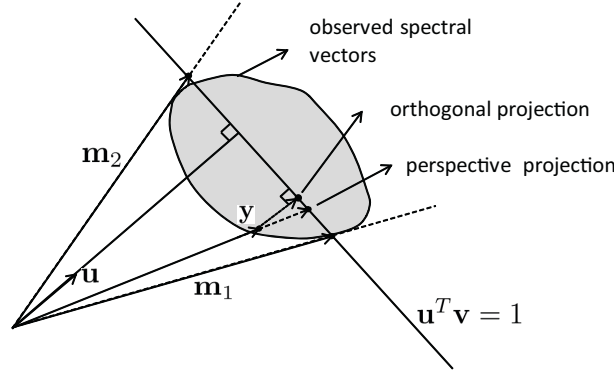


Figure 5. Projections of the observed data onto an hyperplane. a) Perspective projection: spectral vectors \mathbf{y} are in a cone; the scaling $\mathbf{y}/(\mathbf{y}^T \mathbf{u})$ brings them to the hyperplane defined by $\mathbf{y}^T \mathbf{u} = 1$. b) Orthogonal projection on an hyperplane.

3.2 Affine set projection

From now on, we assume that the observed data set has been projected onto the signal subspace and, for lightness of notation, we still represent the projected vectors as in (3), that is

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{w}, \quad (4)$$

where $\mathbf{y}, \mathbf{w} \in \mathbb{R}^p$ and $\mathbf{M} \in \mathbb{R}^{p \times p}$. Since the columns of \mathbf{M} belong to the subspace signal, the original mixing matrix is simply given by $\mathbf{F}\mathbf{M}$.

Model (4) is an oversimplification of reality, as it does not take into account signature variability (from pixel to pixel) due to changes in the composition of substances, surface contaminants and other variations such as age-induced color fading due to oxidation or bleaching, uncompensated atmospheric and environmental effects, and uncompensated errors in the sensor. Signature variability has been studied and accounted for in a few unmixing algorithms (see, e.g.,^{44–46}). This phenomenon is primarily characterized by spectral shape invariance;⁴⁷ *i.e.*, while the spectral shapes of the endmembers are fairly consistent, their amplitude varies considerably over the scene. This means that the endmember signatures are affected by a positive scale factor that varies from pixel to pixel.

In the aforementioned circumstances and in the absence of noise, the observed spectral vectors are no longer included in a simplex but rather in a cone as illustrated in Fig. 5. Notice that, if we preserve the observation model $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}$ (noise is neglected), the components of the fractional abundance vector $\boldsymbol{\alpha}$ do not sum-to-one anymore, although they are still positive. In order to ensure that the fractional abundances sum to one, and thus to recover the linear model $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ belongs to the probability simplex, one of the following procedures may be applied:

- Perspective projection:** This is the so-called dark point fixed transform (DPFT) proposed in.⁶⁴ For a given observed vector \mathbf{y} , this projection, illustrated in Fig. 5, amounts to rescale \mathbf{y} according to $\mathbf{y}/(\mathbf{y}^T \mathbf{u})$. The affine hyperplane containing the projected vectors is the set of vectors $\mathbf{v} \in \mathbb{R}^p$ satisfying $\mathbf{v}^T \mathbf{u} = 1$.
- Orthogonal projection:** Use PCA to identify the affine set that best represent the observed data in the least squares sense and then compute the orthogonal projection of the observed vectors onto this set (see⁴⁸ for details). This procedure is also illustrated in Fig. 5.

Notice that the procedure a) modifies the direction of the spectral vectors whereas b) do not. Perhaps, the best approach is to identify vector \mathbf{u} using b) and then apply the perspective projection described in a).

4. LINEAR SPECTRAL UNMIXING

Linear spectral unmixing has been intensively researched in the recent years.^{3,14,15,47,49–52} Under this model, and assuming that the number of substances and their reflectance spectra are known, hyperspectral unmixing is a linear problem to which many solutions have been proposed (e.g., maximum likelihood estimation,⁵³ spectral signature matching,⁵⁴ spectral angle mapper,⁵⁵ subspace projection methods,^{56,57} and constrained least squares⁵⁸).

In most cases, the number of substances and their reflectances are not known and, then, hyperspectral unmixing falls into the class of blind source separation problems.⁵⁹ Independent Component Analysis (ICA) has recently been proposed as a tool to blindly unmix hyperspectral data.^{60–62} ICA is based on the assumption of mutually independent sources (abundance fractions), which is not the case of hyperspectral data, since the sum of abundance fractions is constant, implying statistical dependence among them. This dependence compromises ICA applicability to hyperspectral data as shown in.^{15,63} In fact, ICA finds the endmember signatures by multiplying the spectral vectors with an unmixing matrix which minimizes the mutual information among channels. If sources are independent, ICA provides the correct unmixing, since the minimum of the mutual information corresponds to and only to independent sources. This is no longer true for dependent fractional abundances. Nevertheless, some endmembers may be approximately unmixed. These aspects are addressed in.⁶³

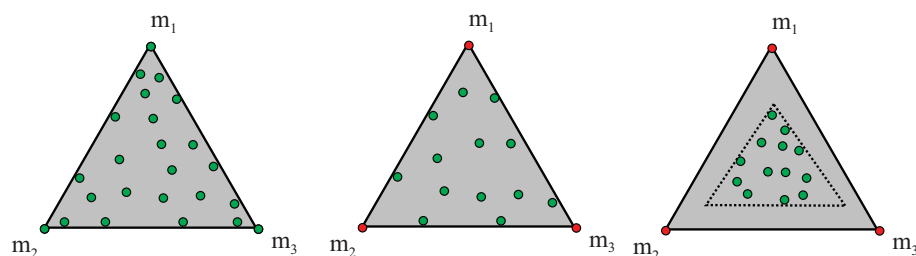


Figure 6. Illustration of the concept of simplex of minimum volume containing the data for three data sets. The endmembers in the left hand side and in the middle are identifiable by fitting a simplex of minimum volume to the data, whereas this is not applicable to the right hand side data set. The former data set correspond to a highly mixed scenario.

