5-2012

Assessing and Enabling Independent Component Analysis As A Hyperspectral Unmixing Approach

Matthew R. Stites
Utah State University

Recommended Citation

Stites, Matthew R., "Assessing and Enabling Independent Component Analysis As A Hyperspectral Unmixing Approach" (2012). *All Graduate Theses and Dissertations*. 1199.
https://digitalcommons.usu.edu/etd/1199

ASSESSING AND ENABLING INDEPENDENT COMPONENT ANALYSIS AS A HYPERSPECTRAL UNMIXING APPROACH

by

Matthew R. Stites

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

Approved:

Dr. Jacob H. Gunther Major Professor	Dr. Todd K. Moon Committee Member
Dr. Scott E. Budge Committee Member	Dr. Donald Cripps Committee Member
Dr. Gustavious P. Williams Committee Member	Dr. Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2012

Abstract

Assessing and Enabling Independent Component Analysis as a Hyperspectral Unmixing Approach

by

Matthew R. Stites, Doctor of Philosophy
Utah State University, 2012

Major Professor: Dr. Jacob H. Gunther
Department: Electrical and Computer Engineering

As a result of its capacity for material discrimination, hyperspectral imaging has been utilized for applications ranging from mining to agriculture to planetary exploration. One of the most common methods of exploiting hyperspectral images is spectral unmixing, which is used to discriminate and locate the various types of materials that are present in the scene. When this processing is done without the aid of a reference library of material spectra, the problem is called blind or unsupervised spectral unmixing. Independent component analysis (ICA) is a blind source separation approach that operates by finding outputs, called independent components, that are statistically independent. ICA has been applied to the unsupervised spectral unmixing problem, producing intriguing, if somewhat unsatisfying results. This dissatisfaction stems from the fact that independent components are subject to a scale ambiguity which must be resolved before they can be used effectively in the context of the spectral unmixing problem.

In this dissertation, ICA is explored as a spectral unmixing approach. Various processing steps that are common in many ICA algorithms are examined to assess their impact on spectral unmixing results. Synthetically-generated but physically-realistic data are used to allow the assessment to be quantitative rather than qualitative only. Additionally, two algorithms, class-based abundance rescaling (CBAR) and extended class-based abundance rescaling (CBAR-X), are introduced to enable accurate rescaling of independent components. Experimental results demonstrate the improved rescaling accuracy provided by the CBAR and CBAR-X algorithms, as well as the general viability of ICA as a spectral unmixing approach.

(123 pages)

Public Abstract

Assessing and Enabling Independent Component Analysis as a Hyperspectral Unmixing Approach

by

Matthew R. Stites, Doctor of Philosophy

Utah State University, 2012

Major Professor: Dr. Jacob H. Gunther
Department: Electrical and Computer Engineering

Perhaps the most common way to distinguish materials is by color. For example, this is typically how one determines, from some distance, whether a material on the ground is grass (green), soil (brown), or asphalt (black). To accomplish this, most digital cameras (along with the human eye) produce images that are comprised of three different colors, called spectral bands: red, green, and blue. The combination of these bands enables material discrimination. Working in the same way, but on a much larger scale, hyperspectral imaging sensors produce images that are comprised of hundreds of spectral bands. This combination of bands enables more accurate and sensitive discrimination of the materials in a scene.

One of the most common ways to make hyperspectral images useful is to perform spectral unmixing. This process can determine what types of materials are in the image as well as where those materials are located within the image. When this is done without access to some sort of reference library of material spectra (i.e. material colors), the processing is called blind or unsupervised spectral unmixing. One of many methods for performing blind spectral unmixing is independent component analysis (ICA). ICA is an unmixing approach that produces outputs, called independent components, that are statistically independent from one another. One problem associated with ICA in the context of the spectral unmixing problem is scale ambiguity. The problem arises because multiplication by a constant value does not affect the independence of two random variables. Scale ambiguity hinders interpretation of spectral unmixing results by preventing comparison of different materials (since they may be scaled differently).

In this dissertation, ICA is examined as a spectral unmixing approach. Various processing steps that are common in many ICA algorithms are assessed to determine their impact on spectral unmxing results. Synthetically-generated, but physically-realistic data are used to allow the assessment to be quantitative rather than qualitative only. Additionally, two algorithms, class-based abundance rescaling (CBAR) and extended class-based abundance rescaling (CBAR-X), are introduced to enable accurate rescaling of independent components. Experimental results demonstrate the improved rescaling accuracy provided by the CBAR and CBAR-X algorithms, as well as the general viability of ICA as a spectral unmixing approach.

For Kristin and our kids...

Acknowledgments

I always imagined the pursuit of a Ph.D. to be a very individual effort. It turns out it has been anything but. It would be impossible to recognize everyone that contributed to this work, but I feel obligated to at least try.

First, I would to thank my committee: Dr. Jake Gunther for always being upbeat, willing to listen, and an advocate for students; Dr. Todd Moon for providing sage advice and setting a standard of diligence; Dr. Gus Williams for introducing me to and helping me feel a part of a community that I was previously unaware of; Dr. Scott Budge for being willing to discuss a variety of topics, scholastic and otherwise; and Dr. Don Cripps for always being interested in my progress and for stepping in on short notice.

I would also like to thank the office staff of the Electrical and Computer Engineering department: Kathy, Tricia, and Mary Lee for taking care of so much behind the scenes. Thank you for doing your jobs so well; it has made my job that much easier.

This work would not have been possible without financial assistance. I gratefully acknowledge the generous fellowship provided by the Space Dynamics Laboratory, as well as funding from the Department of Energy.

For mentoring me through an internship experience that was more interesting and enjoyable than I ever could have envisioned, I would like to thank Dr. Kevin Mitchell at Los Alamos National Laboratory.

On a more personal note, I would like to thank my extended family and friends. They have provided encouragement, support, distraction (the good kind), motivation, and even occasional IT assistance. I especially appreciate the concern and unique support of my grandparents, Ray and Sue Tanner.

Finally—and most importantly—I would like to thank my wife, Kristin, and kids, Jack, Ava, and Clara, for their love, patience, and constant eagerness to help in any way that they can. Every day they helped me move forward, whether they realized it or not. They

have made a number of sacrifices so that I could pursue this endeavor. I could not imagine doing this without them.

Matthew R. Stites

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Motivation

Hyperspectral imaging is a passive remote sensing technique which collects images in a large number of electromagnetic wavelengths, referred to as spectral bands. An intuitive way of visualizing hyperspectral images is as a stack of hundreds of images, producing a three-dimensional "cube" of data. Two of the cube dimensions represent the usual spatial dimensions encountered in most imaging methods. The third dimension represents the spectral content. Thus, for every pixel in a hyperspectral image a spectral profile—or simply spectrum—is obtained. This concept is illustrated in Fig. 1.1.

A distinguishing feature of hyperspectral images is their large quantity of spectral bands which typically number in the hundreds. These bands tend to be narrow (∼10 nm) and closely spaced, providing a high-resolution spectrum for each pixel in an image. This is important because the way in which matter reflects or emits electromagnetic radiation is dependent on, among other things, its chemical and physical composition. Thus, a spectral profile can be indicative of a specific material. This makes hyperspectral images especially useful for applications that require discrimination of the various materials in a scene.

Utilization of hyperspectral image data can be found in a broad variety of disciplines. Hyperspectral data have been employed for geological characterization of the surface of the Earth [1–4] as well as the surface of Mars [5,6]. They are also used to assist in "precision-crop management" [7–9] and ecological management efforts [10–12]. Targets such as land mines and camouflage nets can be located using hyperspectral image data [13, 14]. Another field employing hyperspectral imaging on a slightly different scale is food safety and inspection [15, 16].

One of the most common methods of extracting information from hyperspectral images is hyperspectral unmixing, more frequently called spectral unmixing. Although it can be

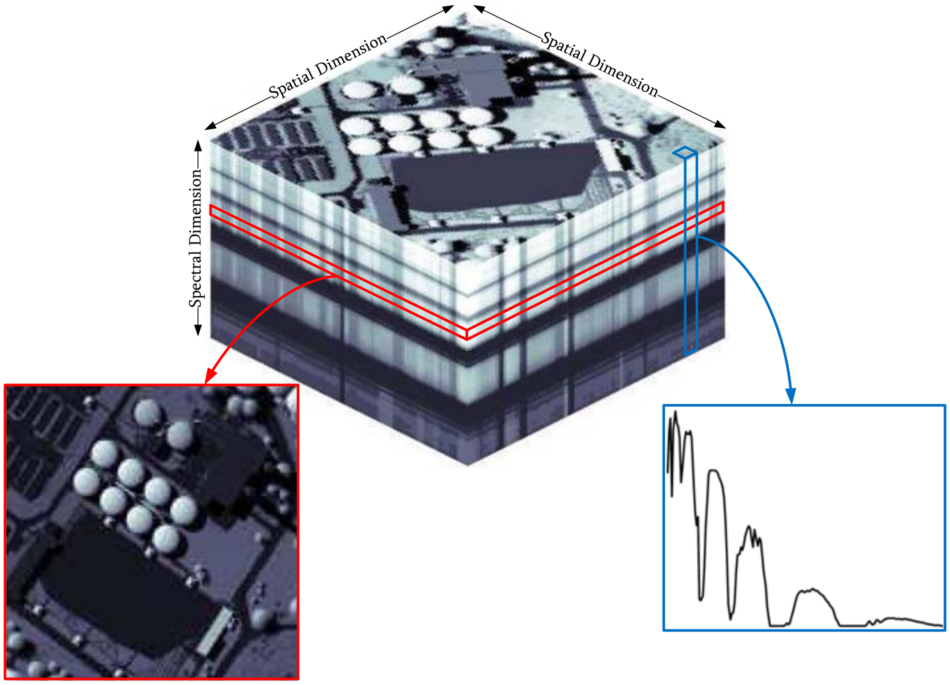


Fig. 1.1: The concept of a hyperspectral data cube.

performed in a variety of ways, spectral unmixing ultimately produces two fundamental results. The first is a set of spectra, referred to as endmembers. An endmember is a spectrum that is representative of a class of materials in a scene. How endmembers are defined (associated with materials) depends on the data and the application. Material classes can be broad and limited to those materials that appear frequently in the scene, resulting in a short list of endmembers, such as vegetation, soil, rock, and shade. Alternatively, a narrow, one-to-one association can be made resulting in an endmember for each and every discernible material in the scene. Such an association is assumed in the remainder of this document. Because of this, the terms endmember and material are used somewhat interchangeably. It is important to remember, however, that in other applications endmembers and materials are not necessarily the same thing.

The second spectral unmixing result is a set of images called abundance maps. An abundance map is produced for each endmember in the scene, and indicates the degree to which that endmember occupies each pixel. In order to be physically meaningful, abundance map values are typically required to fall between zero (endmember not present the pixel) and one (endmember fills the pixel).

There are two general approaches to spectral unmixing. The first is to unmix the data through some process of comparison with a known library of reference spectra. This is referred to as supervised spectral unmixing. The second approach, known as unsupervised (or blind) spectral unmixing, requires the data to be unmixed without any access to reference spectra. The focus of this dissertation is on the unsupervised problem. There are a number of motivations for taking an unsupervised approach. First, and perhaps most obvious, is that access to a spectral library may not be available. Second, a spectral library may be incomplete, leading to erroneous or misleading results. Third, the type of measurement made by a hyperspectral sensor to produce a hyperspectral image may be different from the type of measurement made to produce a reference spectrum. This necessitates some type of additional processing before any type of comparison or matching can be performed. Finally, spectral libraries tend to contain spectra of ideal or pristine materials obtained under laboratory conditions. These may not be representative of real-world materials and/or may not adequately describe the spectral variability of materials observed in the real world which can result from surface contamination, oxidation, and bleaching, among other things [17].

This dissertation explores the blind hyperspectral unmixing problem, focusing specifically on one proposed unmixing approach called independent component analysis (ICA). This exploration includes both assessing and enabling ICA. A quantitative assessment is performed to determine whether ICA can be reasonably applied to the spectral unmixing problem. The enabling component of this work consists of developing algorithms to overcome some of the common criticisms levied against ICA in the context of spectral unmixing. The specific contributions of this dissertation include:

- A thorough and quantitative assessment of the impact of dimension reduction using principal component analysis (PCA) on spectral unmixing results (Chapter 4);

- A thorough and quantitative assessment of the impact of orthogonalization on spectral unmixing results (Chapter 4);
- The identification of a “splitting” behavior frequently exhibited by ICA wherein distinct regions from a single abundance map appear in separate ICA outputs (Chapter 4);
- A quantitative assessment of ICA as a hyperspectral unmixing approach that considers materials with broad spatial coverage, as well as those with limited spatial coverage (Chapter 4);
- The development and assessment of the class-based abundance rescaling (CBAR) algorithm which is used to mitigate the problem of scale ambiguity associated with ICA and significantly out-performs existing approaches (Chapter 5);
- The development and assessment of the extended CBAR (CBAR-X) algorithm which is capable of retaining abundance information that would be lost by other rescaling methods (Chapter 6).

All of these results combine to show that ICA—aided by rescaling algorithms—can do a reasonably good job of unmixing those materials in a scene with limited spatial coverage. This behavior is complementary to most existing spectral unmixing approaches which are designed to extract the dominant materials.

The remainder of this dissertation is outlined as follows. First, two important models are presented in Chapter 2. These models provide valuable insight into spectral unmixing algorithms and also the relationship between material properties and measured hyperspectral data. Additionally, many of the important terms and quantities associated with hyperspectral imaging are defined. Chapter 3 provides a survey of blind hyperspectral unmixing approaches. This survey does not cover every single approach in the published literature since such a survey would be unbearably long and almost immediately out of date. Instead, various types of existing unmixing approaches are organized and characterized based on the general assumptions underlying them, providing a “map” of unmixing approaches. In

Chapters 4–6 the dissertation focus is narrowed to the more specific problem of using ICA as a spectral unmixing approach. These three chapters are presented as three independently publishable papers, resulting in some minor repetition of content. Chapter 4 introduces ICA and some of the processing steps common to many ICA algorithms. A number of experiments are described which allow for quantification of the effect of ICA processing steps on unmixing performance. The problem of rescaling ICA results so that they are physically meaningful as abundances is addressed in Chapters 5 and 6. A rescaling algorithm, called class-based abundance rescaling (CBAR), is developed in Chapter 5 and compared against existing approaches. This algorithm is extended in Chapter 6 to address certain circumstances where abundance information can be lost during CBAR rescaling. Finally, Chapter 7 provides some summarizing remarks and potential directions for future research.

References

[1] F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1388–1400, June 2003.

[2] R. G. Resmini, M. E. Kappus, W. S. Aldrich, J. C. Harsanyi, and M. Anderson, "Mineral mapping with HYperspectral Digital Imagery Collection Experiment (HYDICE) sensor data at Cuprite, Nevada, U.S.A," *International Journal of Remote Sensing*, vol. 18, no. 7, pp. 1553–1570, May 1997.

[3] P. Bierwirth, D. Huston, and R. Blewett, "Hyperspectral mapping of mineral assemblages associated with gold mineralization in the central Pilbara, Western Australia," *Economic Geology*, vol. 97, no. 4, pp. 819–826, 2002.

[4] B. Martini, E. Silver, W. Pickles, and P. Cocks, "Hyperspectral mineral mapping in support of geothermal exploration: Examples from Long Valley Caldera, CA and Dixie Valley, NV, USA," *Geothermal Resources Council Transactions*, vol. 27, pp. 657–662, 2003.

[5] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10-12, pp. 2194–2208, June 2008.

[6] A. Gendrin, N. Mangold, J.-P. Bibring, Y. Langevin, B. Gondet, F. Poulet, G. Bonello, C. Quantin, J. Mustard, R. Arvidson, and S. LeMouelic, “Sulfates in Martian layered terrains: The OMEGA/Mars Express view,” *Science*, vol. 307, no. 5715, pp. 1587–1591, 2005.

[7] M. Moran, Y. Inoue, and E. Barnes, “Opportunities and limitations for image-based remote sensing in precision crop management,” *Remote Sensing of Environment*, vol. 61, no. 3, pp. 319–346, 1997.

[8] D. W. Lamb and R. B. Brown, “Remote-sensing and mapping of weeds in crops,” *Journal of Agricultural Engineering Research*, vol. 78, no. 2, pp. 117–125, 2001.

[9] L. Serrano, I. Filella, and J. Penuelas, “Remote sensing of biomass and yield of winter wheat under different nitrogen supplies,” *Crop Science*, vol. 40, no. 3, pp. 723–731, 2000.

[10] M. Mirik, J. E. Norland, R. L. Crabtree, and M. E. Biondini, “Hyperspectral one-meter-resolution remote sensing in Yellowstone National Park, Wyoming: II. Biomass,” *Rangeland Ecology and Management*, vol. 58, no. 5, pp. 459–465, 2005.

[11] T. Schmid, M. Koch, J. Gumuzzio, and P. M. Mather, “A spectral library for a semi-arid wetland and its application to studies of wetland degradation using hyperspectral and multispectral data,” *International Journal of Remote Sensing*, vol. 25, pp. 2485–2496, July 2004.

[12] Y. Suzuki, H. Okamoto, K. Tanaka, W. Kato, and T. Kataoka, “Estimation of chemical composition of grass in meadows using hyperspectral imaging,” *Environment Control in Biology*, vol. 46, no. 2, pp. 129–137, 2008.

[13] B. H. P. Maathuis and J. L. van Genderen, “A review of satellite and airborne sensors for remote sensing based detection of minefields and landmines,” *International Journal of Remote Sensing*, vol. 25, pp. 5201–5245, Dec. 2004.

[14] C. M. Stellman, G. G. Hazel, F. Bucholtz, J. V. Michalowicz, A. Stocker, and W. Schaaf, “Real-time hyperspectral detection and cuing,” *Optical Engineering*, vol. 39, no. 7, pp. 1928–1935, 2000.

[15] M. S. Kim, A. Lefcourt, K. Chao, Y. Chen, I. Kim, and D. Chan, “Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part I. Application of visible and near-infrared reflectance imaging,” *Transactions of the American Society of Agricultrual and Biological Engineering*, vol. 45, no. 6, pp. 2027–2037, 2002.

[16] K. Chao, P. Mehl, and Y. Chen, “Use of hyper- and multi-spectral imaging for detection of chicken skin tumors,” *Applied Engineering in Agriculture*, vol. 18, no. 1, pp. 113–119, 2002.

[17] G. A. Shaw and H. K. Burke, “Spectral imaging for remote sensing,” *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.

Chapter 2

Models

Before describing spectral unmixing approaches, it is informative to present the data models that are assumed by these approaches. First, a model that describes sensor-reaching radiance is introduced. This model describes the relationship between observed endmember spectra and the underlying material properties that are often of interest. This makes it especially useful in identifying endmember spectra. The linear mixing model (LMM)—which motivates almost every spectral unmixing algorithm—is then presented. It describes how multiple endmembers located in a single pixel combine to produce the observed data for that pixel.

2.1 Model of Radiance Reaching the Sensor

In this section a formulation for the spectral radiance (a term for the quantity that is measured by hyperspectral sensors) arriving at the sensor is presented. This presentation summarizes a more thorough development by Schott. The reader is referred to his text [1] or those by Wolfe [2, 3] for further details.

There are a number of terms that need to be defined prior to developing a radiance model (including radiance). First are those that deal with scene illumination. These terms originate in radiometry, the field of measuring radiation. The rate of flow of light energy is called *radiant flux* or *power* and is given units of watts (W). The power incident onto a surface, normalized by the surface area, is called *irradiance* and has units of watts per square meter (W/m^2). Similarly, the power exiting a surface, normalized by the surface area is called *exitance* and also has units of watts per square meter (W/m^2). *Intensity* describes the power from a point source into a particular direction. It is defined as power, normalized by solid angle, and has units of watts per steradian (W/sr). Solid angle, with units of steradian, is a two-dimensional analog of angle and can be thought of as the projection of an area onto the unit sphere. The most useful of radiometric terms is *radiance*. It is defined as the power, normalized both by surface area and solid angle, and has units of watts per square meter per steradian ($\text{W/m}^2\text{/sr}$). Radiance is useful because it can describe power per unit solid angle from, through, or onto a surface. A summary of these quantities as well as the symbols commonly used to represent them is given in Table 2.1.

Normalizing any of these quantities by the wavelength of the light yields a spectral, or wavelength dependent, quantity. For example, normalizing radiance by wavelength yields *spectral radiance*, with units of watts per square meter per micron per steradian ($\text{W/m}^2/\mu\text{m/sr}$). As stated earlier, spectral radiance is the quantity that is measured by hyperspectral sensors. A subscript λ is sometimes attached to a spectral quantity to underscore its wavelength dependence.

The remaining definitions needed to construct a radiance model center around the interactions of radiation and matter. All matter emits electromagnetic radiation. This radiation is associated with the energy inherent in matter due to temperature and is called thermal or self-emitted radiation. It was shown by Planck that the spectral exitance of an ideal emitter, called a blackbody, is given by

$$M_\lambda^{\text{BB}} = M_\lambda^{\text{BB}}(T) = \frac{2\pi hc^2}{\lambda^5\left(e^{\frac{hc}{\lambda kT}} - 1\right)}, \tag{2.1}$$

where h is the Planck constant (6.6256×10^{-34} joules $\cdot$ s), c is the speed of light in a vacuum (2.9979×10^8 m/s), k is the Boltzmann gas constant (1.38×10^{-23} joules/K), λ

Table 2.1: Common radiometric quantities.

Name	Symbol	Units
Radiant flux (power)	Φ	W
Irradiance	E	W/m^2
Exitance	M	W/m^2
Intensity	I	W/sr
Radiance	L	$\text{W/m}^2\text{/sr}$

is the wavelength in microns, and T is the temperature in Kelvin. The formula in (2.1) is often denoted $B(\lambda, T)$ or $B_\lambda(T)$ and referred to as the blackbody function. The dependence on wavelength can be eliminated by integrating over wavelength, so that the exitance of a blackbody is given by

$$M^{\mathrm{BB}} = M^{\mathrm{BB}}(T) = \int_\lambda \frac{2\pi hc^2}{\lambda^5 \left(e^{\frac{hc}{\lambda kT}} - 1\right)} d\lambda. \tag{2.2}$$

A fundamental quantity of matter that is related to self-emitted radiation is *emissivity* which describes how effectively an object emits radiation. It is defined as the ratio of an object's exitance, $M(T)$, to the exitance of an ideal emitter at the same temperature, $M^{\mathrm{BB}}(T)$, i.e.,

$$\varepsilon = \frac{M(T)}{M^{\mathrm{BB}}(T)}. \tag{2.3}$$

The definition of spectral emissivity, is identical with spectral exitances in place of exitances,

$$\varepsilon_\lambda = \frac{M_\lambda(T)}{M_\lambda^{\mathrm{BB}}(T)}. \tag{2.4}$$

There are three other fundamental quantities that describe how matter interacts with external or incident radiation. As with the quantities described above, each of these can be defined with or without a dependence on wavelength. For brevity, formal definitions are presented only for the case not dependent on wavelength. The associated spectral definition follows the by replacing the terms in each formula with their spectral counterparts, as shown for emissivity in (2.3) and (2.4).

Transmittance describes how effectively radiation propagates through a material and is defined as the ratio of the exitance on the back of an object, M_{back}, to the irradiance on the front, E_{front}, i.e.,

$$\tau = \frac{M_{\mathrm{back}}}{E_{\mathrm{front}}}. \tag{2.5}$$

Reflectance describes how effectively radiation is turned back into the region from whence it came and is defined as the ratio of the exitance on the front of an object, M_{front}, to the irradiance on the front, E_{front}, i.e.,

$$\rho = \frac{M_{\text{front}}}{E_{\text{front}}}. \tag{2.6}$$

Absorptance describes the ability of a material to convert incident radiation into another form of energy, such as thermal energy. It is defined as the ratio of the power per unit area converted to another form of energy, $M_{\text{converted}}$, to the irradiance on the front, E_{front}, i.e.,

$$\alpha = \frac{M_{\text{converted}}}{E_{\text{front}}}. \tag{2.7}$$

Each of these quantities is unitless and ranges from zero to one. Additionally, the conservation of energy requires that all incident radiation be transmitted, reflected, or absorbed, so that

$$M_{\text{back}} + M_{\text{front}} + M_{\text{converted}} = E_{\text{front}}, \tag{2.8}$$

and

$$\tau + \rho + \alpha = 1. \tag{2.9}$$

Emissivity is associated with the internal energy of matter, while transmittance, reflectance, and absorptance are associated with external energy incident upon matter. But, these quantities can be related, at least when an object is in thermodynamic equilibrium. In that case emissivity is equal to absorption [4], which yields the relationships

$$\alpha = \varepsilon, \tag{2.10}$$

and

$$\tau + \rho + \varepsilon = 1. \tag{2.11}$$

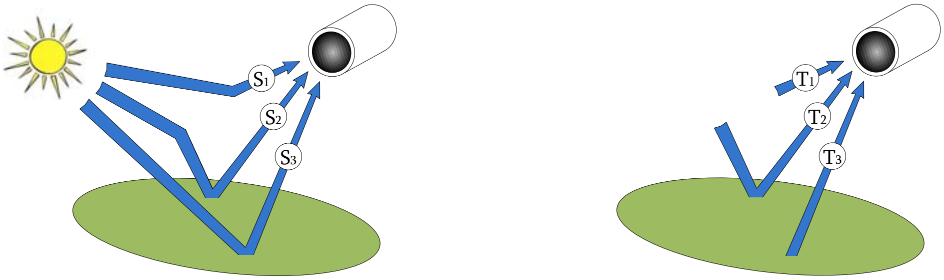
The concepts of illumination, radiometry, and the interaction of radiation with matter described above are sufficient to develop a model for the spectral radiance that reaches the sensor. This is done by considering individual radiance paths, i.e., all sources of radiation and all of the paths by which that radiation reaches the sensor. The total spectral radiance reaching the sensor is the summation of all of these paths. It can be useful to group paths by source. The two most common sources are the sun (assuming a daytime scene) and thermal emission from the objects in the scene. Thus, the spectral radiance reaching the sensor can be formulated as

$$L_{\text{sensor}} = L_S + L_T + L_{\text{other}}, \tag{2.12}$$

where L_S is total spectral radiance from solar radiance paths, L_T from thermal radiance paths, and L_{other} the contribution from radiance paths associated with other sources. Notice that although spectral radiance is being described, the λ subscript has been dropped. This will be the case for the remainder of this development to simplify notation. Only solar and thermal paths are considered here. They tend to be the most ubiquitous and provide the most significant contribution to the total radiance.

The specific solar and thermal paths examined here are shown in Fig. 2.1. The path S_1 is called solar upwelling radiance. It is the solar radiation that is scattered by the atmosphere into the sensor without ever reaching the ground (assuming the ground to be the object or surface of interest). The path S_2 is called reflected solar downwelling radiance. This is the solar radiation that is scattered by the atmosphere onto the ground, which is subsequently reflected back to the sensor. The last path, S_3, is the most obvious—reflected solar radiance. This is the direct solar radiation that is reflected by the ground to the sensor. The three thermal paths T_1, T_2, and T_3 are analogous to the three solar paths with the source of radiation being self-emitted radiation rather than solar radiation. So, T_1 and T_2 arise from the energy of the atmosphere being emitted into the sensor or emitted onto the ground and then reflected into the sensor, and T_3 is the energy emitted by the ground that reaches the sensor.

There are a number of assumptions that must be made before developing specific formulations for each of the radiance paths in Fig. 2.1. First, and foremost, it is assumed that the six paths shown are the only significant paths in the scene. Perhaps the most glaring omission here is the contribution of adjacency paths. These are paths that result



(a) (b)

Fig. 2.1: Significant radiance paths used to calculate the radiance reaching the sensor. (a) Solar radiance paths. (b) Thermal radiance paths.

when objects in the scene reflect solar energy and emit thermal energy onto the ground. This energy can then be reflected from the ground to the sensor. The contribution from adjacency paths may or may not be significant, depending on the geometry of the scene. The reason these paths are omitted is because they do not lend themselves to a general model. The contribution of these paths can vary greatly from scene to scene and from pixel to pixel within a scene. If intimate knowledge of scene geometry, i.e., size, shape, orientation, and location of every object in the scene, is available, then that knowledge can be utilized to determine the contribution of adjacency paths. But, that information is seldom available. Similarly, without knowledge of scene geometry it is difficult to determine what fraction of the sky is obscured by objects in the scene for a given pixel. So, again because of a lack of scene knowledge, a second assumption is that every pixel has a full view of the sky. This assumption precludes the presence of clouds in the scene. It also implies that every pixel lies in the same plane, i.e., the ground is flat. A third assumption is that the composition of the atmosphere is constant across the scene. Depending on the size of the scene, this tends to be a safe assumption. Similarly, it is assumed that for a given pixel the area on the ground imaged by that pixel is homogeneous in composition and temperature.

A final, simplifying assumption has to do with the reflectance (or emittance) of the materials in the scene. When reflectance was defined in (2.6), it was done by calculating the reflected exitance of the object, which does not depend on direction. But reflectance is a directional property (consider shining a flashlight at a mirror). The observed energy reflected by a material depends both on the angle of incident illumination as well as the angle of observation. There are a number of ways to characterize this behavior in laboratory and field measurements. To avoid the complexity associated with this directional variability, the reflectance is assumed to be uniform in all directions, i.e., Lambertian. Under this assumption, the orientation of the material with respect to the illumination source and sensor does not matter. For many natural materials this is a realistic assumption, as long as glancing angles are avoided.

With these simplifications and assumptions in place, the equations for each of the radiance paths are fairly straight forward. The equations for solar radiance paths are given in (2.13)–(2.16).

$$L_{S_1} = L_{u_S} \tag{2.13}$$

$$L_{S_2} = E_{d_S}\rho_g\tau_2 \tag{2.14}$$

$$L_{S_3} = E_S \cos\sigma\tau_1\rho_g\tau_2 \tag{2.15}$$

$$L_S = \left(E_S \cos\sigma\tau_1 + E_{d_S}\right)\rho_g\tau_2 + L_{u_S} \tag{2.16}$$

The spectral radiance of path S_1 is defined in (2.13) simply to be the upwelling solar radiance, L_{u_S}, a term that is dependent on solar irradiance and atmospheric transmission and scattering properties. The reflected solar downwelling radiance of path S_2 is defined in (2.14). The downwelling solar irradiance is denoted by E_{d_S} and, like the solar upwelling radiance, depends on both the solar irradiance and atmospheric transmission and scattering properties. The term τ_2 denotes the atmospheric transmittance between the ground and the sensor and depends on the composition of the atmosphere. The remaining term in this equation, ρ_g, is the ground reflectance normalized based on the assumption that the ground reflects uniformly in all directions. The final solar path, S_3, is represented in (2.15). The three new terms in this equation are E_S, the solar irradiance at the top of the atmosphere, σ, the angle of solar declination from zenith, and τ_1, the transmittance of the atmosphere from the top of the atmosphere to the ground. These results are combined in (2.16), giving a formula for the total spectral radiance arriving at the sensor due to the sun.

The equations for the thermal paths T_1, T_2, and T_3, are almost identical to those for the solar paths and are given in (2.17)–(2.19).

$$
\begin{aligned}
L_{T_1} &= L_{u_T} && (2.17)\\
L_{T_2} &= E_{d_T}\rho_g\tau_2 && (2.18)\\
L_{T_3} &= B(T_g)\varepsilon_g\tau_2 && (2.19)\\
L_T &= \left[B(T_g) + \left(E_{d_T} - B(T_g)\right)\rho_g\right]\tau_2 + L_{u_T} && (2.20)
\end{aligned}
$$

The solar upwelling radiance and downwelling irradiance terms are replaced by their thermal emission analogs, L_{u_T} and E_{d_T}, and the solar irradiance is replaced by the exitance of a blackbody at the temperature of the ground, T_g. Notice that this term is multiplied by the emissivity of the ground rather than the reflectance, which was the case for the corresponding solar path. Using the relationship $\varepsilon = 1 - \rho$ for an opaque object, i.e., $\tau = 0$, the combined spectral radiance reaching the sensor from self-emission can be written in terms of ρ_g as in (2.20).

The equations for total solar radiance and and total self-emitted radiance can be added to calculate the total spectral radiance reaching the sensor. But, in most cases this is unnecessary. Figure 2.2 shows the spectral transmittance of the atmosphere from the ground to a sensor at an elevation of 100 km. This was calculated using the MODTRAN atmospheric modeling code [5]. Displayed along the top of the figure is a common partitioning of this region of the electromagnetic spectrum. The acronyms for the four divisions stand for, from left to right, visible/near infrared (VNIR), short-wave infrared (SWIR), mid-wave infrared (MWIR), and long-wave infrared (LWIR). The two shortest wavelength regions are often grouped together as VNIR/SWIR. Each of these partitions contains at least one spectral region where the atmosphere is sufficiently transparent to collect hyperspectral data. Hyperspectral sensors are typically designed to collect data in only one of these three divisions (VNIR/SWIR, MWIR, or LWIR). In the VNIR/SWIR wavelengths the magnitude of solar radiance is so much larger than the magnitude of self-emitted radiance that the self-emitted radiance paths can be ignored for all but the most precise of applications. In the LWIR wavelengths, just the opposite is true. So, for VNIR/SWIR, $L \approx L_S$ and for LWIR, $L \approx L_T$. It is only in the MWIR region where the magnitudes of solar radiance and self-emitted radiance are comparable that all of the radiance paths must be included.

2.2 Linear Mixing Model

The model of radiance reaching the sensor developed in the previous section describes how to model a “pure” endmember. It does not, however describe endmember mixing. For this, the linear mixing model (LMM) is most frequently used. This model treats an observed mixed pixel as a linear combination of the constituent endmembers in the pixel. The weights of the linear combination are the fractional area of the pixel occupied by each endmember [6]. The validity of this model depends on the nature of the distribution of the materials within a pixel. If the distribution is such that radiation is most likely to interact with only one of the materials on its way to the sensor, i.e., the materials in a scene comprise a macroscopic mixture and are expressly separated one from another, then the linear model tends to be accurate [7,8].

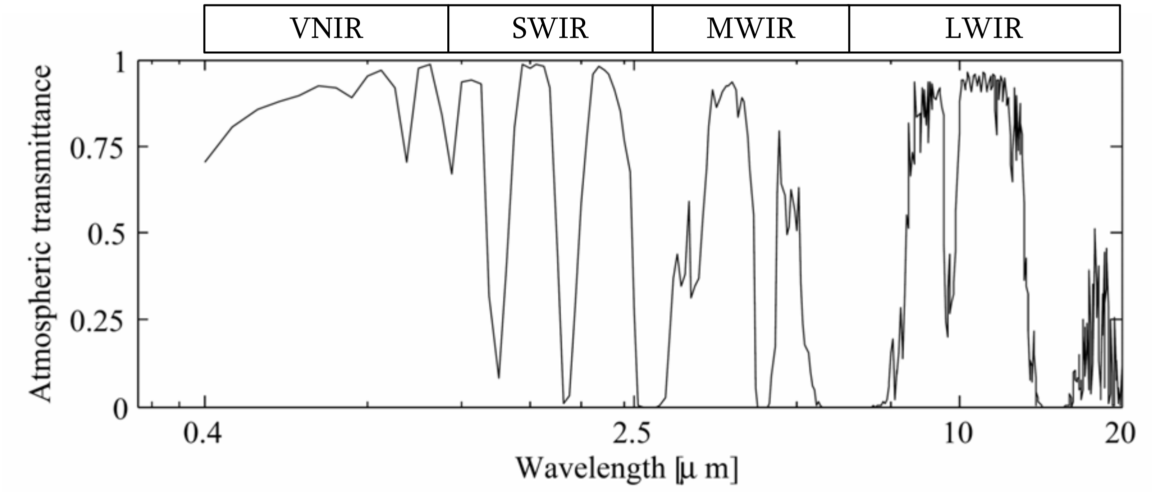


Fig. 2.2: Spectral transmittance of the atmosphere.

Representing spectra as column vectors, the linear mixing model is formulated as

$$\mathbf{x}(t) = \sum_{p=1}^{P} a_p(t)\mathbf{m}_p + \mathbf{n}(t), \tag{2.21}$$

where $\mathbf{x}(t)$ is the measured spectrum of pixel t, $a_p(t)$ is the fractional abundance of endmember p in pixel t, $\mathbf{m}_p$ is the spectrum associated with endmember p, and P is the number of endmembers in the scene. The last term, $\mathbf{n}(t)$, is an additive noise term that is frequently—but not necessarily—included in the model. The dimensions of both $\mathbf{x}(t)$ and $\mathbf{m}_p$ are $K \times 1$, where K is the number of spectral bands collected by the sensor. For simplicity, a single index, t is used to indicate the spatial location of a pixel within the image, e.g., using a raster-scan ordering. So, for a hyperspectral cube with spatial dimensions of $M \times N$, $t \in \{1, 2, \ldots, T = MN\}$.

To ensure the abundances are physically meaningful, the LMM almost always includes two constraints. The nonnegativity constraint given by

$$a_p(t) \geq 0 \quad \forall\ p, t, \tag{2.22}$$

prevents endmembers from filling a negative area of a pixel. The additivity constraint,

$$\sum_{p=1}^{P} a_p(t) = 1 \quad \forall\ t, \tag{2.23}$$

requires the combination of endmembers in a pixel to fill nothing more or less than the area of one pixel.

Equation (2.21) can be rewritten by assembling the individual endmember spectra into an endmember matrix, $\mathbf{M}$, and by stacking the abundance values into an abundance vector, $\mathbf{a}$, yielding

$$\mathbf{x}(t) = \mathbf{M}\mathbf{a}(t) + \mathbf{n}(t). \tag{2.24}$$

Similarly, the entire data cube can be represented by assembling the pixel vectors into matrices, so that

$$\mathbf{X} = \mathbf{M}\mathbf{A} + \mathbf{N}. \tag{2.25}$$

The two models presented in this chapter are sufficient to explore most hyperspectral unmixing approaches. The linear mixing model is used by the vast majority of unmixing algorithms to describe the relationship between endmember spectra and observed data. The sensor-reaching radiance model can be used to describe the relationship between library spectra, environmental conditions, and endmember spectra. With this background, the next chapter explores spectral unmixing algorithms in more detail.

References

[1] J. R. Schott, *Remote Sensing: the Image Chain Approach.* New York: Oxford University Press, 1997.

[2] W. L. Wolfe, *Introduction to Imaging Spectrometers.* Bellingham, WA: SPIE, 1997.

[3] W. L. Wolfe, *Introduction to Radiometry.* Bellingham, WA: SPIE, 1998.

[4] F. Grum and R. J. Becherer, *Optical Radiation Measurements. Volume 1: Radiometry.* New York: Academic Press, 1979.

[5] A. Berk, L. S. Bernstein, G. P. Anderson, P. K. Acharya, D. C. Robertson, J. H. Chetwynd, and S. M. Adler-Golden, “MODTRAN cloud and multiple scattering upgrades with application to AVIRIS,” *Remote Sensing of Environment*, vol. 65, no. 3, pp. 367–375, 1998.