To overcome the limitations of the ICA-based approaches to spectral unmixing, large research efforts have been devoted in the last decade to the development of unmixing algorithms targeted at spectral unmixing. Most of these algorithms adopt either a geometrical or a statistical framework.⁴⁹ To motivate these two directions, let us consider the three data sets shown in Fig. 6 generated under the linear model

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{w} \quad \text{subject to:} \quad \sum_{i=1}^p \alpha_i = 1, \quad \alpha_i \geq 0, \quad (5)$$

where the noise is assumed to be negligible. As shown in Section 2, the spectral vectors generated according to (5) are in a simplex whose vertices correspond to the endmembers. The left hand side data set contains pure pixels, *i.e.*, for any of the p endmembers there is at least one pixel containing only the correspondent material; the data set in the middle does not contain pure pixels but contains at least $p - 1$ spectral vectors on each facet. In both data sets (left and middle), the endmembers may be inferred by fitting a minimum volume (MV) simplex to the data; this rather simple and yet powerful idea, introduced by Craig in his seminal work,⁶⁴ underlies several geometrical based unmixing algorithms.

The MV simplex shown in the right hand side example of Fig. 6 is smaller than the true one. This situation corresponds to a highly data set where there are no spectral vectors near the facets. For these class of problems, we usually resort to the statistical framework in which the estimation of the mixing matrix and of the fractional abundances are formulated as a statistical inference problem by adopting suitable probability models for the parameters involved, namely for the fractional abundances.

In the Subsections 4.1-4.2 and 4.3, we address in more detail, respectively, the geometrical and the statistical unmixing approaches. In the case of the geometrical approach, we consider separately the presence (Subsection 4.1) and the absence (Subsection 4.2) of pure pixels in the data set, as the algorithms for the pure pixel scenario are

much simpler than those for non-pure pixel case. In Subsection 4.4, we address sparse regression-based unmixing, a very recent spectral unmixing research front, which avoids, in given circumstances, the direct inference of the mixing matrix.

4.1 Geometrical based approaches: pure pixel based algorithms

The pure pixel based algorithms still belong to the MV class but assume the presence in the data of at least one pure pixel per endmember, meaning that there is at least one spectral vector on each vertex of the data simplex. This assumption, though enabling the design of very efficient algorithms from the computational point of view, is a strong requisite that may not hold in many datasets. In any case, these algorithms find the set of most pure pixels in the data. Anyway, perhaps due to its computational lightness and clear conceptual meaning, they are, by far, the class most used in linear hyperspectral unmixing applications.

- The pixel purity index (PPI) algorithm^{65,66} uses MNF as a preprocessing step to reduce dimensionality and to improve the SNR. PPI projects every spectral vector onto skewers (large number of random vectors). The points corresponding to extremes, for each skewer direction, are stored. A cumulative account records the number of times each pixel (i.e., a given spectral vector) is found to be an extreme. The pixels with the highest scores are the purest ones.
- N-FINDR⁶⁷ is based on the fact that in spectral dimensions, the volume defined by a simplex formed by the purest pixels is larger than any other volume defined by any other combination of pixels. This algorithm finds the set of pixels defining the largest volume by inflating a simplex inside the data.
- The iterative error analysis (IEA) algorithm⁶⁸ implements a series of linear constrained unmixings, each time choosing as endmembers those pixels which minimize the remaining error in the unmixed image.
- The vertex component analysis (VCA) algorithm⁶⁹ iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are exhausted.
- The simplex growing algorithm (SGA)⁷⁰ iteratively grows a simplex by finding the vertices corresponding to the maximum volume.
- The sequential maximum angle convex cone (SMACC) algorithm⁷¹ is based on a convex cone for representing the spectral vectors. The algorithm starts with a single endmember and increases incrementally in dimension. A new endmember is identified based on the angle it makes with the existing cone. The data vector making the maximum angle with the existing cone is chosen as the next endmember to enlarge the endmember set. The algorithm terminates when all of the data vectors are within the convex cone, to some tolerance.

4.2 Geometrical based approaches: Minimum volume based algorithms

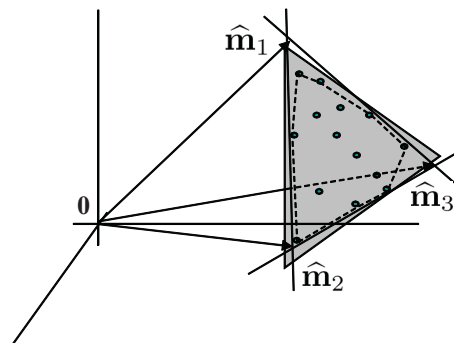


Figure 7. Illustration of the concept of simplex of minimum volume containing the data.

The MV approaches aim at finding the mixing matrix \mathbf{M} minimizing the volume of the simplex defined by its columns and containing the observed spectral vectors. This is a nonconvex optimization problem much harder than those considered in the previous subsection in which the endmembers belong to the data set.