[6] N. Keshava and J. Mustard, “Spectral unmixing,” *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[7] P. E. Johnson, M. O. Smith, S. Taylor-George, and J. B. Adams, “A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures,” *Journal of Geophysical Research*, vol. 88, pp. 3557–3562, Apr. 1983.

[8] K. Guilfoyle, M. L. Althouse, and C.-I. Chang, “Further investigations into the use of linear and nonlinear mixing models for hyperspectral image analysis,” *Proceedings of SPIE*, vol. 4725, pp. 157–167, 2002.

Chapter 3

Survey of Spectral Unmixing Approaches

The spectral unmixing problem can be partitioned into three distinct steps or operations. These operations are dimension reduction, endmember extraction, and inversion. Although dimension reduction is a common unmixing step—typically the first—it is not always employed and might be optional in some situations. In contrast with dimension reduction, some form of endmember extraction and inversion are always required in the spectral unmixing process. In many cases these two steps are coupled, so that unmixing is performed through an iterative process of computing endmembers based on abundances and vice versa. Such coupling makes it difficult to discuss endmember extraction separately from inversion. Each of these steps is discussed in more detail in the sections that follow. The most common approaches to dimension reduction are introduced in Section 3.1. A handful of commonly-used inversion approaches are described in Section 3.2. A brief survey of endmember extraction algorithms is given in Section 3.3. Inversion approaches that are uniquely coupled to a single endmember extraction approach are described along with that endmember extraction approach in Section 3.3 rather than Section 3.2.

3.1 Dimension Reduction

The operation of dimension reduction is typically employed as a preprocessing step in the spectral unmixing process. As such, it is often considered to be optional. The goal of dimension reduction is not only to reduce the number of dimensions used to represent the data, but to do so in a way that retains as much useful information as possible. This reduces the computational complexity of subsequent algorithms and requires less physical storage. It can also improve the signal-to-noise ratio (SNR) of the data so long as the data that are discarded correspond to noise. Despite its wide application to the spectral unmixing problem, dimension reduction can produce undesired results by eliminating data which are important [see Chapter 4]. The three most frequently used dimension reduction approaches are principal component analysis (PCA) [1], maximum noise fraction (MNF) transformation [2,3], and singular value decomposition (SVD) [4]. For a more broad survey of dimension reduction approaches, the reader is referred to the paper by Bioucas-Dias [5].

3.2 Inversion

Inversion is the process of estimating the abundances, given a set of endmember spectra. Although it is typically the final step in the spectral unmixing process, there are a few inversion approaches that are frequently used which warrant presentation before endmember extraction is discussed. The most common inversion approaches are least-squares approaches based on the linear mixing model in (2.21). Using the LMM notation, and ignoring constraints, the least-squares problem is $\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{M}\mathbf{a}\|_2^2$, which is solved by

$$\hat{\mathbf{a}}_{\mathrm{LS}} = \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T\mathbf{x}, \tag{3.1}$$

as long as $\mathbf{M}$ has full column rank. A closed-form solution is also available when the additivity constraint (2.23) is enforced:

$$\hat{\mathbf{a}}_{\mathrm{ACLS}} = \hat{\mathbf{a}}_{\mathrm{LS}} - \mathbf{z}\left(\mathbf{1}^T\hat{\mathbf{a}}_{\mathrm{LS}} - 1\right), \tag{3.2}$$

where $\mathbf{1}$ is a P-element column vector of ones, and $\mathbf{z} = \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{1}\left[\mathbf{1}^T\left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{1}\right]^{-1}$ [6].

The additivity constraint can also be incorporated in a less strict fashion by augmenting the data vector and endmember matrix so that $\tilde{\mathbf{x}} = \begin{bmatrix}\mathbf{x} \\ \delta\end{bmatrix}$ and $\tilde{\mathbf{M}} = \begin{bmatrix}\mathbf{M} \\ \delta\mathbf{1}^T\end{bmatrix}$ where δ is a positive scalar value [7]. The least-squares solution is then calculated according to (3.1) using the augmented data. In this formulation the additivity constraint is not strictly enforced, rather the value of δ is used to weight the importance of the constraint relative to the minimizing of the error in the representation.

The nonnegativity constraint (2.22) is incorporated less frequently because there is no closed-form solution associated with it. One approach is to use one of the least-squares approaches described above and then set all the negative elements of the solution to zero, rescaling if necessary to enforce the additivity constraint [7]. Solutions to a least-squares formulation that incorporates the nonnegativity constraint can be obtained using quadratic programming methods. One such algorithm that uses an active set approach is the nonnegatively constrained least squares (NCLS) method of Chang and Heinz [8]. This approach has subsequently been combined with the augmented additivity constrained least squares approach [7].

3.3 Endmember Extraction

As a step in the spectral unmixing process, the role of endmember extraction—as the name makes obvious—is to produce a set of endmembers from a hyperspectral image. The goal of finding a data set that is representative of an image is one that also appears in the context of dimension reduction and image compression. But, in contrast with those problems, the set of endmembers is required to be physically meaningful. Rather than describing the variance or energy in an image as might be the case with a data set obtained for dimension reduction or image compression, endmembers must describe the materials in the image. Because the spectra of materials can be quite similar, endmembers can be strongly correlated, which is usually not the case in dimension reduction and image compression.

3.3.1 Non-Statistical Endmember Extraction

One non-statistical approach to endmember extraction is to find the spectra in the image that are most extreme, based on some measure of extremity. Such an approach can be justified by the geometry of the linear mixing model, although these approaches do not explicitly exploit that geometry. In the absence of noise the LMM constrains mixed pixels to fall inside a simplex whose vertices are the image endmembers, making the endmembers the most extreme pixels in the data.

Pixel purity index (PPI) finds the extreme pixels in a scene by projecting all of the pixels onto a large number of randomly-generated vectors [9,10]. The pixels which produce the minimum and maximum projected values for each vector are identified and a counter associated with them is incremented. Those pixels which have the largest resulting counter values after all of the projections are complete are selected as the endmembers.

An algorithm that is similar to PPI which incorporates spatial context is spatial-spectral endmember extraction (SSEE) [11]. Instead of randomly-generated vectors, pixels are projected onto the eigenvectors obtained by performing singular value decompositions on square, non-overlapping subsets of the image. Those pixels corresponding to extreme projections are retained as endmember candidates. A window is then scanned across the image and any pixels in the window that are spectrally similar to endmember candidates in the window are averaged together. This scanning and averaging process is repeated multiple times.

Automated morphological endmember extraction (AMEE) uses the morphological operations of dilation and erosion to extract endmembers [12]. As defined for hyperspectral data, dilation finds the pixel that is most spectrally distinct from the other pixels in a spatial kernel, and erosion finds the pixel which best represents all of the pixels in the spatial kernel. Beginning with a kernel of some minimum size, the morphological eccentricity index (MEI) is calculated. The MEI is an image which describes the distance between the two pixels obtained from dilation and erosion. This is repeated for larger and larger kernels, updating the MEI at each iteration until a maximum kernel size is reached. Endmembers are then selected by thresholding the MEI image.

Other non-statistical approaches directly exploit the geometry of the linear mixing model. Visual representations of some of the geometric constructs used by these methods are shown in Fig. 3.1. These approaches include minimum-volume transformations (MVT) which attempt to find a simplex (whose vertices are the endmembers) of minimum volume which completely envelops the observed data. When the vertices of a simplex are stacked into a matrix, **A**, the volume enclosed by that simplex is given by

$$V(\mathbf{A}) \propto \det\left(\begin{bmatrix} \mathbf{a}_1 - \mathbf{a}_0 & \mathbf{a}_2 - \mathbf{a}_0 & \dots & \mathbf{a}_N - \mathbf{a}_0 \end{bmatrix}\right). \tag{3.3}$$

For the determinant above to be defined, **A** must be a $N \times N + 1$ matrix. For this reason, dimension reduction is always performed for a volume-based method. It is also common for a projection or normalization step to be performed so that the length of each observed pixel is identical. This is a simplification to the problem which effectively eliminates spectral variability due to scene topography.

MVT approaches include the dark-point-fixed transform (DPFT) which assumes a known dark point that represents the sensor output when no energy is incident upon it. The data are translated so this point is at the origin, making $\mathbf{a}_0 = \mathbf{0}$ in (3.3). The volume is then

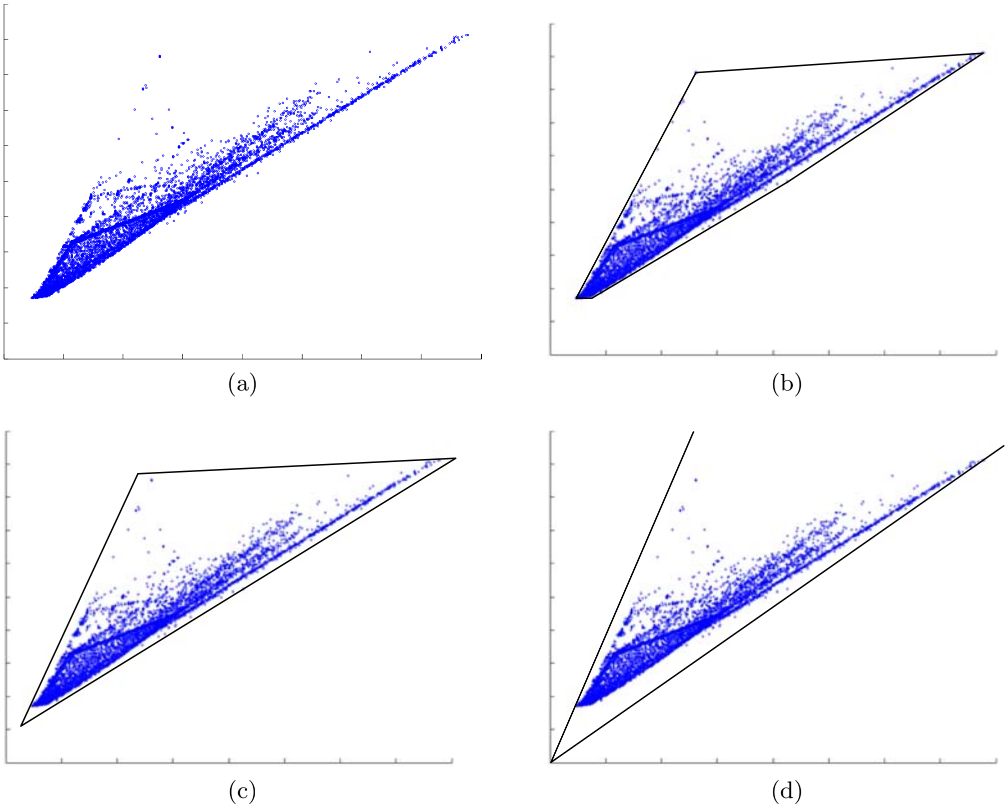


Fig. 3.1: Geometric constructs used in non-statistical endmember extraction. (a) A two-dimensional scatter plot of hyperspectral data. Hyperspectral data enclosed by (b) a convex hull, (c) a simplex, and (d) a convex cone.

minimized, one face of the simplex at a time, subject to the observed data being enclosed by the simplex. This constrained optimization is performed using linear programming [13]. A similar approach is the fixed-point-free transform (FPFT) which makes no assumption of a known dark point [13]. In this case, a row of some constant value is appended to the dimension-reduced endmember matrix. This has the effect of creating a new, higher-dimensional (by one) simplex with a new, fixed vertex at the origin. The volume of this new simplex is proportional to the prior one, so the algorithm proceeds in the same fashion as the DPFT. The simplex identification via split augmented Lagrangian (SISAL) is similar to the DPFT, but uses soft nonnegativity constraints and convex approximations to reduce computational complexity [14]. Another MVT approach is the shrink-wrap algorithm [15] which minimizes the cost function

$$J(\mathbf{M}, \mathbf{X}) = V(\mathbf{M}) + \lambda \sum_{p} \sum_{t} \frac{1}{a_p(t)}, \tag{3.4}$$

where the values of $a_p(t)$ are obtained from the augmented least-squares inversion described in Section 3.2. The first term in the cost function is the volume enclosed by the simplex, and the second term describes the distance of the observed data from a face of the simplex. The relative emphasis of these two terms is controlled by λ. The algorithm begins with a large value of λ and proceeds with the optimization, slowly lowering the value of λ. This forces the simplex to be far from the data initially, so that the cost function can only be reduced by reorienting the simplex to a more accurate position. Then, as λ decreases the simplex is allowed to shrink to fit the data more tightly. This is done to make the approach more robust to initialization, as local minima can be problematic for MVT approaches. Another MVT approach that is intended to be robust to local minima is based on simulated annealing [16].

N-FINDR is another geometric approach that utilizes the volume of a simplex, but instead of shrinking a simplex to envelop the data, it inflates a simplex that is wholly contained within the data [17]. The volume calculation is formulated just as with the FPFT, but rather than doing a constrained minimization, a combinatorial approach is taken. Pixels are randomly selected from the observed data to be endmember candidates. Then, each pixel is substituted into the first endmember location, with the other endmembers remaining fixed. If the substituted pixel increases the volume of the resulting simplex, the pixel is retained, otherwise it is discarded. This process is repeated for each location in the endmember matrix. A family of algorithms (AVMAX, SVMAX, and WAVMAX) based on N-FINDR have been developed to improve convergence, especially in the presence of noise [18].

Vertex component analysis (VCA) [19] attempts to find the vertices of the simplex by using randomly-generated orthogonal projections. After normalizing the data, an iterative process randomly generates a vector, makes it orthogonal to previously generated vectors, and projects each pixel onto that vector. The pixels corresponding to the most extreme projection are selected as endmembers.

Yet another geometric approach is convex cone analysis (CCA) [20]. This approach attempts to find the vertices of a polygon that encloses all possible linear combinations of the first N eigenvectors of the spectral correlation matrix. This is done by solving a large number of systems of equations, making the approach very computationally expensive. Although the method for finding a solution is quite different, the geometry of the problem is nearly identical to the DPFT.

Other non-statistical approaches have been developed which attempt to find an accurate representation of the data. Iterative error analysis (IEA) is such an approach [21]. The IEA algorithm proceeds by performing constrained inversion assuming the mean of the data is the only endmember. The error of the representation is calculated and the pixel corresponding to the largest error is selected as an endmember, replacing the mean. The process is repeated and at each iteration one new pixel is appended to the endmember matrix used for inversion. The process terminates when a specified number of endmembers has been obtained or when the representation error falls below a certain threshold.

Fuzzy k-means (also called fuzzy c-means) is a classification approach that attempts to find a minimum error representation for the data [22]. Unlike many other classification classification approaches fuzzy k-means allows for fractional class membership, where the fractions must sum to one. This constraint is similar to the LMM additivity constraint, motivating its application to hyperspectral data [23]. Unlike other approaches, fuzzy k-means does not assume any sort of linear mixing model.

All of the non-statistical approached described above—except for fuzzy k-means classification—make use either implicitly or explicitly of the linear mixing model. Most of these approaches assume that pure endmember pixels are present in the observed data. Many are sensitive to initialization due to local minima and non-unique solutions [24].

3.3.2 Statistical Endmember Extraction

Statistical endmember extraction approaches can be divided into parametric and non-parametric categories. Non-parametric statistical methods are those that make an explicit assumption that the data are random, without necessarily determining the specific probability distribution that describes that random behavior. Instead, these approaches tend to use statistical quantities such as means and variances to obtain a solution.

Independent component analysis (ICA) is used to describe a variety of approaches that separate mixed "sources" by forcing the unmixed data to be as independent as possible [25]. A linear mixing model is typically assumed, and frequently it is assumed that the mixing matrix is square, necessitating some form of dimension reduction. In one case a contextual ICA approach was employed where the endmembers were treated as the sources (which are assumed to be independent), and the abundances as the mixing matrix [26]. Other applications of ICA to the hyperspectral unmixing problem treat the abundances as the sources and the endmembers as the mixing matrix. These include joint cumulant-based ICA [27], joint approximate diagonalization of eigen-matrices (JADE) [28], and FastICA [28–31]. A modified version of FastICA that forces the separated sources to be orthogonal to one another is linear spectral random mixture analysis (LSRMA) [32]. One of the complications associated with ICA is scale ambiguity. This arises because multiplication by a scalar does not affect the independence of two signals. In light of this, it has been suggested that ICA is not applicable to the abundance quantification problem, but is better suited to classification and target detection [33]. These topics are addressed in more detail in Chapters 4–6.

Iterated constrained endmembers (ICE) is a statistical approach that is similar to both representation-based and geometric non-statistical approaches [34]. The algorithm assumes a representation error that is approximately Gaussian, motivating a least-squares approach to representing the data. To prevent the simplex enclosed by the estimated endmembers from becoming too small, i.e., due to a lack of pure endmembers, a penalty function, $V(\mathbf{M})$, is defined to be the sum of the variances of the endmembers. This leads to a cost function of the form

$$J(\mathbf{M}, \mathbf{A}) = (1 - \lambda) \left\| \mathbf{X} - \mathbf{M}\mathbf{A} \right\|_2^2 + \lambda V(\mathbf{M}). \tag{3.5}$$

This function is minimized in alternating fashion over $\mathbf{M}$ (using a closed-form solution) and $\mathbf{A}$ (using quadratic programming to enforce additivity and nonnegativity constraints). An extension of ICE that attempts to arrive at sparse estimates for $\mathbf{A}$ is sparsity promoting iterated constrained endmembers (SPICE) [35]. The SPICE algorithm adds an additional term to the ICE cost function in (3.5) to produce

$$J(\mathbf{M}, \mathbf{A}) = (1 - \lambda) \left\| \mathbf{X} - \mathbf{M}\mathbf{A} \right\|_2^2 + \lambda V(\mathbf{M}) + c \sum_t \sum_p \gamma_p a_{t,p}. \tag{3.6}$$

The constant c is used generally to control the degree to which abundance values are driven to zero, while the value of γ_p changes based on the abundance estimates to drive small abundance values more quickly to zero. This additional term can be viewed as assuming a Laplacian prior for the abundances. As with ICE, the minimization of the cost function is performed iteratively alternating between $\mathbf{M}$ and $\mathbf{A}$. SPICE introduces an additional step of "pruning" endmembers whose maximum abundance proportions drop below some threshold.

Algorithms have also been developed for hyperspectral unmixing that attempt to find signals that are minimally complex, or alternatively, that are maximally predictable [36]. These algorithms maximize a cost function that is a ratio of the variance of the signal and its predictability which is defined as the difference between a pixel and an average of its neighbors. By this definition, predictability might also be thought of as smoothness.

The cost function is maximized using a gradient ascent algorithm, resulting in the spatial complexity blind source separation (SCBSS) algorithm [36]. An extension of this algorithm adds a term to the cost function to encourage spectral smoothness to produce the spectral and spatial complexity blind source separation (SSCBSS) algorithm [36].

Nonnegative matrix factorization (NMF) is another approach that has been utilized for hyperspectral unmixing. NMF encompasses a wide variety of problems in which the goal is to find two nonnegative matrices, $\mathbf{W} \in \mathbb{R}_{+}^{m\times k}$ and $\mathbf{H} \in \mathbb{R}_{+}^{k\times n}$, whose product provides a reasonable approximation of another matrix, $\mathbf{Y} \in \mathbb{R}_{+}^{m\times n}$, i.e., $\mathbf{Y} \approx \mathbf{WH}$ [37]. At their root, NMF hyperspectral unmixing approaches assume gaussian-distributed noise, which motivates (in an ML sense) minimizing the Frobenius norm of the error in the representation, $\|\mathbf{X} - \mathbf{MA}\|_F^2$. This minimization is subject to the nonnegativity of $\mathbf{M}$ and $\mathbf{A}$. A common approach to solving this problem is through iterative multiplicative updates [38]. Other approaches use an additive update rule combined with setting negative results to zero, since the additive steps do not guarantee nonnegative results [39]. In order to improve the performance of NMF approaches, application-specific auxiliary constraints are often incorporated into the problem formulation, resulting in cost functions of the form

$$J(\mathbf{M}, \mathbf{A}) = \|\mathbf{X} - \mathbf{MA}\|_F^2 + \sum_i \lambda_i J_i(\mathbf{M}, \mathbf{A}). \tag{3.7}$$

These cost functions are typically minimized using modified multiplicative or additive update rules. Under certain circumstances, they can also be solved using an alternating least-squares (ALS) approach [39]. Additional constraints incorporated into the cost function for hyperspectral unmixing include additivity [40–42], smoothness [43], and simplex volume [44].

Parametric methods assume specific distributions to describe the data and attempt to find the best parameters for those distributions, usually using a maximum likelihood (ML) or maximum a posteriori (MAP) approach. Gaussian class estimation is one example of a parametric approach [45]. This approach models the observed data using a finite gaussian mixture model. This model assumes that there are a finite number of “pure” classes, i.e., endmembers, which are gaussian distributed. Linear mixtures of these endmembers are also gaussian distributed. To make the problem more tractable the mixing coefficients are discretized to produce a finite number of mixture classes. A variation of the expectation-maximization (EM) algorithm [46] known as stochastic expectation-maximization (SEM) [47] is used to determine class means and covariances and assign pixels into the most likely class. Given initial values for all of the prior class probabilities, and means and covariances for the endmember classes, posterior class probabilities are computed for each pixel. Based on these posterior probabilities, each pixel is randomly assigned to a class. The prior, mean and covariance of each class is then updated based on the pixels which have been assigned to it. It has been suggested that this approach works better when the prior probabilities are not updated from iteration to iteration, but instead remain constant [48].

A similar approach is dependent component analysis (DECA) [49, 50]. This approach assumes that the abundances follow a K-component Dirichlet finite mixture:

$$p_A\left(\mathbf{a}|\boldsymbol{\theta}\right) = \sum_{q=1}^{K} \epsilon_q D\left(\mathbf{a}|\boldsymbol{\theta}_q\right). \tag{3.8}$$

The observed data are assumed to be i.i.d. samples from a random variable $\mathbf{X}$. Using a change of variables, a log-likelihood is formulated in terms of $p_A(\mathbf{a}|\boldsymbol{\theta})$. The parameters to be estimated are $\mathbf{W}$, a square unmixing matrix, and the Dirichlet mixing model parameters, $\boldsymbol{\theta} = \{\epsilon_1, \ldots, \epsilon_K, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K\}$. This estimation is done using the EM algorithm, where the hidden data are a set of labels for each observation indicating to which class the observation belongs. At each iteration of the algorithm the parameter values are updated sequentially, first the values of ϵ_q, then the values of $\boldsymbol{\theta}_q$, and finally $\mathbf{W}$. The algorithm is actually a generalized EM (GEM) algorithm as the update approaches for $\boldsymbol{\theta}_q$ and $\mathbf{W}$ do not maximize the Q-function, they only ensure that it does not decrease. This approach has been subsequently modified to infer then number of Dirichlet classes and to reduce the computational complexity of the GEM algorithm [51].

A MAP framework has been proposed [52] based on the model in (2.24) where

$$p(\mathbf{M}, \mathbf{A}|\mathbf{X}) \propto p_n(\mathbf{X}|\mathbf{M}, \mathbf{A}) p_m(\mathbf{M}) p_a(\mathbf{A}). \tag{3.9}$$

The prior on the abundances is chosen to be zero if the additivity constraint is violated and constant otherwise. The prior on the spectra is based on a linear auto-regressive model. The noise is assumed to be gaussian distributed. The log of the posterior distribution is treated as a cost function and iteratively maximized. Each iteration consists of a gradient step for $\mathbf{A}$ and explicit constrained least-squares solution for $\mathbf{M}$ until the algorithm has converged.

Bayesian positive source separation (BPSS) utilizes a hierarchical Bayesian model to perform linear unmixing subject to nonnegativity and additivity constraints [53]. The approach assumes gaussian-distributed noise, leading to a gaussian-distributed likelihood. The endmembers and abundances are both assigned gamma distributions for priors. The parameters of the gamma distributions are assumed constant for each endmember, but may change from one endmember to another. The same is true for the abundances corresponding to each endmember. Uninformative priors are chosen for the hyperparameters. The MAP estimates for the endmembers and abundances are obtained using a Gibbs sampler [54, 55] to sample from the posterior density. BPSS is applied to hyperspectral image data in [56].

An approach that is similar to BPSS but perhaps more uniquely tailored for hyperspectral unmixing is joint Bayesian endmember extraction (JBEE) [57]. As with BPSS, the likelihood is assumed to be gaussian. The endmembers are assumed to follow truncated multivariate gaussian distributions which are proportional to standard multivariate gaussian distributions, but are zero anywhere any component of the endmember is negative. The means of these distributions are obtained from a simpler endmember extraction algorithm such as VCA [19] or N-FINDR [17]. The variance of these distributions are fixed as large values to indicate relative uncertainty. The prior for the abundances is a uniform distribution over a simplex which satisfies the additivity constraint. This is equivalent to a Dirichlet distribution with all of its parameters equal to one. Finally, the noise variance is assumed to follow an inverse-gamma distribution with one parameter fixed and the other assigned a noninformative Jeffrey's prior [54]. Again, a Gibbs sampler is use to find the MAP estimates of the model.

Bayesian approaches are appealing because they provide a framework for incorporating all of the assumptions and constraints associated with a problem. Rather than a single estimated value, these approaches provide a distribution allowing the confidence in an estimate to be quantified. Their performance has been shown to be superior to other approaches such as VCA and N-FINDR [57]. The trade though, is computational complexity. For example, JBEE has been shown to be slower than N-FINDR by a factor of ∼65 and slower than VCA by a factor of ∼1500! To unmix a hyperspectral cube with 200 bands and 128×128 spatial pixels, the BPSS approach required over four and a half days [56]. Such performance makes these approaches computationally prohibitive in many cases.

References

[1] I. Jolliffe, *Principal Component Analysis*, ser. Springer Series in Statistics. New York: Springer-Verlag, 1986.

[2] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[3] J. Lee, A. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295–304, May 1990.

[4] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing.* Upper Saddle River, NJ: Prentice Hall, 2000.

[5] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[6] J. J. Settle and N. A. Drake, "Linear mixing and the estimation of ground cover proportions," *International Journal of Remote Sensing*, vol. 14, no. 6, pp. 1159–1177, Apr. 1993.

[7] D. Heinz and C.-I. Chang, “Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[8] C.-I. Chang and D. Heinz, “Constrained subpixel target detection for remotely sensed imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1144–1159, May 2000.

[9] J. W. Boardman, F. A. Kruse, and R. O. Green, “Mapping target signatures via partial unmixing of AVIRIS data,” *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, pp. 23–26, 1995.

[10] C.-I. Chang and A. Plaza, “A fast iterative algorithm for implementation of pixel purity index,” *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 63–67, Jan. 2006.

[11] D. 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.

[12] A. Plaza, P. Martinez, R. Perez, and J. Plaza, “Spatial/spectral endmember extraction by multidimensional morphological operations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sept. 2002.

[13] M. Craig, “Minimum-volume transforms for remotely sensed data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

[14] J. Bioucas-Dias, “A variable splitting augmented lagrangian approach to linear spectral unmixing,” *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS ’09*, pp. 1–4, Aug. 2009.

[15] D. R. Fuhrmann, “A simplex shrink-wrap algorithm,” *Proceedings of SPIE*, vol. 3718, pp. 501–511, 1999.

[16] C. Bateson, G. Asner, and C. Wessman, “Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.

[17] M. E. Winter, “N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data,” *Proceedings of SPIE*, vol. 3753, pp. 266–275, 1999.

[18] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, “A simplex volume maximization framework for hyperspectral endmember extraction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4177 –4193, Nov. 2011.

[19] J. Nascimento and J. Dias, “Vertex component analysis: a fast algorithm to unmix hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[20] 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, Mar. 1999.

[21] R. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, “Automatic endmember extraction from hyperspectral data for mineral exploration,” *Proceedings of 21st Canadian Symposium on Remote Sensing*, pp. 891–897, 1999.

[22] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York: Plenum Press, 1981.

[23] G. M. Foody and D. P. Cox, “Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions,” *International Journal of Remote Sensing*, vol. 15, no. 3, pp. 619–631, 1994.

[24] A. Plaza and C.-I. Chang, “Impact of initialization on design of endmember extraction algorithms,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3397–3407, Nov. 2006.

[25] A. Cichocki and S. ichi Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* West Sussex, UK: John Wiley & Sons, 2002.

[26] J. Bayliss, J. A. Gualtieri, and R. F. Cromp, “Analyzing hyperspectral data with independent component analysis,” *Proceedings of SPIE*, vol. 3240, pp. 133–143, 1997.

[27] X. Zhang and C. H. Chen, “New independent component analysis method using higher order statistics with application to remote sensing images,” *Optical Engineering*, vol. 41, pp. 1717–1728, 2002.

[28] J. Nascimento and J. Dias, “Does independent component analysis play a role in unmixing hyperspectral data?” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[29] T. Tu, “Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis approach,” *Optical Engineering*, vol. 39, no. 4, pp. 897–906, Apr. 2000.

[30] B. R. Foy and J. Theiler, “Scene analysis and detection in thermal infrared remote sensing using independent component analysis,” *Proceedings of SPIE*, vol. 5439, pp. 131–139, 2004.

[31] J. Wang and C.-I. Chang, “Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 260–2616, Sept. 2006.

[32] C.-I. Chang, S. S. Chiang, J. A. Smith, and I. W. Ginsberg, “Linear spectral random mixture analysis for hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 2, pp. 375–392, 2002.

[33] S. Chiang, C.-I. Chang, and I. Ginsberg, “Unsupervised hyperspectral image analysis using independent component analysis,” *IEEE Geoscience and Remote Sensing Symposium*, vol. 7, pp. 3136–3138, 2000.

[34] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. 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, Oct. 2004.

[35] 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, July 2007.

[36] S. Jia and Y. Qian, “Spectral and spatial complexity-based hyperspectral unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.

[37] A. Cichocki, R. Zdunek, A. H. Phan, and S. ichi Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multiway Data Analysis and Blind Source Separation.* West Sussex, UK: John Wiley & Sons, 2009.

[38] D. D. Lee and H. S. Seung, “Algorithms for nonnegative matrix factorization,” *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[39] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, “Algorithms and applications for approximate nonnegative matrix factorization,” *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[40] Y. M. Masalmah, M. Velez-Reyes, and S. Rosario-Torres, “An algorithm for unsupervised unmixing of hyperspectral imagery using positive matrix factorization,” *Proceedings of SPIE*, vol. 5806, pp. 703–710, 2005.

[41] Y. M. Masalmah and M. Veléz-Reyes, “A full algorithm to compute the constrained positive matrix factorization and its application in unsupervised unmixing of hyperspectral imagery,” *Proceedings of SPIE*, vol. 6966, no. 1, p. 69661C, 2008.

[42] Q. Du, I. Kopriva, and H. Szu, “Investigation on constrained matrix factorization for hyperspectral image analysis,” *IEEE Geoscience and Remote Sensing Symposium*, vol. 6, pp. 4304–4306, July 2005.

[43] V. P. Pauca, J. Piper, and R. J. Plemmons, “Nonnegative matrix factorization for spectral data analysis,” *Linear Algebra and its Applications*, vol. 416, no. 1, pp. 29–47, 2006.

[44] L. Miao and H. Qi, “Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[45] A. D. Stocker and A. P. Schaum, “Application of stochastic mixing models to hyperspectral detection problems,” *Proceedings of SPIE*, vol. 3071, pp. 47–60, 1997.

[46] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[47] P. Masson and W. Pieczynski, “SEM algorithm and unsupervised statistical segmentation of satellite images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 3, pp. 618–633, May 1993.

[48] M. T. Eismann and R. C. Hardie, “Stochastic spectral unmixing with enhanced endmember class separation,” *Applied Optics*, vol. 43, no. 36, pp. 6596–6608, Dec. 2004.

[49] J. Nascimento and J. Bioucas-Dias, “Hyperspectral unmixing algorithm via dependent component analysis,” *IEEE Geoscience and Remote Sensing Symposium*, pp. 4033–4036, July 2007.

[50] J. Nascimento and J. M. Bioucas-Dias, “Hyperspectral unmixing based on mixtures of dirichlet components,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1 –16, 2011.

[51] J. Nascimento and J. Bioucas-Dias, “Learning dependent sources using mixtures of dirichlet: Applications on hyperspectral unmixing,” *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS ’09*, pp. 1–5, Aug. 2009.

[52] L. Parra, C. Spence, P. Sajda, A. Ziehe, and K.-R. Müller, “Unmixing hyperspectral data,” *Advances in Neural Information Processing Systems (NIPS)*, vol. 12, pp. 942–948, 1999.

[53] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, “Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

[54] J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing.* New York: Springer-Verlag, 1996.

[55] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms.* Cambridge, UK: Cambridge University Press, 2003.

[56] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. Benediktsson, “On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation,” *Neurocomputing*, vol. 71, no. 10-12, pp. 2194–2208, June 2008.

[57] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. Hero, “Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery,” *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

Chapter 4

A Quantitative Assessment of Independent Component Analysis as a Hyperspectral Unmixing Approach[1]

4.1 Introduction

Hyperspectral imaging is a remote sensing approach that simultaneously collects both spatial and spectral data. Spectral data are collected in hundreds of narrow contiguous bands that may cover the visible, near-infrared, and short-wave infrared (0.4–2.5 μm), the mid-wave infrared (3–5 μm), and/or the long-wave infrared (8–14 μm). Although the size of a pixel on the ground varies, spatial measurements typically consist of hundreds of pixels in both spatial dimensions. Such images contain a wealth of information and have found application in a broad range of fields such as food safety [1], agriculture [2], mineralogy [3], ecology [4], and target detection [5], as well as many others.

One method of exploiting hyperspectral image data is through spectral unmixing. This process refers to one or both of two fundamental operations. The first is the identification of spectra that are representative of the distinct materials in the scene. These spectra are referred to as endmembers and the problem of identifying them is called endmember extraction. The endmember spectrum associated with a material should not be confused with the spectral signature for that material found in a library of reference spectra, as the two are almost always different. It is possible that an endmember spectrum may not be found in the observed data. This occurs when the material associated with that endmember does not completely fill any single pixel in the image. In that case, the endmember spectrum will only be found in the observed data in combination with other endmember spectra.

---

[1] M.R. Stites, J.H. Gunther, T.K. Moon, and G.P. Williams (to be submitted for publication)

Because an endmember is uniquely associated with a specific material, the terms endmember and material are used interchangeably throughout the remainder or this chapter.

The second operation is abundance quantification, which entails determining the proportion of each endmember in each pixel of the image. Abundance maps provide useful visualizations of this data, showing where each endmember is located in an image and how completely each pixel is filled by that endmember. Depending on the algorithm and the application, endmembers may be determined first and subsequently utilized for abundance quantification, the endmembers and abundances may be found simultaneously, or abundances may be computed without any prior endmember information [6].

There are a wide variety of algorithms that have been developed to unmix hyperspectral data. Some of these attempt to find the most "extreme" pixels in the image, based on some definition of distance or extremity. These include pixel purity index (PPI) [7], iterative error analysis (IEA) [8], and automated morphological endmember extraction (AMEE) [9]. Other, similar approaches exploit the geometry imposed by the linear mixing model (LMM), described later. Minimum volume transformation (MVT) [10], N-FINDR [11], convex cone analysis (CCA) [12], vertex component analysis (VCA) [13], and shrink-wrapping [14] are all algorithms of this type. Statistical approaches which assume specific probability distributions for the observed hyperspectral data and/or abundances have also been developed, such as dependent component analysis (DECA) [15], and Bayesian positive source separation (BPSS) [16].

A statistical unmixing approach that does not assume a specific distribution for the data is independent component analysis (ICA) [17]. This approach attempts to unmix the data by finding maximally independent abundances. A variety of ICA algorithms have been applied to hyperspectral unmixing including contextual ICA [18], joint cumulant-based ICA [19], joint approximate diagonalization of eigen-matrices (JADE) [20], and FastICA [20–23]. ICA has also been employed as a hyperspectral classification approach [24, 25].

Whenever spectral unmixing algorithms are assessed, two types of experiments are typically performed. In the first, synthetic images are created according to a simple generative model—usually the linear mixing model. The complexity of these images varies, but they are typically composed of 2–10 endmembers whose spectra are obtained from a real hyperspectral image or from a spectral reference library. In many cases spatial contiguity is incorporated using abundance maps consisting of simple square or circular regions. These kinds of test images are fairly common in the spectral unmixing literature [9,11,12,26,27]. Since many spectral unmixing approaches do not consider spatial context, synthetic images can also be produced using randomly generated abundances which adhere to some probability distribution. In these cases a generative model is used which incorporates other interesting behavior, such as topographic variation and endmembers with spectral variability [13, 20]. In the majority of these cases the endmembers are generated in relative proportion one with another. That is, there is no single material which dominates the scene spatially and no material that is present in only a very small fraction of pixels. These images are useful because they are relatively simple to generate, and because complete ground truth data are available, including abundance maps accurate to small fractions of a pixel. Spectral unmixing results can then be compared against the ground truth data to provide quantitative assessments of spectral unmixing algorithms.

The second type of experiment is to test an algorithm by unmixing a real hyperspectral data set. The results of the unmixing are often assessed visually by recognizing landmarks in the original image and in the unmixed data [19,22]. In some cases ground truth data are available and can be compared to unmixing results [18, 26]. Unfortunately, these ground truth data often only provide information for a handful of the materials in the scene and may be incomplete for certain areas/materials in the image. They do not provide the fine abundance resolution of synthetic images and are not available for every image which might be of interest.

Both of the experimental approaches described above are useful and even essential to assessing the usefulness and behavior of a hyperspectral approach. There is, however, a third approach that can be seen as something of a middle ground between the two. This approach utilizes synthetic images that more closely approximate real data by modeling scene geometry, material properties, sensor behavior, atmospheric contributions, and so forth. Complex scene geometry is desirable because it produces images that have regions of spatial contiguity, topographic variation, and endmember spectral variability. It also leads to broad variations in the spatial coverage of individual materials. Because the images are synthetic, complete ground truth data are still available. Such an approach is not intended to be a replacement for the existing methods described above. Instead, it should be treated as a complementary approach, allowing for unique insights and observations to be explored.

This complementary approach could be employed to explore a variety of hyperspectral unmixing algorithms. Throughout the remainder of this chapter, it is used to assess the behavior of ICA—specifically FastICA. Although application of FastICA to the problem has been explored before [20–23], there are still questions regarding its utility as a hyperspectral unmixing approach. A common opinion—though not a consensus—is that ICA can produce interesting and useful results, but that it is common for some materials to be incorrectly unmixed [20, 22, 23]. Thus, further exploration is warranted to confirm existing assertions regarding FastICA and also provide further insight into the behavior of the algorithm.

The remainder of the chapter proceeds as follows. Section 4.2 provides a basic overview of ICA and the FastICA algorithm. It also outlines the ICA data model and compares it with the linear mixing model used to describe hyperspectral data. Section 4.3 explains the approach taken to generate synthetic—but realistic—hyperspectral data cubes. Examples of both image data and abundance maps are shown. Section 4.4 describes the experiments performed, presents the results of those experiments, and provides insight into those results. Finally, Section 4.5 contains a few concluding observations and remarks.