Fig. 7 further illustrates the concept of simplex of minimum volume containing the data. The estimated mixing matrix $\widehat{\mathbf{M}} \equiv [\widehat{\mathbf{m}}_1, \widehat{\mathbf{m}}_2, \widehat{\mathbf{m}}_3]$ differs slightly from the true mixing matrix because there is not enough data points per facet (necessarily $p - 1$ per facet) to define the true simplex. The simplex $\mathbf{conv}(\mathbf{M})$, *i.e.*, the convex hull of the columns of $\mathbf{M} \in \mathbb{R}^{p \times p}$, lives in an affine set in \mathbb{R}^p and therefore it has volume zero in \mathbb{R}^p . To obtain a nondegenerated volume, the extended simplex $\mathbf{M}_0 \equiv [\mathbf{0}, \mathbf{M}]$, containing the origin, is usually considered. We recall that the volume of $\mathbf{conv}(\mathbf{M}_0)$, the convex hull of \mathbf{M}_0 , is given by

$$V(\mathbf{M}_0) \equiv \frac{|\det(\mathbf{M})|}{p!}.$$

Note that minimizing the volume of $\mathbf{conv}(\mathbf{M})$ in \mathbb{R}_{p-1} and minimizing $V(\mathbf{M}_0)$, both subject to the constraint that the data set belongs to $\mathbf{conv}(\mathbf{M})$ are equivalent problems.

Assuming that the vectors $\mathbf{m}_i - \mathbf{m}_1 \in \mathbb{R}^p$ for $i = 2, \dots, p$, are affinely independent, an alternative to computing the volume of the simplex $\mathbf{conv}(\mathbf{M}_0)$ consists in first representing these $p - 1$ vectors in the $(p - 1)$ -dimensional subspace $\text{span}\{\mathbf{m}_i - \mathbf{m}_1, i = 2, \dots, p\}$ and then computing

$$\begin{aligned} V(\mathbf{M}) &\equiv \frac{1}{(p-1)!} |\det[\mathbf{m}_2 - \mathbf{m}_1, \dots, \mathbf{m}_p - \mathbf{m}_1]| \\ &= \frac{1}{(p-1)!} \left| \det \begin{bmatrix} 1 & \cdots & 1 \\ \mathbf{m}_0 & \cdots & \mathbf{m}_p \end{bmatrix} \right|. \end{aligned}$$

Craig's work,⁶⁴ published in 1994, put forward the seminal concepts regarding the algorithms of MV type. After identifying the subspace and applying projective projection (DPFT), the algorithm iterates each facet of the simplex such that the volume

$$V(\widehat{\mathbf{M}}_0) \equiv \frac{\text{abs}(|\widehat{\mathbf{M}}|)}{p!}$$

is minimized and all spectral vectors are belongs to this simplex; *i.e.*, $\widehat{\mathbf{M}}^{-1} \mathbf{y}_i \succeq 0$, for $i = 1, \dots, n$. In a more formal way: **for** $t = 1, \dots$,

$$\begin{aligned} \widehat{\mathbf{M}}_0^{t+1} &= \arg \min_{\mathbf{M}_0} V(\mathbf{M}_0) \\ \text{s.t.:} & \quad \text{facets}(\mathbf{M}_0) = \text{facets}(\widehat{\mathbf{M}}_0^t), \text{ except for facet } i = (t \bmod p) \\ \text{s.t.:} & \quad \mathbf{M}^{-1} \mathbf{y}_i \succeq 0, \text{ for } i = 1, \dots, n. \end{aligned}$$

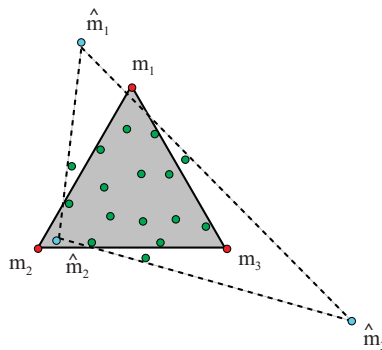


Figure 8. Noisy data. The dashed simplex represents the simplex of minimum volume; by allowing violations to the positivity constraint, the MVSA and SISAL algorithms yield a simplex very close to the true one.

The minimum volume simplex analysis (MVSA)⁷² and the simplex identification via variable splitting and augmented Lagrangian (SISAL)⁷³ implement a robust version of the MV concept. The robustness is introduced by allowing the positivity constraint to be violated. To grasp the relevance of this modification, we show in Fig. 8, a set of noisy spectral vectors. Due to the presence of noise, or any other perturbation source, the spectral vectors may lie outside of the true data simplex. The application of the a MV algorithm would lead to the dashed estimate, which is may be far from the original.

In order to get estimated endmembers closer to original ones, MVSA/SISAL allows violations to the positivity constraint by weighting these violations with a soft constraint given by the hinge loss function ($\text{hinge}(x) = 0$ if $x \geq 0$ and $-x$ if $x < 0$). After identifying the signal subspace and applying an affine projection, MVSA/SISAL aims at solving the following optimization problem:

$$\begin{aligned} \hat{\mathbf{Q}} &= \arg \max_{\mathbf{Q}} \log(|\det(\mathbf{Q})|) - \lambda \mathbf{1}_p^T \text{hinge}(\mathbf{Q}\mathbf{Y})\mathbf{1}_n \\ \text{s.t.:} \quad & \mathbf{1}_p^T \mathbf{Q} = \mathbf{q}_m, \end{aligned} \quad (6)$$

where $\mathbf{Q} \equiv \mathbf{M}^{-1}$, $\mathbf{1}_p$ and $\mathbf{1}_n$ are column vectors of ones of sizes p and n (n stands for the number of spectral vectors), respectively, $\mathbf{q}_m \equiv \mathbf{1}_p^T \mathbf{Y}_p^{-1}$ with \mathbf{Y}_p being any set of linearly independent spectral vectors taken from the data set $\mathbf{Y} \equiv [\mathbf{y}_1, \dots, \mathbf{y}_n]$, and λ is a regularization parameter.

We make the following two remarks: a) maximizing $\log(|\det(\mathbf{Q})|)$ is equivalent to minimizing $V(\mathbf{M}_0)$; b) the term $-\lambda \mathbf{1}_p^T \text{hinge}(\mathbf{Q}\mathbf{Y})\mathbf{1}_n$ weights the positivity violations. As λ approaches to infinity, we recover the hard constraint.

The minimum volume enclosing simplex (MVES)⁷⁴ aims at solving the optimization problem (6) with $\lambda = \infty$, *i.e.*, for hard positivity constraints. MVES implements a cyclic minimization using linear programs (LPs). Although the optimization problem (6) is nonconvex, it is proved in (6) that the existence of pure pixels is a sufficient condition for MVES to perfectly identify the true endmembers.

The minimum volume transform-nonnegative matrix factorization (MVC-NMF)⁷⁵ solves the following optimization problem applied to the original data set, *i.e.*, without dimensionality reduction ($\mathbf{Y} \in \mathbb{R}^{B \times n}$ and $\mathbf{M} \in \mathbb{R}^{B \times p}$)

$$\begin{aligned} \hat{\mathbf{M}} &= \arg \min_{\mathbf{M} \in \mathbb{R}^{B \times p}} \frac{1}{2} \|\mathbf{Y} - \mathbf{M}\mathbf{S}\|_F^2 + \lambda V^2(\mathbf{M}) \\ \text{s.t.:} \quad & \mathbf{M} \geq 0, \mathbf{S} \geq 0, \mathbf{1}^T \mathbf{S} = \mathbf{1}_n^T, \end{aligned} \quad (7)$$

where $\mathbf{S} \equiv [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n] \in \mathbb{R}^{p \times n}$ is a matrix containing the fractional abundances $\|\mathbf{A}\|_F^2 \equiv \text{tr}(\mathbf{A}^T \mathbf{A})$ is the Frobenius norm of matrix \mathbf{A} and λ is a regularization parameter. The optimization (7) minimizes a two term objective function, where the term $(1/2)\|\mathbf{Y} - \mathbf{M}\mathbf{S}\|_F^2$ measures the approximation error and the term $V^2(\mathbf{M})$ measures the square of the volume of the simplex defined by the columns of \mathbf{M} . The regularization parameter λ controls the tradeoff between the reconstruction errors and simplex volumes. MVC-NMF implements a sequence of alternate minimizations with respect to \mathbf{S} (quadratic programming problem) and with respect to \mathbf{M} (nonconvex programming problem). The major difference between MVC-NMF and MVSA/SISAL algorithms is that the latter allows violations of the positivity constraint, whereas the former does not.

The iterative constrained endmembers (ICE) algorithm⁷⁶ aims at solving an optimization problem similar to that of MVC-NMF, where the volume of the simplex is replaced by an approximation that is more manageable: the sum of squared distances between all the simplex vertices. The algorithm implements a sequence of alternate minimizations with respect to \mathbf{S} and with respect to \mathbf{M} . An advantage of ICE over MVC-NMF is that in the former both minimizations are quadratic programming problems, whereas in the MVC-NMF the optimization with respect to \mathbf{M} is a nonconvex problem. The sparsity-promoting ICE (SPICE)⁷⁷ is an extension of the ICE algorithm that incorporates sparsity-promoting priors aiming at finding the correct number of endmembers.

ORASIS³⁵ is a hyperspectral framework developed by the Naval Research Laboratory consisting of several algorithms organized in six modules: exemplar selector, adaptive learner, demixer, knowledge base/spectral library, and spatial postprocessor. The first step consists in flat fielding the spectra. Next, the exemplar selection

module is used to select spectral vectors that best represent the smaller convex cone containing the data. The other pixels are rejected when the spectral angle distance is less than a given threshold. The procedure finds the basis for a subspace of a lower dimension using a modified Gram-Schmidt orthogonalization. The selected vectors are then projected onto this subspace, and a simplex is found by a MV procedure.

To finish this section, we refer the convex cone analysis (CCA),⁷⁸ which finds the boundary points of the data convex cone (it does not apply affine projection), what is very close to MV concept. CCA starts by selecting the eigenvectors corresponding to the largest eigenvalues. These eigenvectors are then used as a basis to form linear combinations that have only nonnegative elements, thus belonging to a convex cone. The vertices of the convex cone correspond to spectral vectors contains as many zero elements as the number of eigenvectors minus one.

4.3 Statistical methods

When the spectral mixtures are highly mixed, the geometrical based methods yields poor results because there are not enough spectral vectors in the simplex facets. In these cases, the statistical methods are a powerful alternative, which, usually, comes with price: higher computational complexity, when compared with the geometrical based approaches.

Under the statistical framework, spectral unmixing is formulated as a statistical inference problem. Adopting a Bayesian perspective, the inference engine is the posterior density of the random objects to be estimated. Since our unknowns are the mixing matrix \mathbf{M} and the abundance fractions \mathbf{S} , we have, by using the Bayes law,

$$p_{M,S|Y}(\mathbf{M}, \mathbf{S}|\mathbf{Y}) = p_{Y|M,S}(\mathbf{Y}|\mathbf{M}, \mathbf{S})p_M(\mathbf{M})p_S(\mathbf{S})/p_Y(\mathbf{Y}), \quad (8)$$

where the notation p_A and $p_{A|B}$ stands for the probability density function (pdf) of \mathbf{A} and of \mathbf{A} given \mathbf{B} , respectively, and it was assumed that \mathbf{M} and \mathbf{S} are statistically independent. Popular estimators are⁷⁹ the *maximum a posteriori probability* (MAP) given by

$$(\widehat{\mathbf{M}}, \widehat{\mathbf{S}}) \equiv \arg \max_{\mathbf{M}, \mathbf{S}} p_{M,S|Y}(\mathbf{M}, \mathbf{S}|\mathbf{Y}) \quad (9)$$

$$= \arg \min -\log p_{Y|M,S}(\mathbf{Y}|\mathbf{M}, \mathbf{S}) - \log p_M(\mathbf{M}) - \log p_S(\mathbf{S}) \quad (10)$$

and the *minimum mean square error* (MMSE) given by $(\widehat{\mathbf{M}}, \widehat{\mathbf{S}}) \equiv \mathbb{E}[\mathbf{M}, \mathbf{S}|\mathbf{Y}]$.