4.2 Independent Component Analysis

The generalized blind source separation (BSS) problem is modeled as

$$\mathbf{x}(t) = f\left(\mathbf{s}(t)\right), \tag{4.1}$$

where $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \ldots\ x_K(t)]^T$ is the observed data vector, $\mathbf{s}(t) = [s_1(t)\ s_2(t)\ \ldots\ s_L(t)]^T$

is a vector of the sources of interest, and $f(\cdot)$ describes the mixing process which operates on the sources to create the observed data. The observations and sources are indexed by t which, depending on the application, may represent time, spatial location, or some other quantity. In the case of hyperspectral unmixing, t is used to index spatial location, i.e., individual pixels. The goal of BSS is to estimate the original sources from the observed data with limited or no knowledge of either $f(\cdot)$ or $\mathbf{s}(t)$. The estimation process is often referred to as unmixing. BSS has found application in many varied areas including biomedical signal processing [28, 29], telecommunications [30, 31], and finance [32, 33].

ICA is an approach that attempts to perform BSS by exploiting the statistical independence of the original sources. While this can be accomplished in a number of ways, many ICA algorithms invoke the central limit theorem [34], observing that the distribution of mixed random variables tends toward a Gaussian distribution. Hence, sources can be separated by optimizing a cost function that reflects some measure of gaussianity. Commonly used cost functions include kurtosis, a fourth-order cumulant, and negentropy which is the difference between the entropy of the data and a Gaussian random vector of the same correlation. Negentropy can be approximated using a non-quadratic, symmetric function. Other ICA approaches include minimization of mutual information [35], and joint diagonalization of eigenmatrices [36].

Although nonlinear ICA methods exist [37, 38], linear mixing is most commonly assumed. In this case the mixing is represented by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad t = 1 \ldots T, \tag{4.2}$$

where $\mathbf{x}(t)$ is $K \times 1$, $\mathbf{s}(t)$ is $L \times 1$, the mixing matrix, $\mathbf{A}$, is $K \times L$ and T is the total number of observations (pixels). Stacking the observed and source data as $\mathbf{X} = [\mathbf{x}(1)\ \mathbf{x}(2)\ \ldots\ \mathbf{x}(T)]$ and $\mathbf{S} = [\mathbf{s}(1)\ \mathbf{s}(2)\ \ldots\ \mathbf{s}(T)]$, the model becomes

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \tag{4.3}$$

with the $K \times T$ observation matrix $\mathbf{X}$, and $L \times T$ source matrix $\mathbf{S}$. Notice that using this notation $\mathbf{s}(t)$, which is a column of $\mathbf{S}$, references all of the sources at a specific location. Alternatively, a row of $\mathbf{S}$, denoted $\mathbf{s}_i^T$, is used to describe a single source over all locations.

The mixed data must satisfy two important conditions for ICA to be a valid unmixing approach. First, since ICA attempts to unmix the data by exploiting the independence of the sources, the sources must be independent. Second, because the methods of separation utilized by ICA algorithms attempt to maximize nongaussianity (based on the central limit theorem), no more than one source may be gaussian distributed.

4.2.1 FastICA

FastICA is an ICA algorithm that assumes the linear mixing model in (4.3) with the additional constraint that $K = L$, making the mixing matrix $\mathbf{A}$ square. The unmixing model then becomes $\mathbf{Y} = \mathbf{B}\mathbf{X}$, where $\mathbf{Y}$ contains the estimates of the original sources. Defining the unmixing matrix to be

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_K^T \end{bmatrix}, \tag{4.4}$$

a single independent component can be obtained as

$$y_i(t) = \mathbf{b}_i^T \mathbf{x}(t), \tag{4.5}$$

or equivalently,

$$\mathbf{y}_i^T = \mathbf{b}_i^T \mathbf{X}. \tag{4.6}$$

Since neither reordering nor scaling of the estimates affects their independence, ICA outputs are subject to scale ambiguity and order uncertainty. Because of this the unmixing matrix, $\mathbf{B}$, is not necessarily the inverse of $\mathbf{A}$. Rather, $\mathbf{BA} = \mathbf{DP}$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{P}$ is a permutation matrix. Due to this ambiguity, any result of the form $\mathbf{y}_i^T = \gamma \mathbf{s}_j^T$, where γ is a constant scalar value, is generally considered a success.

Prior to performing any source separation the observed data are whitened so that $\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t)$, where $E\left[\mathbf{z}\right] = \mathbf{0}$, and $E\left[\mathbf{z}\mathbf{z}^T\right] = \mathbf{I}$. Incorporating the whitened data, the unmixing model becomes $\mathbf{Y} = \mathbf{WZ} = \mathbf{WVX}$, and $\mathbf{B} = \mathbf{WV}$, where $\mathbf{W}$ is comprised of stacked vectors as $\mathbf{B}$ in (4.4).

As part of the whitening process the dimension of the observed data is reduced via principle component analysis (PCA). Unless specified by the user, the number of dimensions is determined automatically from the relative magnitudes of the eigenvalues of the covariance matrix of the observed data. This dimension reduction step is an attempt to estimate the number of sources and make the mixing matrix square, as required by the FastICA model.

After whitening and dimension reduction, the source separation is achieved by using a simple fixed-point algorithm to maximize a cost function. Thus, the source separation problem becomes

$$\max_{\mathbf{w}_i} E\left\{G\left(\mathbf{w}_i^T \mathbf{z}(t)\right)\right\}, \quad i = 1, \ldots, K. \tag{4.7}$$

Typically, $G(\cdot)$ in (4.7) is defined to be

$$G_1(y) = y^4, \tag{4.8}$$

$$G_2(y) = \frac{1}{a_1} \log\cosh(a_1 y), \tag{4.9}$$

or

$$G_3(y) = -\frac{1}{a_2} \exp(-a_2 y^2/2). \tag{4.10}$$

The derivatives of these functions are

$$g_1(y) = 3y^3, \tag{4.11}$$

$$g_2(y) = \tanh(a_1 y), \tag{4.12}$$

and

$$g_3(y) = y \exp(-a_2 y^2/2). \tag{4.13}$$

The first function is an approximation of the kurtosis of y. Incorporating either of the other two functions gives an approximation of the negentropy of y.

Because the whitening step effectively orthogonalizes the observed data, the unmixing matrix, $\mathbf{W}$ is constrained to be an orthogonal matrix with $\mathbf{W}\mathbf{W}^T = \mathbf{W}^T\mathbf{W} = \mathbf{I}$. This constraint is enforced at each iteration of the cost function optimization in one of two ways. If the components are extracted one at a time, deflationary orthogonalization is performed. This approach updates a single unmixing vector using the gradient optimization algorithm. That vector is then made orthogonal to all of the previously computed unmixing vectors:

$$\widetilde{\mathbf{w}}_i = \mathbf{w}_i - \sum_{j=1}^{i-1} (\mathbf{w}_i^T \widetilde{\mathbf{w}}_j)\widetilde{\mathbf{w}}_j. \tag{4.14}$$

The unmixing vector is then normalized as

$$\widetilde{\mathbf{w}}_i = \widetilde{\mathbf{w}}_i / \|\widetilde{\mathbf{w}}_i\|. \tag{4.15}$$

Alternatively, if all of the components are estimated simultaneously then symmetric orthogonalization is performed. In this case all L unmixing vectors are updated and subsequently orthogonalized using the update formula

$$\widetilde{\mathbf{W}} = \left(\mathbf{W}\mathbf{W}^T\right)^{-1/2} \mathbf{W}. \tag{4.16}$$

4.2.2 Application to Hyperspectral Data

One approach to modeling the radiance of a single pixel in a hyperspectral image is the linear mixing model [39]. This model is typically formulated as

$$\mathbf{x}(t) = \sum_{l=1}^{L} \mathbf{m}_l a_l(t) + \mathbf{n}(t) = \mathbf{M}\mathbf{a}(t) + \mathbf{n}(t). \tag{4.17}$$

In this model $\mathbf{x}(t)$ is the observed $K \times 1$ pixel where K is the number of spectral bands of the sensor. As described previously, the index t is used to indicate the spatial location of the pixel. The $K \times 1$ vector $\mathbf{m}_l$ represents an endmember spectrum and $a_l(t)$ is the fractional abundance of that endmember in the pixel. The total number of endmembers is L. Instrument noise and model error are represented by $\mathbf{n}(t)$. The $K \times L$ matrix $\mathbf{M}$ is the endmember matrix and contains the L individual endmembers in its columns. The $L \times 1$ abundance vector, $\mathbf{a}(t)$, is formed by stacking the relative abundances. The relative abundances are subject to two constraints:

$$a_l(t) \geq 0, \quad l = 1, \ldots, L, \tag{4.18}$$

$$\sum_{l=1}^{L} a_l(t) = 1. \tag{4.19}$$

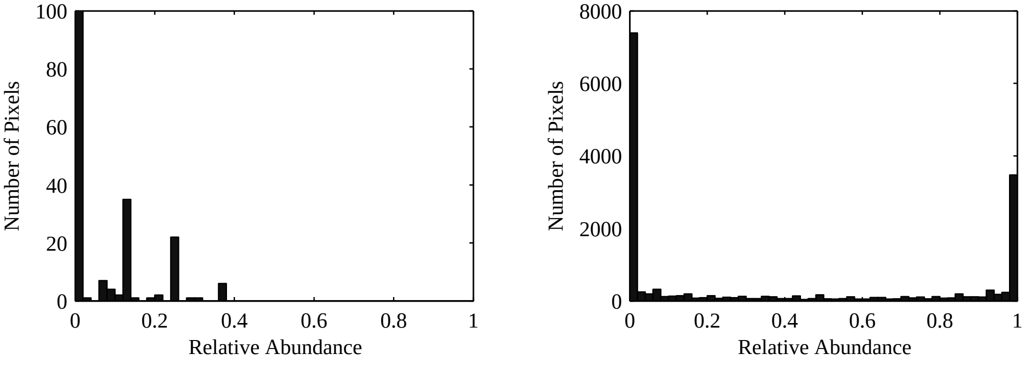
These constraints impose the physically meaningful requirements that the fractional abundances be nonnegative and sum to one. This model is valid only when the materials in the pixel are well-partitioned from one another [40]. Even though this is not always the case in nature, this model is still widely used.

The pixels in the observed cube can be indexed in row-scanned order so that each spectral band is represented as a one-dimensional vector, rather than a two-dimensional image. Then, the terms on both sides of (4.17) can be stacked as

$$\mathbf{X} = \mathbf{MA} + \mathbf{N}, \tag{4.20}$$

where $\mathbf{X}$ and $\mathbf{N}$ are $K \times T$ matrices, $\mathbf{A}$ is an $L \times T$ matrix, and T is the total number of pixels in the image. In this arrangement a column of $\mathbf{X}$ is the spectrum of a specific pixel in the image and a row of $\mathbf{X}$ contains all of the pixels from one spectral band of the data, in row-scanned order. Similarly, a column of $\mathbf{A}$ describes the fractional abundances for every material in a single pixel while a row of $\mathbf{A}$ contains the fractional abundance in every pixel of a single material again in row-scanned order. Each row of $\mathbf{A}$ can be re-indexed into an image to create a material abundance map.

The hyperspectral mixing model in (4.20) is structurally similar to the linear ICA model in (4.3). The endmember matrix is analogous to the mixing matrix and the abundance matrix corresponds to the source matrix. The one difference is the addition of noise in the hyperspectral model. If the SNR is sufficiently large, this term may be safely ignored, in which case the models are identical. Otherwise, the noise effects could be minimized by smoothing, dimension reduction, or some other preprocessing step. Recall that the ICA model requires the sources to be nongaussian, implying that the fractional abundances for each material must not have a Gaussian distribution. This requirement is satisfied as abundance values tend to accumulate near zero or one depending on their spatial coverage and have a predominantly one-sided distribution. Figure 4.1 shows histograms for abundance maps of two different materials generated from a three-dimensional model of a real-world scene which demonstrate this behavior. The other requirement imposed by ICA is that the sources be independent. For the hyperspectral data model the abundance of each material is required to be independent of every other material. This requirement is violated by the additivity constraint in the linear mixing model (4.19). Although this is a violation of the ICA assumptions, as the number of endmembers and/or signature variability increases, the statistical dependence of the sources decreases and ICA performance improves [20].



(a) (b)

Fig. 4.1: Histograms of row-scanned abundance maps for (a) a sparse material, and (b) a dense material. Both of these are distributed in a way that is clearly nongaussian. Notice the change of scale in (a) required to display the non-zero abundance values. The left-most bin corresponding to zero actually extends above 16,000 pixels.

4.3 Experimental Data Description

In order to perform the kind of complementary experiments described earlier, a means of producing realistic images and the associated ground truth data is needed. This section describes the tool employed to produce the synthetic data that were incorporated into the experiments described in subsequent sections of this chapter.

The Digital Image and Remote Sensing Image Generation (DIRSIG) software is a physics-based image simulation tool developed at the Rochester Institute of Technology (RIT) [41]. The tool allows the user to describe complex scene geometry, viewing geometry, and the spectral and thermal properties of materials in the scene. The user can also describe a variety of sensor properties including sensor type, scan behavior, focal length, detector layout, and spectral and spatial response [42]. MODTRAN [43] is incorporated to simulate realistic atmospheric behavior from user-provided atmospheric and weather information. The software makes use of all of this information to generate realistic remote sensing images. Additionally, DIRSIG can also export the ground truth associated with each image.

For our experiments, two test images were generated using DIRSIG. Both images incorporate the “MegaScene” geometric scene description, which models a 0.6 square mile area of Rochester, New York. A pushbroom spectrometer model that incorporates a spectral response between 0.4 μm and 2.5 μm with 224 bands was used. The spectral response is similar to the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [44]. The altitude of the sensor was 2 km. With these settings in place, 1024×1024 pixel cubes and truth maps were generated with a ground sampling distance (GSD) of 0.25 m. These were then binned spatially to produce 128×128 pixel cubes and truth maps with a GSD of 2.0 m. The binning was performed to produce data with the desired linear mixing behavior.

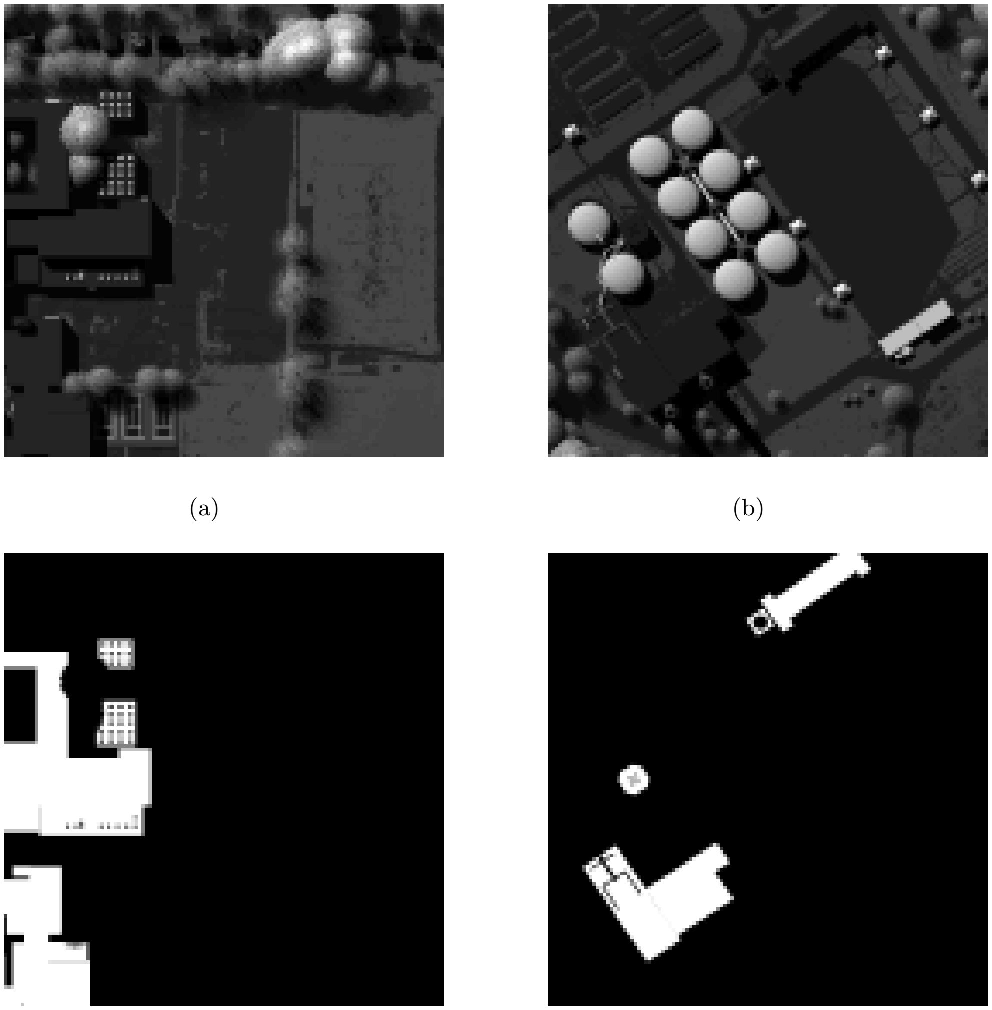
The first cube generated is referred to as “Mega1” because of its location within the first tile of the MegaScene. The scene is dominated by two large buildings surrounded by a parking lot. At the top of the image is a residential road with homes on either side that are mostly obscured by trees. Three tennis courts are located at the bottom of the image. The remainder of the scene is grass. There are 43 unique materials in this scene.

The second cube comes from the fourth MegaScene tile and is aptly named "Mega4." This scene contains ten large industrial tanks surrounded by some buildings and parking lots. Around the periphery of the scene are areas of trees and grass. This scene contains 21 unique materials. Examples of the synthetic data are shown in Fig. 4.2.

A list of the materials contained in each scene is provided in Appendix A. These materials are sorted by the number of pixels in which they appear and are loosely segregated into four categories based on their spatial coverage in the image. Super-sparse materials are those with a combined coverage of less than one pixel. Materials in the sparse category typically are present in 1% or less of the image pixels and cover less than 0.5% of the image. They may or may not appear in the image as pure pixels. Dense materials appear in over half of the pixels in the image and consequently also constitute a large number of pure pixels. Materials falling between the sparse and dense categories are classified as intermediate materials. This categorization is used to analyze how materials of varying spatial distribution are affected in the spectral unmixing process. This is an example of the type of assessment that is not ususally made in the two most common experimental scenarios referred to in Section 4.1.

4.4 Experimental Results

Three sets of experiments were performed to characterize the utility of FastICA as a spectral unmixing approach. The first set of experiments examined the impact of dimension reduction on the best-case unmixing scenario. Second, the effects of orthogonalization were explored, again considering a best-case unmixing scenario. Because dimension reduction and orthogonalization are not unique to FastICA, these two experiments are of interest beyond the scope of FastICA. In the final set of experiments, unmixing was performed using FastICA. The results of these experiments are quantified by comparing estimated material abundances with corresponding abundance ground truth. The quality of endmember extraction was not considered in these experiments. Some observations are made in the following narrative on the effects of adding noise to the synthetic images, but a characterization of the impact of noise on the unmixing process is beyond the scope of this chapter.



(a)

(b)

(c)

(d)

Fig. 4.2: Examples of the test images generated in DIRSIG. (a) RMS image of Mega1. (b) RMS image of Mega4. (c) Mega1 abundance map for “Roof, Gravel, Gray.” (d) Mega4 abundance map for “Roof, Gravel, Gray.”

For the remainder of this chapter, whenever performance is plotted versus material, i.e., the x-axis is "Material Number," the materials are numbered according to the lists in Table A.1 and Table A.2. The first (left-most) material in the plot is the most sparse and the last (right-most) is the most dense. Markers are used to denote the four categories of spatial coverage of materials. A circle ($\circ$) is used to identify super-sparse materials, an x ($\times$) for sparse materials, a diamond ($\Diamond$) for intermediate materials, and a square ($\Box$) for dense materials.