Under the linear observation model (5) and assuming the noise random vector \mathbf{w} is Gaussian with covariance matrix $\sigma^2 \mathbf{I}$, then, we have $-\log p_{Y|M,S}(\mathbf{Y}|\mathbf{M}, \mathbf{S}) = (1/(2\sigma^2))\|\mathbf{Y} - \mathbf{M}\mathbf{S}\|_F^2 + \text{const}$. It is then clear that ICE/SPICE and MVC-NMF algorithms, which have been classified as geometrical, can also be classified as statistical, yielding MAP estimates. In all these algorithms, the estimates are obtained by minimizing a two-term objective function: one term plays the role $-\log p_{Y|M,S}(\mathbf{Y}|\mathbf{M}, \mathbf{S})$ and the other plays the role of $-\log p_M(\mathbf{M}) - \log p_S(\mathbf{S})$. The parameter λ , termed the regularization parameter, set the relative weight between the two terms.

The work⁸⁰ introduces a Bayesian approach where the linear observation model observation model (5) with zero-mean white Gaussian noise of covariance $\sigma^2 \mathbf{I}$ is assumed, the fractional abundances are uniformly distributed on the simplex, and the prior on \mathbf{M} is an autoregressive model.

The Bayesian approaches introduced in^{81–83} have all the same flavor: The posterior density is build under a hierarchical Bayesian model, which assumes conjugate prior distributions for some unknown parameters, accounts for nonnegativity and full additivity constraints. A Gibbs sampler is then used to overcome the complexity of evaluating the resulting posterior distribution and approximating the MMSE estimate.

In many application, the additive noise term may neglected because the noise power is very small. When that is not the case but the signal subspace has much lower dimension than the number of bands, then, as seen in Section 3.1, the projection onto the signal subspace largely reduces the noise power. Under this circumstances, and assuming that $\mathbf{M} \in \mathbb{R}^{p \times p}$ is invertible and the observed spectral vectors are independent, then we can write

$$p_{Y|M}(\mathbf{Y}|\mathbf{M}) = \left(\prod_{i=1}^n p_{\alpha}(\mathbf{M}^{-1} \mathbf{y}_i) \right) |\det(\mathbf{M}^{-1})|^n,$$

where p_α is the fractional abundance pdf, and compute the em maximum likelihood (ML) estimate of $\mathbf{W} \equiv \mathbf{M}^{-1}$. This is precisely the ICA line of attack, under the assumption that the fractional abundances are independent, *i.e.*, $p_\alpha = \prod_{k=1}^P p_{\alpha_k}$. The fact that this assumption is not valid in hyperspectral applications⁶³ has promoted research on suitable statistical models for hyperspectral fractional abundances and in effective algorithms to infer the mixing matrices. The is the case with DECA;^{84,85} the abundance fractions are modeled as mixtures of Dirichlet densities, thus, automatically enforcing the constraints on abundance fractions imposed by the acquisition process, namely nonnegativity and constant sum. A cyclic minimization algorithm is developed where: 1) the number of Dirichlet modes is inferred based on the minimum description length (MDL) principle; 2) a generalized expectation maximization (GEM) algorithm is derived to infer the model parameters; 3) a sequence of augmented Lagrangian based optimizations are used to compute the signatures of the endmembers.

Finally, some methods have targeted the integration of spectral and spatial information in statistical fashion. For instance, extended morphological operations have been used as a baseline to develop an automatic morphological endmember extraction (AMEE) algorithm⁸⁶ for spatial-spectral endmember extraction. Also, spatial averaging of spectrally similar endmember candidates found via singular value decomposition (SVD) was used in the spatial spectral endmember extraction (SSEE) algorithm.⁸⁷ Recently, a spatial preprocessing (SPP) algorithm⁸⁸ has been proposed which estimates, for each pixel vector in the scene, a spatially-derived factor that is used to weight the importance of the spectral information associated to each pixel in terms of its spatial context. The SPP is intended as a preprocessing module that can be used in combination with an existing spectral-based endmember extraction algorithm.

4.4 Sparse regression based unmixing

The spectral unmixing problem have recently been approached in a semi-supervised fashion, by assuming that the observed image signatures can be expressed in the form of linear combinations of a number of pure spectral signatures known in advance^{89–91} (*e.g.*, spectra collected on the ground by a field spectro-radiometer). Unmixing then amounts to finding the optimal subset of signatures in a (potentially very large) spectral library that can best model each mixed pixel in the scene. In practice, this is a combinatorial problem which calls for efficient linear sparse regression techniques based on sparsity-inducing regularizers, since the number of endmembers participating in a mixed pixel is usually very small compared with the (ever-growing) dimensionality and availability of spectral libraries.³ Linear sparse regression is an area of very active research with strong links to compressed sensing, basis pursuit, basis pursuit denoising, and matching pursuit.⁹²

Let us assume then that the spectral endmembers used to solve the mixture problem are no longer extracted nor generated using the original hyperspectral data as input, but instead selected from a library $\mathbf{A} \in \mathbb{R}^{B \times m}$ containing a large number of spectral samples, say m , available a priori. In this case, unmixing amounts to finding the optimal subset of samples in the library that can best model each mixed pixel in the scene. Usually, we have $m > B$ and therefore the linear problem at hands is under-determined. With the aforementioned assumptions in mind, let $\mathbf{x} \in \mathbb{R}^m$ denote the fractional abundance vector with regards to the library \mathbf{A} . With these definitions in place, we can now write our sparse regression problem as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{Ax}\|_2 \leq \delta, \mathbf{x} \geq \mathbf{0}, \quad (11)$$

where $\|\mathbf{x}\|_0$ denotes the number of non-zero components of \mathbf{x} and $\delta \geq 0$ is the error tolerance due to noise and modeling errors.