4.4.1 Computation of Optimal Estimates

Because complete ground truth abundance maps are available, the optimal, linear unmixing vector and corresponding abundance estimate can be calculated for each material. This was done prior to performing any experiments. These results constitute a best-case unmixing scenario, i.e., the best result FastICA could produce, and provide a baseline against which experimental results can be compared. A common metric used in such comparisons is mean-square error (MSE)

$$MSE \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\hat{a}(t) - a(t)\right)^2, \tag{4.21}$$

where $\hat{a}(t)$ is an estimated abundance and $a(t)$ is the ground truth abundance. However, MSE is not invariant to scaling, which is essential when considering ICA outputs, since they are subject to scale ambiguity. Thus, a preferred metric to MSE is the correlation coefficient, defined as

$$r(\hat{\mathbf{a}}, \mathbf{a}) \equiv \left(\frac{\hat{\mathbf{a}} - \mu_{\hat{\mathbf{a}}}}{\|\hat{\mathbf{a}} - \mu_{\hat{\mathbf{a}}}\|}\right)^T \left(\frac{\mathbf{a} - \mu_{\mathbf{a}}}{\|\mathbf{a} - \mu_{\mathbf{a}}\|}\right), \tag{4.22}$$

where $\mu_{\hat{\mathbf{a}}}$ and $\mu_{\mathbf{a}}$ are the sample means of $\hat{\mathbf{a}}$ and $\mathbf{a}$, respectively. The absolute value of this metric is invariant to scaling of the arguments, as desired. Conveniently, it also always falls in the range $[0, 1]$. It is used throughout the remaining experiments to quantify performance.

The unmixing formula (4.5) in combination with the linear mixing model for hyperspectral data (4.17) provides a formula for extracting individual abundances, $\hat{a}_i(t) = \mathbf{b}_i^T \mathbf{x}(t)$.

Stacking this result to eliminate the spatial indexing yields $\hat{\mathbf{a}}_i^T = \mathbf{b}_i^T\mathbf{X}$. The unmixing vector that maximizes $r(\hat{\mathbf{a}}, \mathbf{a})$ is given by

$$\check{\mathbf{b}} = \left[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{X} - \mu_{\mathbf{X}})^T\right]^{-1} (\mathbf{X} - \mu_{\mathbf{X}}) \frac{(\mathbf{a} - \mu_{\mathbf{a}})}{\|\mathbf{a} - \mu_{\mathbf{a}}\|}. \tag{4.23}$$

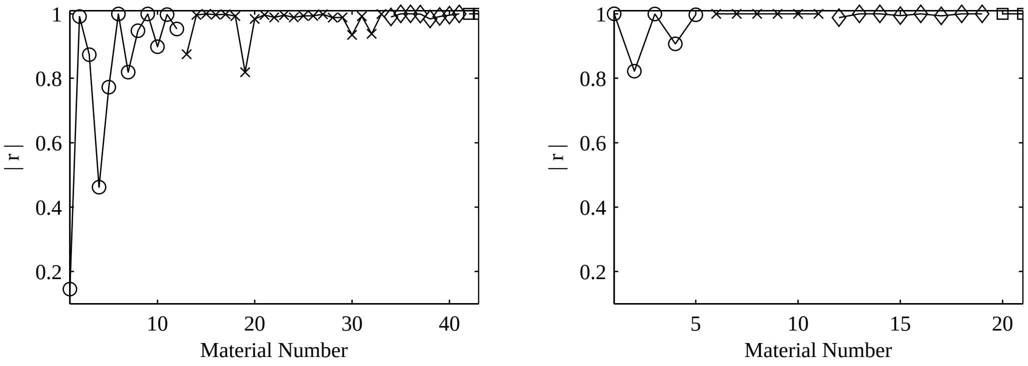
The optimal abundance estimate is then

$$\check{\mathbf{a}}^T = \check{\mathbf{b}}^T\mathbf{X}. \tag{4.24}$$

The optimal unmixing vectors and abundance estimates were calculated according to (4.23) and (4.24), respectively, for every material in both of the test cubes. In the absence of noise, as shown in Fig. 4.3, the maximum correlation coefficient, $r(\check{\mathbf{a}}_i, \mathbf{a}_i)$, overall is very high. It can be seen that the correlation coefficient tends to improve with an increase in spatial coverage. The fact that the correlation coefficient is not exactly one for every material in the scene stems from illumination, endmember, and atmospheric variability in the DIRSIG-generated cubes. Figure 4.4 provides a visual comparison between ground truth and optimal estimates from Mega1 for one material from each of the four material coverage classifications. From these images it can be seen that material locations can be clearly discerned for values of $|r| \geq 0.8$. Below this threshold, the material locations are less clear and background artifacts become more obvious. Depending on the spatial coverage and congruency of a material, correlation coefficient values as low as 0.5 may be useful.

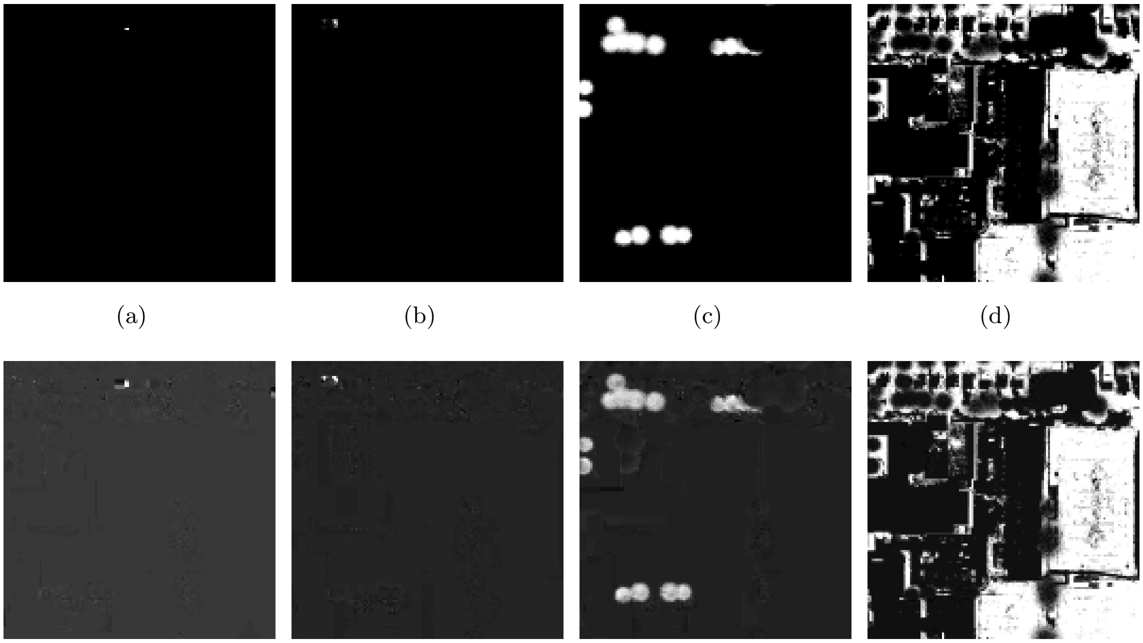
4.4.2 Dimension Reduction

Because it is typically used as a preprocessing step in a variety of spectral unmixing approaches, including FastICA, an experiment was performed to examine the effect of dimension reduction on the best-case unmixing scenario. To do this, the maximum correlation abundance estimates were calculated using dimension reduced data obtained from PCA. The same maximum correlation formulas (4.23) and (4.24) were used, replacing $\mathbf{X}$ with the dimension-reduced data, $\mathbf{X}_N$, given by $\mathbf{X}_N = \mathbf{V}_N^T\mathbf{X}$, where $\mathbf{V}_N$ is the $K \times N$



(a) (b)

Fig. 4.3: Correlation coefficient between optimal estimates and corresponding ground truth. (a) Mega1 results. (b) Mega4 results. Note that Mega1 contains twice as many materials as Mega4.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 4.4: A comparison of material truth maps (first row) with their maximum correlation estimates (second row). (a) and (e) Material 4, Siding, Cedar, Stained Dark Brown, Fair, $r = 0.4617$. (b) and (f) Material 19, Roof Shingle, Asphalt, Eclipse Sample Board, Twilight Gray, $r = 0.8185$. (c) and (g) Material 38, Tree, Norway Maple, Leaf, $r = 0.9840$. (d) and (h) Material 43, Grass, Brown and Green w/ Dirt, $r = 0.9999$.

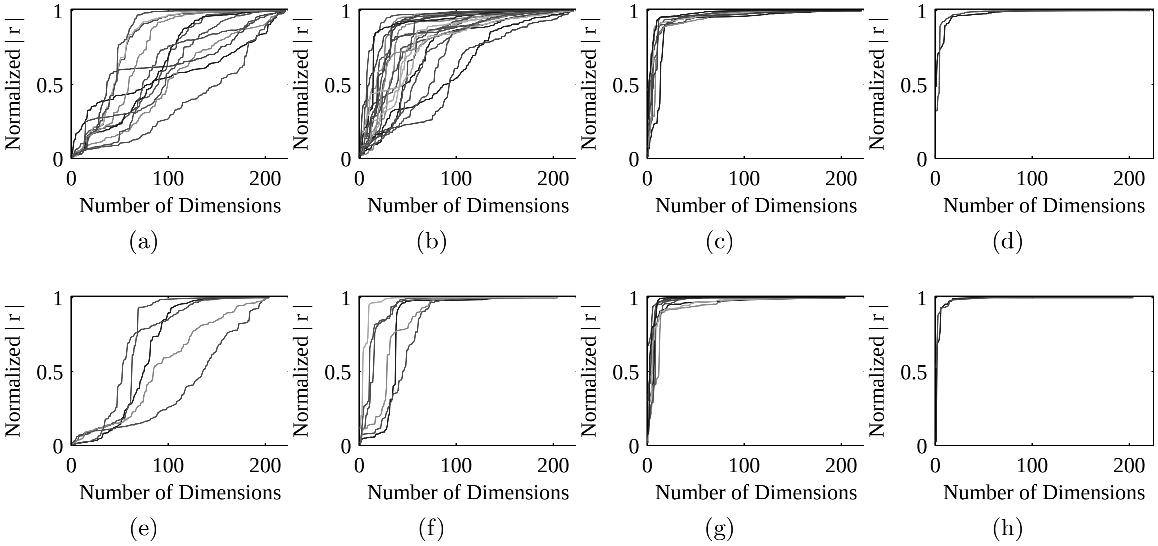
whitening matrix associated with the N most energetic principal components of $\mathbf{X}$.

The results of this experiment are shown in Fig. 4.5 and Table 4.1. The plots in Fig. 4.5 demonstrate how the correlation coefficient of the optimal estimate with the ground truth decreases as the dimensionality of the data is reduced. The correlation coefficient (y-axis) in these plots is normalized by the correlation coefficient obtained when there has been no dimension reduction. The slope of each curve illustrates the contribution of individual principal components to the correlation coefficient of the optimal estimates for a specific material. It is clear from the sharp jumps in the correlation for the dense and intermediate materials that they are well described by the first several principal components. What is also clear is that there is no similar region for the sparse and super-sparse materials. The information associated with these materials appears to be almost uniformly scattered across all of the principal components. For this reason, a relatively large number of dimensions must be retained to achieve near-optimal estimates of these materials. Table 4.1 underscores this conclusion, showing the average number of dimensions that must be kept to obtain 95% and 75% levels of the correlation coefficient obtained when no dimension reduction was performed.

One approach to determining the number of dimensions that should be retained when doing PCA is to keep as many dimensions as are needed to retain some percentage of the total variance in the image. Retaining 99.9% of the total variance in the Mega1 and Mega4 images requires only six and five dimensions, respectively. Based on the results in Table 4.1, that would allow only the dense materials to be extracted at near-optimal levels.

Table 4.1: Number of dimensions necessary to obtain 95% and 75% levels of optimal correlation, by material classification.

	Mega1		Mega4	
	95%	75%	95%	75%
super-sparse	150	110	138	102
sparse	116	53	46	29
intermediate	36	10	23	8
dense	16	6	12	4



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 4.5: Normalized correlation coefficient of the maximum correlation estimates obtained using dimension reduced data. The first row shows the Mega1 results and the second shows the results for Mega4. (a) and (e) Super-sparse materials. (b) and (f) Sparse materials. (c) and (g) Intermediate materials. (d) and (h) Dense materials.

4.4.3 Orthogonalization

Next, the effect of constraining the unmixing vectors to be orthogonal was examined. Because the PCA and whitening step decorrelates the observed data, it is expected that the unmixing vectors for the whitened data should be orthogonal. In the FastICA implementation, this constraint is enforced on the unmixing vectors at the end of each iteration of the cost function optimization.

To apply the orthogonality constraint to the optimal unmixing vectors requires a minor modification to the orthogonalization formula, since the optimal vectors were not calculated using whitened data. When the data are not whitened, the formulas for deflationary orthogonalization (4.14) and (4.15) become

$$\widetilde{\mathbf{b}}_i = \mathbf{b}_i - \sum_{j=1}^{i-1}(\mathbf{b}_i^T \mathbf{C}_{\mathbf{x}} \widetilde{\mathbf{b}}_j)\widetilde{\mathbf{b}}_j, \tag{4.25}$$

and

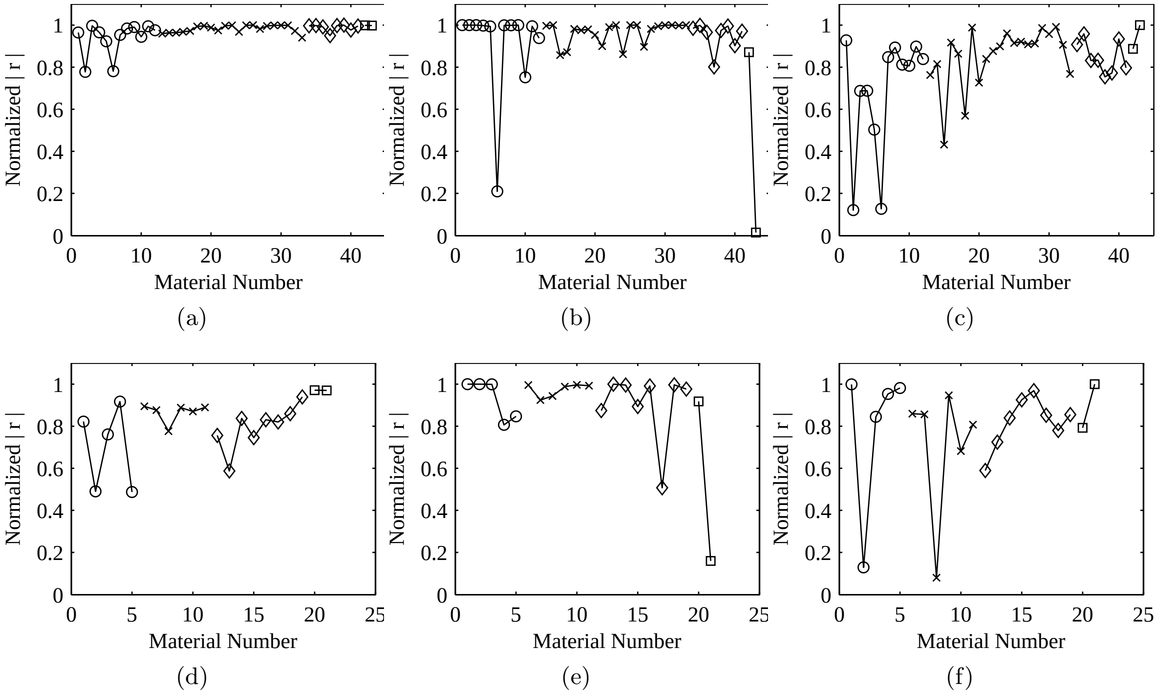
$$\widetilde{\mathbf{b}}_i = \frac{\widetilde{\mathbf{b}}_i}{\sqrt{\widetilde{\mathbf{b}}_i^T \mathbf{C}_\mathbf{x} \widetilde{\mathbf{b}}_i}}, \tag{4.26}$$

respectively, where $\mathbf{C}_\mathbf{x}$ is the covariance matrix of $\mathbf{X}$. The symmetric orthogonalization formula (4.16) changes in a similar way,

$$\widetilde{\mathbf{B}} = \left(\mathbf{B}\mathbf{C}_\mathbf{x}\mathbf{B}^T\right)^{-1/2} \mathbf{B}. \tag{4.27}$$

These changes result from the fact that orthogonality of the unmixing vectors of whitened data is equivalent to $\mathbf{B}\mathbf{C}_\mathbf{x}\mathbf{B}^T = \mathbf{I}$, where $\mathbf{B}$ contains the unmixing vectors of the unwhitened data.

The optimal unmixing vectors calculated by (4.23) were forced to be orthogonal using the formulas above. Abundance estimates were then calculated from the orthogonalized



(a) (b) (c)

(d) (e) (f)

Fig. 4.6: Normalized correlation coefficient of estimates obtained by orthogonalizing the optimal unmixing vectors for Mega1 (first row) and Mega4 (second row). (a) and (d) Symmetric orthogonalization. (b) and (e) Deflationary orthogonalization (sparse to dense). (c) and (f) Deflationary orthogonalization (dense to sparse).

vectors. The effect on the correlation coefficient of the estimates due to orthogonalization is shown in Fig. 4.6. Because the deflationary orthogonalization approach is sequential, the ordering of the vectors matters. The deflation was performed in both ascending and descending material order (most sparse to most dense and vice versa). As would be expected, the results show that better estimates are obtained for those materials that are used earlier in the deflation process. So, to obtain better estimates of a material, it would be desirable for the cost function optimization algorithm to extract the unmixing vector corresponding to that material before any others. The results also show that deflating the estimates for more sparse materials first has less of an effect on the more dense materials than deflating in the opposite order. The symmetric approach is something of a compromise, balancing the negative effects of the orthogonalization across all of the materials.

The results show that, in most cases, orthogonalization does not cause significant degradation of the estimates. This is true even in the presence of additive noise. There are a few exceptions, however, where the degradation is noticeable. Obvious examples of this are materials two and six in the Mega1 results. When symmetric orthogonalization is used, both show an appreciable decrease from the optimal correlation. When the ascending deflationary approach is used, material two is fine, but material six shows significant loss. Both are affected when the deflation is performed in descending material order. This behavior implies that there must be some information shared between the two materials. Thus, if material two is extracted first, it leads to a degradation when extracting material six and vice versa. Both experience degradation when the symmetric approach is used. This pattern can be explained by looking at an image representation of the matrix $\mathbf{B}\mathbf{C}_{\mathbf{x}}\mathbf{B}^T$, shown in Fig. 4.7. If the materials were truly uncorrelated when whitened, then the image would be that of a diagonal matrix with white pixels on the diagonal and the remainder black. But, the off-diagonal bright spots in Fig. 4.7 are indications of correlation between the optimal unmixing vectors, even when the data are whitened. It turns out that material two only shows up in one pixel and material six only shows up in two pixels, one of which is shared with the lone material two pixel. Wherever there is a drop in correlation due to orthogonalization similar results are found, i.e., a more sparse material shows up entirely in a subset of the pixels containing a more dense material. In these cases the additivity constraint in (4.19) leads to stronger correlation than for those materials which share pixels with many different materials. So, while it is true that as the number of endmembers in the data increases, the statistical dependence among sources decreases and ICA performs better [20], co-located materials with limited spatial coverage will still be poorly estimated.

4.4.4 FastICA Performance

As a final experiment, FastICA was used to generate abundance maps for the Mega1 and Mega4 data. Each of the three cost functions in (4.8)–(4.10) was considered, as well as both symmetric and deflationary orthogonalization. The number of components was left to be determined by the algorithm. In each case the algorithm was initialized with a random matrix. As noted earlier, there is a scale ambiguity associated with the FastICA outputs. To be useful in abundance quantification these outputs should fall in the range $[0, 1]$. The best method of rescaling the outputs is not explored in this chapter. Instead, a metric that is invariant to scale is used to assess the results.

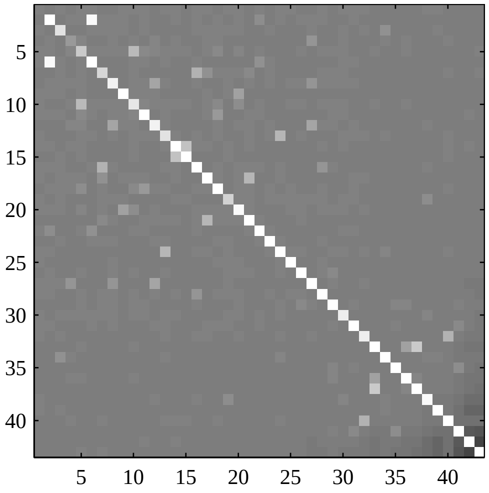


Fig. 4.7: An image representation of the correlation coefficient of the optimal unmixing vectors for Mega1. Off-diagonal bright spots indicate correlation between the vectors, despite whitening. Notice the dark area in the bottom-right of the image due to the negative correlation between the dense materials.

Because the number of components was not specified, more independent components were generated than there are materials in the scene. A method has been proposed for prioritizing independent components obtained from hyperspectral data [23]. For this experiment the normalized correlation coefficient of every independent component with every material ground truth was calculated, and the maximum was retained for each material. These results are shown in Fig. 4.8. Average performance across material classifications is shown in Table 4.2. Generally, it appears that no single cost function or orthogonalization approach is vastly superior to any other. For extracting dense materials, it seems that the pow3 cost function should be avoided and that deflationary orthogonalization usually outperforms symmetric. This might imply that dense materials tend to be found earlier than materials from other categories. For sparse and super-sparse materials only the gauss cost function combined with symmetric orthogonalization gave consistently poor results.

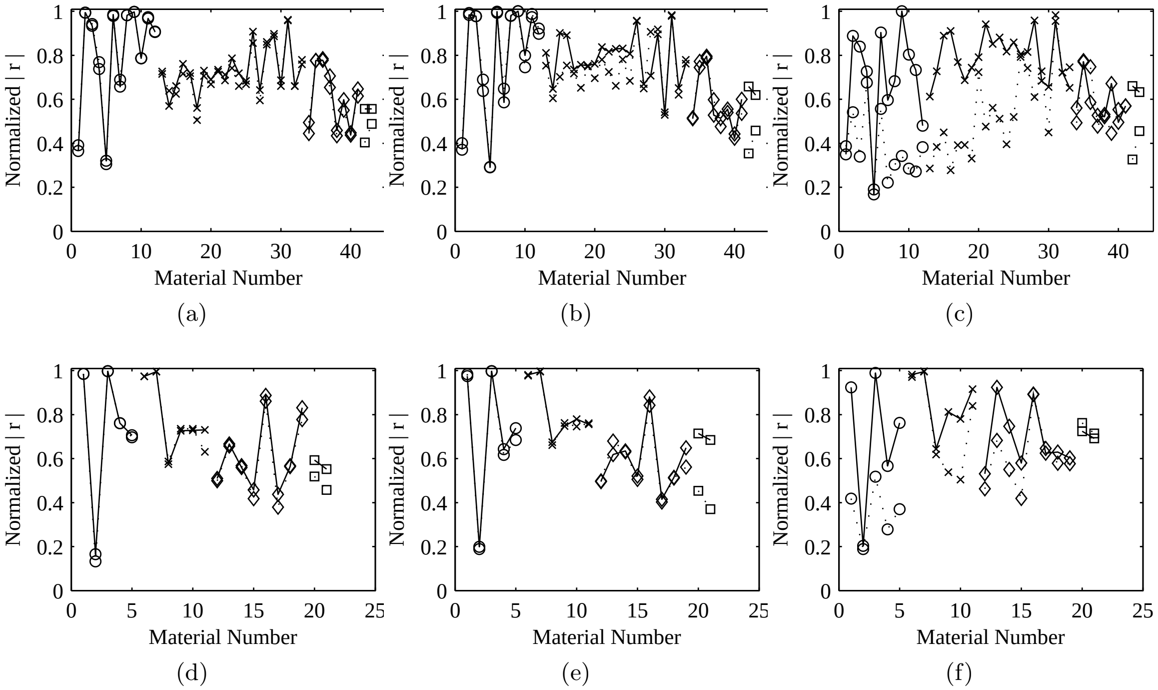


Fig. 4.8: Normalized correlation coefficient of estimates obtained using FastICA for Mega1 (first row) and Mega4 (second row). The deflationary orthogonalization results are shown with a solid line, symmetric orthogonalization with a dotted line. (a) and (d) Cost function "pow3" described by (4.8) and (4.11). (b) and (e) Cost function "tanh" described by (4.9) and (4.12). (c) and (f) Cost function "gauss" described by (4.10) and (4.13).

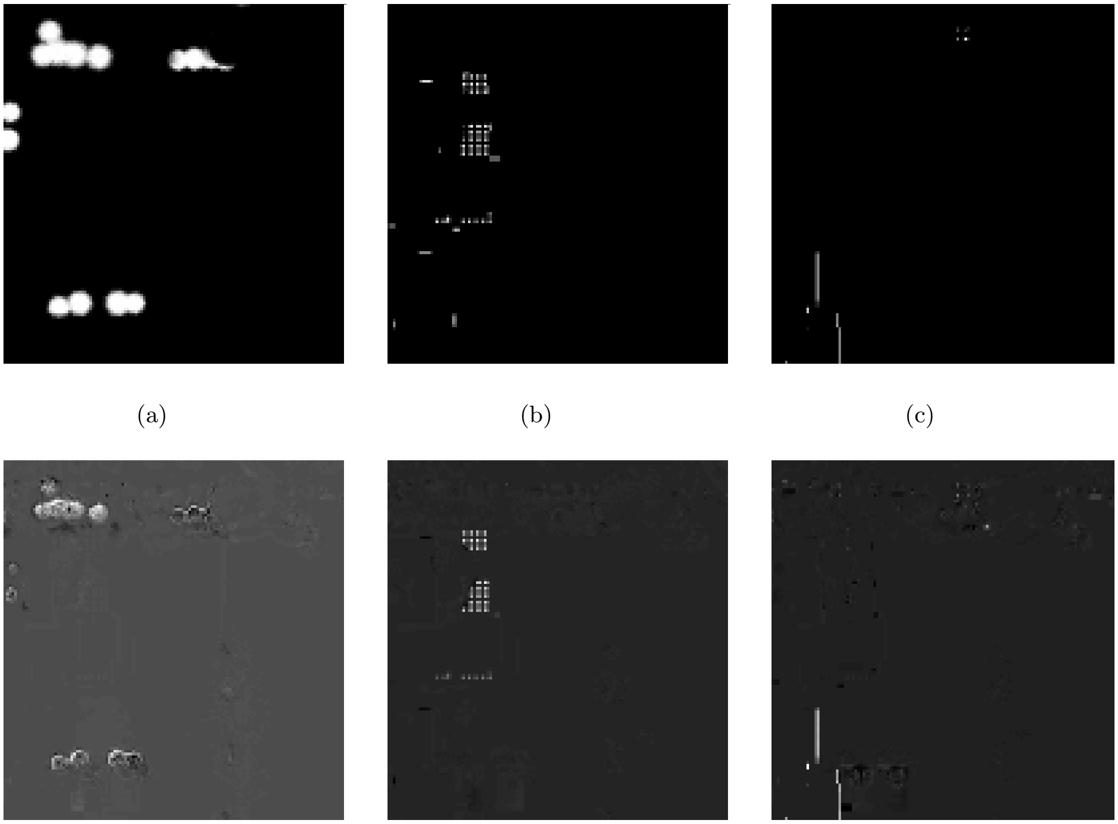
Table 4.2: Average normalized correlation coefficient FastICA based on material classification.

	pow3		tanh		gauss	
	defl	symm	defl	symm	defl	symm
Mega1, super-sparse	0.8027	0.8095	0.7916	0.8023	0.6808	0.3745
Mega4, super-sparse	0.7229	0.7146	0.7106	0.6943	0.6893	0.3547
Mega1, sparse	0.7303	0.7191	0.7871	0.7435	0.7966	0.5463
Mega4, sparse	0.7882	0.7765	0.8222	0.8174	0.8551	0.7446
Mega1, intermediate	0.6055	0.5945	0.5841	0.5837	0.5876	0.5745
Mega4, intermediate	0.6140	0.5909	0.5915	0.5787	0.6912	0.6012
Mega1, dense	0.5566	0.4660	0.6390	0.4060	0.6467	0.3910
Mega4, dense	0.5729	0.4881	0.6996	0.4120	0.7081	0.7384
Mega1, all	0.7192	0.7084	0.7437	0.7145	0.7184	0.4964
Mega4, all	0.6858	0.6636	0.6960	0.6586	0.7392	0.5965

Three ground truth images as well as the independent components most strongly correlated with them are shown in Fig. 4.9. The correlation coefficients of the truth maps and estimates are, from left to right, $|r| = 0.5054$, $|r| = 0.7443$, and $|r| = 0.8853$. These values are not normalized by the best-case coefficients. The images give an idea of the quality of the unmixed data for a range of correlation coefficients.

The images in the first row of Fig. 4.10 show two independent components obtained using FastICA to unmix the Mega1 data. They illustrate two interesting features that have been frequently noticed in the FastICA output. First, there is an intensity gradient across the horizontal dimension of the images. The DIRSIG tool uses a pushbroom sensor model to generate these data with the sensor moving from bottom to top. So, FastICA seems to be extracting information that is associated with the view angle of the sensor and/or path length. Further examination of the associated endmember and atmosphere data is needed to determine exactly what is being highlighted in this gradient. ICA has been shown to extract components corresponding to solar angle effect [16], and this may be something similar.

The second observation is that these two components are both strongly correlated to the same material. The correlation coefficient of the first with the truth map is $|r| = 0.6011$.



(a)

(b)

(c)

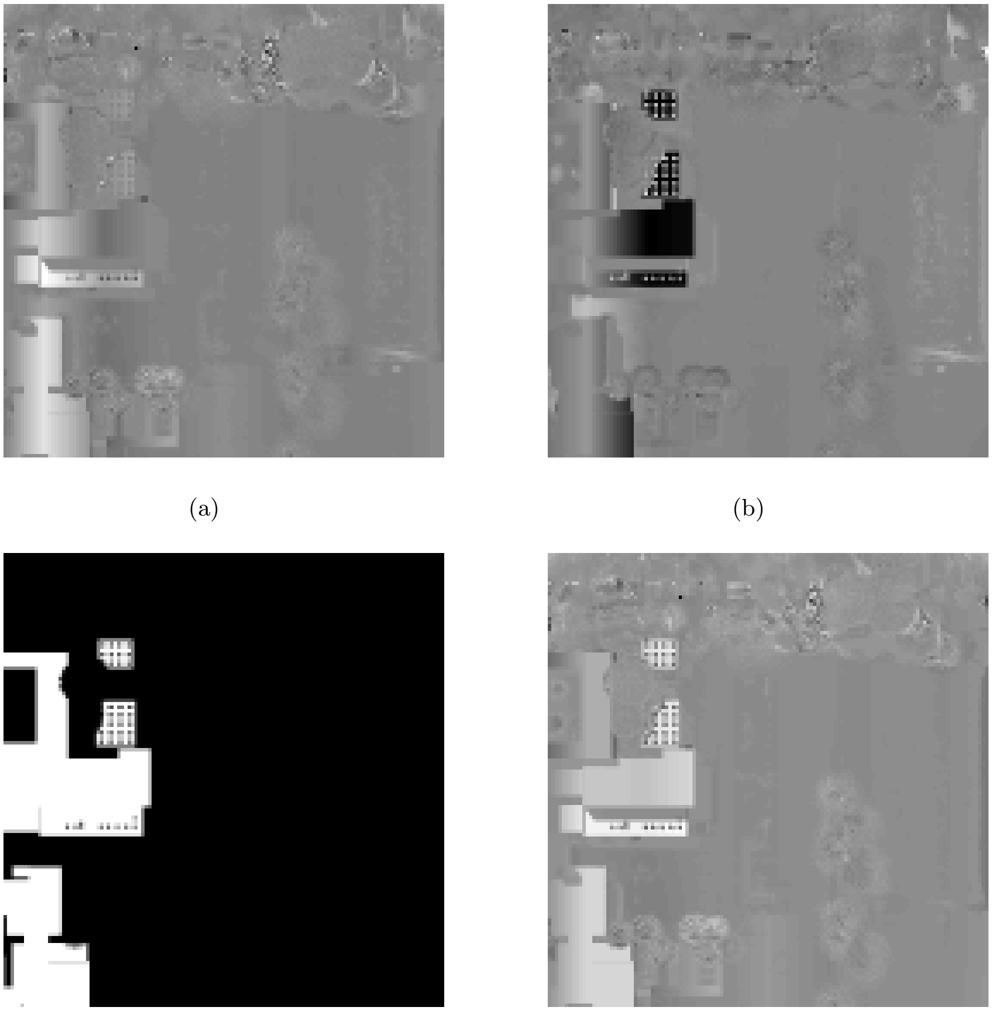
(d)

(e)

(f)

Fig. 4.9: Material truth maps from Mega1 (first row) and the independent components most correlated with them (second row). (a) and (d) Tree, Norway Maple, Leaf truth map and best estimate, $|r| = 0.5054$. (b) and (e) Sheet Metal, White, Fair truth map and best estimate, $|r| = 0.7443$. (c) and (f) Brick, Brampton Brick, Old School, Brown, truth map and best estimate, $|r| = 0.8853$.

For the second, $|r| = 0.4673$. A linear combination of the two can be used to produce the image in Fig. 4.10(d) for which $|r| = 0.7606$. This splitting of a single material into two components seems to occur frequently. The clustering of independent components of hyperspectral images has been examined [45], but an attempt to automate and optimize the process based on results from synthetic data remains a future research effort.



(a) (b) (c) (d)

Fig. 4.10: Two independent components, (a) and (b), that are strongly correlated to the same truth map, shown in (c). A linear combination of the two, (d), provides an improvement to the correlation coefficient.

4.5 Conclusion

The utility of realistic, but synthetic data to assess spectral unmixing approaches was demonstrated using two hyperspectral images generated by DIRSIG. A number of experiments which use this data were then described. In the first, the effect of dimension reduction using PCA was quantified. This experiment demonstrated that to achieve near-optimal results, more dimensions needed to be retained than would be expected based on an analysis of eigenvalues. Just how many more dimensions are necessary depends on the spatial distribution of the materials of interest. In the second experiment the impact of orthogonalization was considered. The impact was found to be minimal except in the case where sparsely distributed materials were found to be consistently co-located. The method of orthogonalziation as well as the order of material extraction determines the severity of the effect. In the final experiment it was shown that FastICA is effective at unmixing some, but not all materials. This is in agreement with an assessment made in existing literature [20]. However, this complementary experimental approach allowed for the identification of a splitting behavior in which FastICA produces multiple outputs containing distinct pieces of a common material. It was shown that these outputs can be merged in a way that produces improved results. The automation of identifying and merging these outputs is an area of future research.

References

[1] M. S. Kim, A. Lefcourt, K. Chao, Y. Chen, I. Kim, and D. Chan, "Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part I. Application of visible and near-infrared reflectance imaging," *Transactions of the American Society of Agricultrual and Biological Engineering*, vol. 45, no. 6, pp. 2027–2037, 2002.

[2] M. Moran, Y. Inoue, and E. Barnes, "Opportunities and limitations for image-based remote sensing in precision crop management," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 319–346, 1997.

[3] F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1388–1400, June 2003.

[4] T. Schmid, M. Koch, J. Gumuzzio, and P. M. Mather, "A spectral library for a semi-arid wetland and its application to studies of wetland degradation using hyperspectral and multispectral data," *International Journal of Remote Sensing*, vol. 25, pp. 2485–2496, July 2004.

[5] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.

[6] N. Keshava, J. P. Kerekes, D. G. Manolakis, and G. A. Shaw, "Algorithm taxonomy for hyperspectral unmixing," *Proceedings of SPIE*, vol. 4049, pp. 42–63, 2000.

[7] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, pp. 23–26, 1995.

[8] R. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," *Proceedings of 21st Canadian Symposium on Remote Sensing*, pp. 891–897, 1999.

[9] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sept. 2002.

[10] M. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

[11] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE*, vol. 3753, pp. 266–275, 1999.

[12] 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, Mar. 1999.

[13] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[14] D. R. Fuhrmann, "A simplex shrink-wrap algorithm," *Proceedings of SPIE*, vol. 3718, pp. 501–511, 1999.

[15] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," *IEEE Geoscience and Remote Sensing Symposium*, pp. 4033–4036, July 2007.

[16] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10-12, pp. 2194–2208, June 2008.

[17] A. Cichocki and S. ichi Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* West Sussex, UK: John Wiley & Sons, 2002.

[18] J. Bayliss, J. A. Gualtieri, and R. F. Cromp, “Analyzing hyperspectral data with independent component analysis,” *Proceedings of SPIE*, vol. 3240, pp. 133–143, 1997.

[19] X. Zhang and C. H. Chen, “New independent component analysis method using higher order statistics with application to remote sensing images,” *Optical Engineering*, vol. 41, pp. 1717–1728, 2002.

[20] J. Nascimento and J. Dias, “Does independent component analysis play a role in unmixing hyperspectral data?” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[21] T. Tu, “Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis approach,” *Optical Engineering*, vol. 39, no. 4, pp. 897–906, Apr. 2000.

[22] B. R. Foy and J. Theiler, “Scene analysis and detection in thermal infrared remote sensing using independent component analysis,” *Proceedings of SPIE*, vol. 5439, pp. 131–139, 2004.

[23] J. Wang and C.-I. Chang, “Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 260–2616, Sept. 2006.

[24] C. Shah, M. Arora, S. Robila, and P. Varshney, “ICA mixture model based unsupervised classification of hyperspectral imagery,” *Proceedings of Applied Imagery Pattern Recognition Workshop*, pp. 29–35, Oct. 2002.

[25] S. Chiang, C.-I. Chang, and I. Ginsberg, “Unsupervised hyperspectral image analysis using independent component analysis,” *IEEE Geoscience and Remote Sensing Symposium*, vol. 7, pp. 3136–3138, 2000.

[26] 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. 650–663, Mar. 2004.

[27] A. Plaza and C.-I. Chang, “Impact of initialization on design of endmember extraction algorithms,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3397–3407, Nov. 2006.

[28] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. Mckeown, V. Iragui, and T. J. Sejnowski, “Removing electroencephalographic artifacts by blind source separation,” *Psychophysiology*, vol. 37, no. 02, pp. 163–178, 2000.

[29] A. Cichocki, S. L. Shishkin, T. Musha, Z. Leonowicz, T. Asada, and T. Kurachi, “EEG filtering based on blind source separation (BSS) for early detection of alzheimer’s disease,” *Clinical Neurophysiology*, vol. 116, no. 3, pp. 729–737, 2005.

[30] J. Joutsensalo and T. Ristaniemi, “Learning algorithms for blind multiuser detection in CDMA downlink,” *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 1040–1044, Sept. 1998.

[31] T. Ristaniemi and J. Joutsensalo, “On the performance of blind source separation in CDMA downlink,” *Proceedings of the International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pp. 437–441, 1999.

[32] A. Back and A. Weigend, “A first application of independent component analysis to extracting structure from stock returns,” *International Journal of Neural Systems*, 1997.