Problem (11) is NP-hard⁹³ and therefore there is no hope in solving it in a straightforward way. Greedy algorithms, such as the orthogonal basis pursuit (OMP),⁹⁴ basis pursuit (BP), and basis pursuit (BPDN)⁹⁵ are three alternative approaches to compute the sparsest solution. The BP and BPDN replaces the ℓ_0 norm with the ℓ_1 norm. In the case of BPDN, we have

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{y} - \mathbf{Ax}\|_2 \leq \delta, \mathbf{x} \geq \mathbf{0}, \quad (12)$$

Contrarily to problem (11), problem (12) is convex and can be solved efficiently.^{96,97} What is, perhaps, totally unexpected is that, in given circumstances related with the coherence of among the columns of matrix

A, problems (11) and (12) have the same solution. This issues are studied in detail in,⁸⁹ where it is concluded that a) the hyperspectral libraries are highly correlated, what is not good for sparse regression, and b) the high level of sparsity of the fractional abundances mitigates the correlation to ensure that (12) yields useful results.

5. SUMMARY

In this paper, we have presented an overview of the state of the art in hyperspectral unmixing. Specifically, four main aspects have been discussed: **i**) mixing models, **ii**) signal subspace identification, **iii**) geometrical-based spectral unmixing, **iv**) statistical-based spectral unmixing, and **v**) sparse regression-based unmixing. In each topic, we describe what physical or mathematical problems are involved and summarize state-of-the-art algorithms to address these problems. The compendium of techniques presented in this work reflects the increasing sophistication of a field that is rapidly maturing at the intersection of many different disciplines, including signal and image processing, physical modeling, linear algebra and computing developments. In the future we will conduct an exhaustive experimental comparison of the techniques discussed in this work.

REFERENCES

1. G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 12–16, 2002.
2. D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 17–28, 2002.
3. N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 44–57, 2002.
4. M. O. Smith, P. E. Johnson, and J. B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal component analysis," in *Proc. of the 15th Lunar and Planetary Sci. Conf., Part 2, Geophys. Res.*, vol. 90, 1985, pp. C797–C804.
5. A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fisher, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California," in *Proc. of the 2nd AVIRIS Workshop, R. O. Green, Ed., Jpl Publ.*, vol. 90-54, 1990, pp. 243–270.
6. G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, and W. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of the Environment*, vol. 44, pp. 127–143, 1993.
7. J. B. Adams and M. O. Smith, "A new analysis of rock and soil types at the viking lander 1 site," *J. of Geophysical Research*, vol. 91, no. B8, pp. 8098–8112, 1986.
8. S. Liangrocapart and M. Petrou, "Mixed pixels classification," in *Proc. of the SPIE Conference on Image and Signal Processing for Remote Sensing IV*, vol. 3500, 1998, pp. 72–83.
9. R. B. Singer and T. B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," in *Proc. of the 10th Lunar and Planetary Sci. Conf.*, 1979, pp. 1835–1848.
10. B. Hapke, "Bidirection reflectance spectroscopy. I. theory," *J. of Geophysical Research*, vol. 86, pp. 3039–3054, 1981.
11. R. N. Clark and T. L. Roush, "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications," *J. of Geophysical Research*, vol. 89, no. B7, pp. 6329–6340, 1984.
12. C. C. Borel and S. A. Gerstl, "Nonlinear spectral mixing models for vegetative and soils surface," *Remote Sensing of the Environment*, vol. 47, no. 2, pp. 403–416, 1994.
13. S. Chandrasekhar, *Radiative Transfer*. New York: Dover, 1960.
14. A. Plaza, G. Martin, J. Plaza, M. Zortea, and S. Sanchez, "Recent developments in spectral unmixing and endmember extraction", in *Optical Remote Sensing - Advances in Signal Processing and Exploitation*, S. Prasad, L. Bruce and J. Chanussot, Eds., Springer-Verlag, 2010 (in press).
15. N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An algorithm taxonomy for hyperspectral unmixing," in *Proc. of the SPIE AeroSense Conference on Algorithms for Multispectral and Hyperspectral Imagery VI*, vol. 4049, 2000, pp. 42–63.
16. C. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, June 2006.
17. S. S. Shen and E. M. Bassett, "Information-theory-based band selection and utility evaluation for reflective spectral systems," in *Proc. of the SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, vol. 4725, 2002, pp. 18–29.
18. I. T. Jolliffe, *Principal Component Analysis*. New York: Spriger Verlag, 1986.
19. L. L. Scharf, *Statistical Signal Processing, Detection Estimation and Time Series Analysis*. Addison-Wesley Pub. Comp., 1991.
20. 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 Sensing*, vol. 26, no. 1, pp. 65–74, 1988.
21. J. B. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 3, pp. 295–304, 1990.
22. J. Bruske and G. Sommer, "Intrinsic dimensionality estimation with optimally topologic preserving maps," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 5, pp. 572–575, 1998.
23. P. Demartines and J. Héroult, "Curvilinear component analysis : A self-organizing neural network for nonlinear mapping of data sets," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 148–154, 1997.
24. M. Lennon, G. Mercier, M. Mouchot, and L. Hubert-Moy, "Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images," in *Proc. of the SPIE Symp. on Remote Sensing Conference on Image and Signal Processing for Remote Sensing VII*, vol. 4541, 2001, pp. 157–169.
25. C. Bachmann, T. Ainsworth, and R. Fusina, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 10, pp. 2786–2803, 2006.

26. —, “Exploiting manifold geometry in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 3, pp. 441–454, 2005.
27. C. Yangchi, M. Crawford, and J. Ghosh, “Applying nonlinear manifold learning to hyperspectral data for land cover classification,” in *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, vol. 6, 2005, pp. 4311–4314.
28. D. Gillis, J. Bowles, G. M. Lamela, W. J. Rhea, C. M. Bachmann, M. Montes, and T. Ainsworth, “Manifold learning techniques for the analysis of hyperspectral ocean data,” in *Proc. of the SPIE conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, S. S. Shen and P. E. Lewis, Eds., vol. 5806, 2005, pp. 342–351.
29. J. Wang and C.-I. Chang, “Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis,” *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 6, pp. 1586–1600, 2006.
30. M. Lennon, M. Mouchot, G. Mercier, and L. Hubert-Moy, “Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images,” in *Proc. of the IEEE Int. Geosci. and Remote Sensing Symp.*, 2001.
31. A. Ifarraguerri and C.-I. Chang, “Unsupervised hyperspectral image analysis with projection pursuit,” *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 6, pp. 127–143, 2000.
32. C. Bachmann and T. Donato, “An information theoretic comparison of projection pursuit and principal component features for classification of landsat tm imagery of central colorado,” *Int. J. Rem. Sens.*, vol. 21, no. 15, pp. 2927–2935(9), 2000.
33. H. Othman and S.-E. Qian, “Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage,” *IEEE Trans. Geosci. Remote Sensing*, vol. 44, no. 2, pp. 397–408, 2006.
34. S. Kaewpitt, J. L. Moigne, and T. El-Ghazawi, “Automatic reduction of hyperspectral imagery using wavelet spectral analysis,” *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 4, pp. 863–871, 2003.
35. J. H. Bowles, J. A. Antoniadis, M. M. Baumbach, J. M. Grossmann, D. Haas, P. J. Palmadesso, and J. Stracka, “Real-time analysis of hyperspectral data sets using NRL’s ORASIS algorithm,” in *Proc. of the SPIE Conference on Imaging Spectrometry III*, vol. 3118, 1997, pp. 38–45.
36. N. Keshava, “A survey of spectral unmixing algorithms,” *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 55–78, 2003.
37. G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
38. J. Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, pp. 465–471, 1978.
39. M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.
40. J. Harsanyi, W. Farrand, and C.-I. Chang, “Determining the number and identity of spectral endmembers: An integrated approach using neyman-pearson eigenthresholding and iterative constrained rms error minimization,” in *Proc. 9th Thematic Conf. Geologic Remote Sensing*, 1993.
41. H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.
42. C.-I. Chang and Q. Du, “Estimation of number of spectrally distinct signal sources in hyperspectral imagery,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004.
43. J. Bioucas-Dias and J. Nascimento, “Hyperspectral subspace identification,” in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46., no. 8, pp. 2435–2445, 2005.
44. C. Bateson, G. Asner, and C. Wessman, “Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis,” *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1083–1094, 2000.
45. F. Kruse, “Spectral identification of image endmembers determined from AVIRIS data,” in *Summaries of the VII JPL Airborne Earth Science Workshop*, 1998.
46. J. Boardman and F. Kruse, “Automated spectral analysis: a geological example using AVIRIS data, northern grapevine mountains, Nevada,” in *Proc. of the 10th Thematic Conference, Geologic Remote Sensing*, 1994.
47. G. Shaw and H. Burke, “Spectral imaging for remote sensing,” *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.
48. T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, “A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing,” *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4418–4432, 2009.
49. M. Parente and A. Plaza, “Survey of geometric and statistical unmixing algorithms for hyperspectral images”, in *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing-WHISPERS’10*, Reykjavik, Iceland, 2010.
50. A. Plaza, P. Martinez, R. Perez, and J. Plaza, “A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 650663, 2004.
51. Y. H. Hu, H. B. Lee, and F. L. Scarpace, “Optimal linear spectral unmixing,” *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 639–644, 1999.
52. M. Petrou and P. G. Foschi, “Confidence in linear spectral unmixing of single pixels,” *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 624–626, 1999.
53. J. J. Settle, “On the relationship between spectral unmixing and subspace projection,” *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 1045–1046, 1996.
54. A. S. Mazer, M. Martin, et al., “Image processing software for imaging spectrometry data analysis,” *Remote Sensing of the Environment*, vol. 24, no. 1, pp. 201–210, 1988.
55. R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, “Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm,” in *Summaries of the 3rd annu. JPL Airborne Geosci. Workshop, R. O. Green, Ed. Publ., 92-14*, vol. 1, 1992, pp. 147–149.
56. J. C. Harsanyi and C.-I. Chang, “Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach,” *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 4, pp. 779–785, 1994.
57. C. Chang, X. Zhao, M. L. G. Althouse, and J. J. Pan, “Least squares subspace projection approach to mixed pixel classification for hyperspectral images,” *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 3, pp. 898–912, 1998.
58. D. C. Heinz, C.-I. Chang, and M. L. G. Althouse, “Fully constrained least squares-based linear unmixing,” in *Proc. of the IEEE International Geoscience and Remote Sensing Symp.*, 1999, pp. 1401–1403.
59. P. Common, “Independent component analysis: A new concept,” *Signal Processing*, vol. 36, pp. 287–314, 1994.
60. J. Bayliss, J. A. Gualtieri, and R. Crompt, “Analysing hyperspectral data with independent component analysis,” in *Proc. of SPIE*, vol. 3240, 1997, pp. 133–143.
61. C. Chen and X. Zhang, “Independent component analysis for remote sensing study,” in *Proc. of the SPIE Symp. on Remote Sensing Conference on Image and Signal Processing for Remote Sensing V*, vol. 3871, 1999, pp. 150–158.

62. T. M. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," *Opt. Eng./SPIE*, vol. 39, no. 4, pp. 897–906, 2000.
63. J. Nascimento and J. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, 2005.
64. M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 99–109, 1994.
65. J. Boardman, "Automating spectral unmixing of AVIRIS data using convex geometry concepts," in *Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, JPL Pub. 93-26, AVIRIS Workshop.*, vol. 1, 1993, pp. 11–14.
66. J. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Summaries of the V JPL Airborne Earth Science Workshop*, vol. 1, 1995, pp. 23–26.
67. M. E. Winter, "N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. of the SPIE conference on Imaging Spectrometry V*, 1999, pp. 266–275.
68. , R.A Neville, K Staenz, T Szeredi, J Lefevbre, and P Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," in *International Airborne Remote Sensing Conference and Exhibition, 4 th/21 st Canadian Symposium on Remote Sensing*, Ottawa, Canada, 1999.
69. J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 8, pp. 898–910, 2005.
70. C.-I. Chang, C.-C. Wu, W. Liu, and Y.-C. Ouyang, "A New Growing Method for Simplex-Based Endmember Extraction Algorithm", in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44., no. 11, pp. 2804–2819, 2006.
71. J. Gruninger, A. Ratkowski, and M. Hoke, The sequential maximum angle convex cone (smacc) endmember model, *Proceedings of SPIE*, vol. 5425, 2004.
72. J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data", in *IEEE Geoscience and Remote Sensing Symposium- IGARSS'08*, Boston, 2008.
73. J. Bioucas-Dias, "Variable splitting augmented Lagrangian approach to linear spectral unmixing," in *First IEEE GRSS Workshop on Hyperspectral Image and Signal Processing-WHISPERS'2009*, Grenoble, France, 2000.
74. T. Chan, C. Chi, Y. Huang, and W. Ma, "Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4418–4432, 2009.
75. L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization" in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.
76. M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F Huntington, "ICE: a statistical approach to identifying endmembers in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, 2004.
77. A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 446–450, Jul. 2007.
78. A. Ifarraguerri and C.-I Chang, "Multispectral and hyperspectral image analysis with convex cones," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 756–770, 1999.
79. J. Bernardo and A. Smith. *Bayesian Theory*, John Wiley & Sons, 1994.
80. L. Parra, K.-R. Mueller, C. Spence, A. Ziehe, and P. Sajda, "Unmixing hyperspectral data" in *Advances in Neural Information Processing Systems (NIPS)*, Cambridge, MA: MIT Press, 2000, vol. 12, pp. 942–948.
81. M. Arngren, M. N. Schmidt, and J. larsen, "Bayesian nonnegative matrix factorization with volume prior for unmixing of hyperspectral images," in *Machine Learning for Signal Processing, IEEE Workshop on (MLSP)*, 2009.
82. N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Processing*, 2009.
83. S. Moussaoui, C. Carteretb, D. Briea, and A. Mohammad-Djafaric, "Bayesian analysis of spectral mixture data using markov chain monte carlo methods," *Chemosometrics and Intelligent Laboratory Systems*, vol. 81, no. 2, pp. 137–148, 2006.
84. J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis", in *IEEE International Geoscience and Remote sensing Symposium -IGARSS2007*, Barcelona, Spain, July 2007.
85. J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing algorithm based on mixtures of Dirichlet components", *IEEE Transactions on Geoscience and Remote Sensing*, 2010 (submitted).
86. A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations", in *IEEE Trans. Geosci. Remote Sensing*, vol. 40, pp. 2025–2041, 2002.
87. D. M. Rogge, B. Rivard, J. Zhang, A. Sanchez, J. Harris, and J. Feng, "Integration of spatial/spectral information for the improved extraction of endmembers," *Remote Sensing of Environment*, vol. 110, no. 3, pp. 287–303, 2007.
88. M. Zortea and A. Plaza, "Spatial preprocessing for endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 2679–2693, 2009.
89. M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data", *IEEE Transactions on Geoscience and Remote Sensing*, 2010 (submitted).
90. D. Rogge, B. Rivard, J. Zhang, and J. Feng, "Iterative spectral unmixing for optimizing per-pixel endmember sets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 12, pp. 3725–3736, 2006.
91. M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "On the Use of Spectral Libraries to Perform Sparse Unmixing of Hyperspectral Data", in *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'10)*, Reykjavik, Iceland, 2010.
92. R. Baraniuk, "Compressive sensing", in *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–, 2007.
93. B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.
94. Y. C. Pati, R. Rezahfar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems and Computers*, Los Alamitos, CA, USA, 2003.
95. S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit" *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.
96. J. Bioucas-Dias and M. Figueiredo, "Alternating Direction Algorithms for Constrained Sparse Regression: Application to Hyperspectral Unmixing", in *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS'10)*, Reykjavik, Iceland, 2010.
97. Z. Guo, T. Wittman, and S. Osher, "L1 Unmixing and its Application to Hyperspectral Image Enhancement," in *Proceedings of SPIE Conference on Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, Orlando, Florida, 2009.