[33] S. Cha and L. Chan, “Applying independent component analysis to factor model in finance,” *Intelligent Data Engineering and Automated Learning-IDEAL 2000. Data Mining, Financial Engineering, and Intelligent Agents*, pp. 161–173, 2009.

[34] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. Boston: McGraw-Hill, 2002.

[35] H. H. Yang and S. Amari, “Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information,” *Neural Computation*, vol. 9, no. 7, pp. 1457–1482, Oct. 1997.

[36] J. F. Cardoso and A. Souloumiac, “Blind beamforming for non gaussian signals,” *IEE Proceedings F, Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[37] P. Pajunen, A. Hyvarinen, and J. Karhunen, “Nonlinear blind source separation by self-organizing maps,” *Proceedings of International Conference on Neural Information Processing*, pp. 1207–1210, 1996.

[38] P. Pajunen and J. Karhunen, “A maximum likelihood approach to nonlinear blind source separation,” *Lecture Notes in Computer Science*, pp. 541–546, 1997.

[39] N. Keshava and J. Mustard, “Spectral unmixing,” *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[40] R. B. Singer and T. B. McCord, “Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance,” *Proceedings of Lunar and Planetary Science Conference*, pp. 1835–1848, 1979.

[41] *The DIRSIG User's Manual*, Rochester Institute of Technology, Oct. 2006. [www.dirsig.org/documentation]

[42] E. J. Ientilucci and S. D. Brown, “Advances in wide-area hyperspectral image simulation,” *Proceedings of SPIE*, pp. 110–121, 2003.

[43] A. Berk, L. S. Bernstein, G. P. Anderson, P. K. Acharya, D. C. Robertson, J. H. Chetwynd, and S. M. Adler-Golden, “MODTRAN cloud and multiple scattering upgrades with application to AVIRIS,” *Remote Sensing of Environment*, vol. 65, no. 3, pp. 367–375, 1998.

[44] R. Green, M. Eastwood, C. Sarture, T. Chrien, M. Aronsson, B. Chippendale, J. Faust, B. Pavri, C. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[45] S. Rajan, "Unmixing of hyperspectral data using independent component analysis," Master's thesis, Utah State University, Logan, UT, 2005.

Chapter 5

An Algorithm to Rescale Independent Components for Abundance Quantification[1]

5.1 Introduction

Hyperspectral imaging is a remote sensing approach that simultaneously collects both spatial and spectral data. Spectral data are collected in hundreds of narrow contiguous bands that may cover the visible, near-infrared, short-wave infrared (0.4–2.5 μm), the mid-wave infrared (3–5 μm), and/or the long-wave infrared (8–14 μm). Because of their high spectral fidelity, hyperspectral images are especially useful in discriminating between the materials in a scene. Although the size of a pixel on the ground varies, spatial measurements typically consist of hundreds of pixels in both spatial dimensions. Such images contain a wealth of information and have found application in a broad range of fields such as food safety [1], agriculture [2], mineralogy [3], ecology [4], and target detection [5], as well as many others.

Spectral unmixing is a common hyperspectral exploitation approach that extracts information about the constituent materials in an image. It is the process of decomposing a hyperspectral image into two products. The first is a set of endmembers, which are the spectra that are representative of the materials (or classes of materials) in a scene. The second spectral unmixing product is a set of abundance maps which are images that show the proportion of each endmember in each pixel of the image.

There are a wide variety of algorithms that have been developed to unmix hyperspectral data. Some of these attempt to find the most “extreme” pixels in the image, based on some definition of distance or extremity. These include pixel purity index (PPI) [6],

---

[1] M.R. Stites, J.H. Gunther, T.K. Moon, and G.P. Williams (to be submitted for publication)

iterative error analysis (IEA) [7], and automated morphological endmember extraction (AMEE) [8]. Other, similar approaches exploit the geometry imposed by a linear mixing model (LMM). Minimum volume transformation (MVT) [9], N-FINDR [10], convex cone analysis (CCA) [11], vertex component analysis (VCA) [12], and shrink-wrapping [13] are all algorithms of this type. Statistical approaches which assume specific probability distributions for the observed hyperspectral data and/or abundances have also been developed, such as dependent component analysis (DECA) [14], and Bayesian positive source separation (BPSS) [15].

A statistical unmixing approach that does not assume a specific distribution for the data is independent component analysis (ICA) [16]. This approach attempts to unmix the data by finding maximally independent abundances. A variety of ICA algorithms have been applied to hyperspectral unmixing including contextual ICA [17], joint cumulant-based ICA [18], joint approximate diagonalization of eigen-matrices (JADE) [19], and FastICA [19–22]. ICA has also been employed as a hyperspectral classification approach [23, 24].

Traditionally, an endmember is expected to represent a general class of spectra, rather than a specific material. Endmembers of this nature are expected to be spatially dominant within the scene. A traditional set of endmembers is relatively small, e.g., grass, trees, roads, and soil. ICA is one of the few spectral unmixing approaches that operates by exploiting abundance properties, rather than endmember properties. Because of this, it can discriminate materials with limited spatial distribution—anomalies—that would not normally be considered endmembers. Applying ICA to the spectral unmixing problem allows for a larger, more varied, and more specific set of endmembers.

There are potential problems associated with using ICA to perform spectral unmixing. One problem is the ICA requirement that the abundances be statistically independent. The linear mixing model that is at the heart of most spectral unmixing approaches guarantees that the abundances are, in fact, statistically dependent [25]. However, it has been shown that this statistical dependence decreases as the number and variability of endmembers increases, and useful results can be obtained from ICA [19].

A second problem arises due to the fact that scalar multiplication does not affect the statistical independence of two random variables. Thus, ICA results are subject to scale (and sign) ambiguity. This ambiguity requires ICA outputs to undergo some sort of rescaling in order to be meaningful. A linear rescaling based on finding the minimum and maximum values of an ICA output has been proposed [22]. However, this approach can perform poorly in the presence of outliers in the ICA output or when there is an unexpected offset in the ICA output. To remedy these shortcomings, a new, nonlinear rescaling approach is presented in this chapter that is based on a statistical model for abundance values. When estimated abundances are compared to true abundances, rescaling based on this model shows significant improvement over other methods. This improvement is observed both in the reduction of mean-square error (MSE) and the increase of the correlation coefficient.

This chapter is organized as follows. The abundance model is constructed in Section 5.2. The rescaling algorithm that has been developed based on the proposed abundance model is described in Section 5.3. Section 5.4 contains experimental description and results. Section 5.5 provides some final observations and conclusions.

5.2 Abundance Model

The development of the abundance model is based on the simple observation that there are three types of pixels in an abundance map. Empty pixels do not contain the material and always take an abundance value of zero. Filled pixels contain only the material and always take an abundance value of one. Finally, mixed pixels are partially filled by the material and have an abundance value between zero and one. This model can be stated probabilistically as

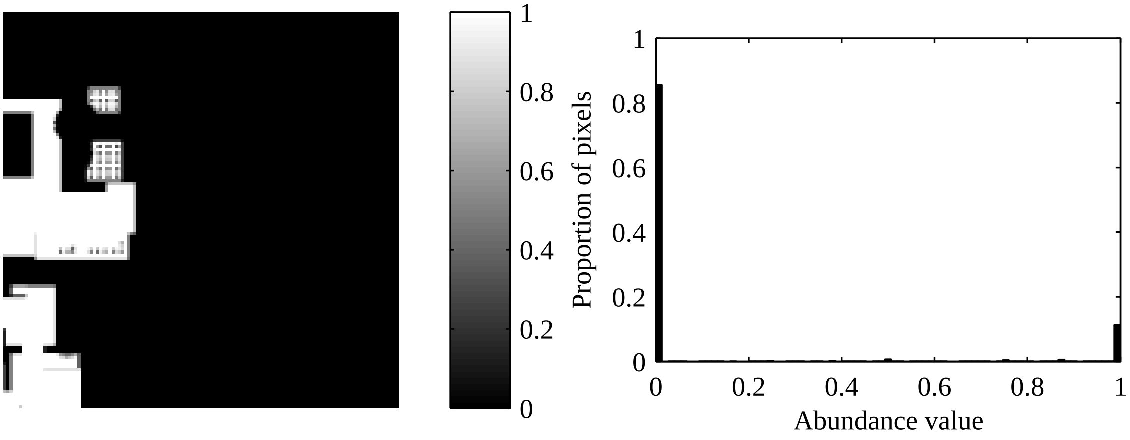
$$p_{\mathrm{ideal}}(a) = P_{\mathrm{e}}\delta(a) + P_{\mathrm{f}}\delta(1-a) + P_{\mathrm{m}}f_{\mathrm{m}}(a), \tag{5.1}$$

where P_{e}, P_{f}, and P_{m} are the proportions of empty, filled, and mixed pixels, respectively, and $f_{\mathrm{m}}(a)$ is a probability density function (pdf) that describes the distribution of the abundance values of mixed pixels. Since every pixel must be either empty, filled, or mixed, P_{e}, P_{f}, and P_{m} must sum to one. The continuous uniform and beta distributions have both been considered as the mixed pixel distribution, $f_{\mathrm{m}}(a)$. The uniform distribution with pdf given by, $f_u(x) = 1,\ 0 < x < 1$, is simple, making it easy to work with. The beta distribution with pdf

$$f_\beta(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1},\ 0 < x < 1, \tag{5.2}$$

is much more complicated, but allows for much more variability in the shape of the distribution.

Figure 5.1(a) shows an abundance map obtained from DIRSIG, a synthetic imagery generation tool [26, 27]. Although the imagery is synthetic, it is generated from a physical model of existing structures, making the abundance distributions very realistic. The material highlighted in Fig. 5.1(a) is a gray roofing gravel. A histogram of the abundances from the image is shown in Fig. 5.1(b). This distribution of abundance values shows large values at zero and one corresponding to empty and filled pixels, and a very small percentage of mixed pixels in the image (less than 3%). Such a distribution—with a small number of mixed pixels—should be expected for any scene where the spatial coverage of a material is significantly larger on average than the spatial resolution of the sensor.



(a)

(b)

Fig. 5.1: DIRSIG-generated abundance truth data for gray roofing gravel. (a) Abundance map image. (b) Histogram of abundance values. This histogram conforms well to the ideal abundance model given by (5.1).

When ICA is used to unmix a hyperspectral image, the ideal abundances are not obtained in the independent components. Instead, each independent component is a corrupted version of the ideal abundances for a given material, modeled as

$$a_{\text{ic}} = c_1 a_{\text{ideal}} + c_0 + \nu. \tag{5.3}$$

In this model, c_1 is a constant that incorporates the scale ambiguity associated with ICA outputs. Occasionally, a small offset is present in the observed data, motivating the inclusion of the constant term, c_0, in the model. The final term, ν, is a zero-mean Gaussian random variable with variance σ^2. This term is included to account for both sensor noise and endmember variability.

The model in (5.3) implies that for an independent component both the empty and filled abundance classes have Gaussian pdfs. The mean of the empty class is at c_0, i.e.,

$$p_{\text{e}}(a) = \phi(a|c_0, \sigma^2), \tag{5.4}$$

and the mean of the filled class is at $c_1 + c_0$, i.e.,

$$p_{\text{f}}(a) = \phi(a|c_1 + c_0, \sigma^2), \tag{5.5}$$

where

$$\phi(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \tag{5.6}$$

The distribution of the mixed class is somewhat more complicated. If the ideal model assumes a uniform distribution, then the independent component model yields a distribution with the form

$$p_{\text{m}}(a) = \frac{1}{2\left(c_1 - c_0\right)} \left[\text{erf}\left(\frac{a - c_0}{\sqrt{2\sigma^2}} \right) - \text{erf}\left(\frac{a - c_1}{\sqrt{2\sigma^2}} \right) \right], \tag{5.7}$$

where $\text{erf}(\cdot)$ is the error function given by

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{5.8}$$

When the ideal model assumes a beta distribution for the mixed class, the pdf for the mixed class in the independent component model is the convolution of a zero-mean Gaussian pdf with a beta pdf, i.e.,

$$p_{\mathrm{m}}(a) = \int_0^1 f_\beta(\tau|\alpha,\beta)\phi(a-\tau|0,\sigma^2)d\tau. \tag{5.9}$$

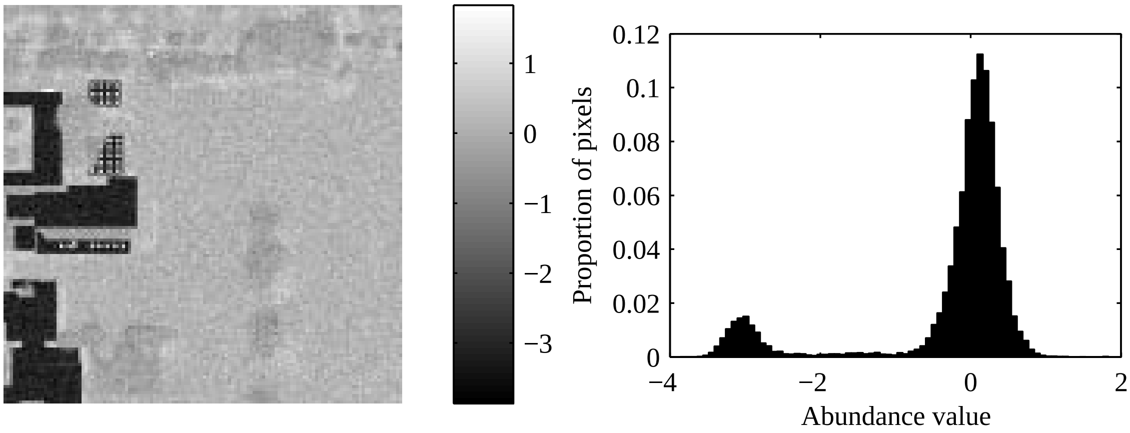
Values for this distribution can be obtained by using numerical convolution methods.

The independent component class distributions described above can be combined with the ideal abundance model (5.1) to produce a pdf for the independent component as

$$p_{\mathrm{ic}}(a) = P_{\mathrm{e}}p_{\mathrm{e}}(a) + P_{\mathrm{f}}p_{\mathrm{f}}(a) + P_{\mathrm{m}}p_{\mathrm{m}}(a). \tag{5.10}$$

This distribution is just a sum of the class distributions scaled by the relative class proportions in the ideal model.

Figure 5.2(a) shows an independent component that is strongly correlated to the abundance map shown in Fig. 5.1(a). The corresponding histogram is shown in Fig. 5.2(b). The histogram of the independent component shows good agreement with the the independent component model (5.10).



(a) (b)

Fig. 5.2: Independent component obtained using ICA corresponding to gray roofing gravel. (a) Independent component image. (b) Histogram of independent component values. This histogram conforms well to the observed abundance model given by (5.3), showing an obvious scaling error and blurring of the sharp peaks of the ideal abundance.

5.3 Rescaling Algorithm

An algorithm has been developed that leverages the independent component model (5.10) to perform the rescaling necessary to make an independent component useful as an abundance map. The algorithm consists of two fundamental steps: 1) parameter estimation; and 2) nonlinear mapping.

5.3.1 Parameter Estimation

The goal of the parameter estimation step is to find values for the parameters in (5.10) to best fit the observed data. These parameters consist of the class probabilities: P_e, P_f, P_m, the scale ambiguity, c_1, the offset, c_0, and the noise variance, σ^2. Note that because the source of the noise is identical for all three classes there is only one noise parameter to estimate. Also, rather than estimating c_1 and c_0 explicitly, we instead estimate the means of the Gaussian classes. We denote these parameters as μ_0 and μ_1. This results in a sign ambiguity—μ_0 could correspond to the empty or the filled class—which is resolved in the nonlinear mapping step. Additionally, there are two parameters, α and β, associated with the beta distribution that must be estimated if that distribution is used to describe the mixed pixel class. Collecting everything, we have the following parameters: $\boldsymbol{\theta}_u = \{P_\mathrm{e}, P_\mathrm{f}, P_\mathrm{m}, \mu_0, \mu_1, \sigma^2\}$ (uniform distribution) or $\boldsymbol{\theta}_\beta = \{P_\mathrm{e}, P_\mathrm{f}, P_\mathrm{m}, \mu_0, \mu_1, \sigma^2, \alpha, \beta\}$ (beta distribution).

The approach for estimating the parameters described above is based on the expectation-maximization (EM) algorithm [28]. The EM algorithm and a slight variation of it, stochastic-expectation-maximization (SEM) [29], have been applied directly to the hyperspectral unmixing problem [30, 31], but not to the problem of rescaling abundances. The EM algorithm is especially useful for maximum-likelihood (ML) estimation problems where there is a many-to-one mapping from an underlying distribution to the distribution governing the observation [32]. In this case, the underlying distribution is given by the class distributions in (5.4), (5.5), and (5.7), and the observation is given by (5.10).

The expectation step of the algorithm consists of computing the expected value of class membership given the observed data and current parameter estimates. In similar problems these values have been referred to as responsibilities [33] or posterior class probabilities [31]. Let $\mathbf{a} \equiv \begin{bmatrix} a(1) & a(2) & \dots & a(T) \end{bmatrix}$ be an observed independent component with T pixels that has been row-scanned to form a vector. The posterior class probabilities can be computed for a given observation, $a(t)$, as

$$\widetilde{\Pi}_0(t) = \hat{P}_0 \cdot \phi(a(t), \hat{\mu}_0, \hat{\sigma}^2), \tag{5.11}$$

$$\widetilde{\Pi}_1(t) = \hat{P}_1 \cdot \phi(a(t), \hat{\mu}_1, \hat{\sigma}^2), \tag{5.12}$$

$$\widetilde{\Pi}_{\mathrm{m}}(t) = \hat{P}_{\mathrm{m}} \cdot p_{\mathrm{m}}(a(t)). \tag{5.13}$$

These responsibilities must be normalized according to

$$\Pi_k(t) = \frac{\widetilde{\Pi}_k}{\widetilde{\Pi}_0(t) + \widetilde{\Pi}_1(t) + \widetilde{\Pi}_{\mathrm{m}}(t)}, \ k \in \{0, 1, \mathrm{m}\}. \tag{5.14}$$

The maximization step of the algorithm computes the ML estimates of the parameters, given the results from the expectation step. The class probabilities are computed as

$$\hat{P}_k = \frac{1}{T} \sum_{t=1}^{T} \Pi_k(t), \ k \in \{0, 1, \mathrm{m}\}, \tag{5.15}$$

where T is the number of pixels in the image. The estimates of the mean are computed according to

$$\hat{\mu}_k = \frac{\sum_{t=1}^{T} \Pi_k(t) \cdot a(t)}{\sum_{t=1}^{T} \Pi_k(t)}, \ k \in \{0, 1\}. \tag{5.16}$$

The estimate of σ^2 can be obtained by estimating σ^2 for each class and combining those estimates, weighted by the estimated class probabilities, i.e.,

$$\hat{\sigma}^2 = \hat{P}_0 \hat{\sigma}_0^2 + \hat{P}_1 \hat{\sigma}_1^2 + \hat{P}_{\mathrm{m}} \hat{\sigma}_{\mathrm{m}}^2. \tag{5.17}$$

For the pure-pixel classes, the estimates of σ^2 are given by

$$\hat{\sigma}_k^2 = \frac{\sum_{t=1}^{T} \Pi_k(t) \cdot (a(t) - \hat{\mu}_k)^2}{\sum_{t=1}^{T} \Pi_k(t)}, \; k \in \{0, 1\}. \tag{5.18}$$

For the mixed-pixel class, estimating σ^2 is somewhat more complicated. In the case of the uniform distribution, it requires equating the derivative of the log-likelihood to zero and solving for σ, i.e.,

$$\left(\frac{2}{\sqrt{\pi}}\right) \sum_{t=1}^{T} \frac{\exp\left[-z_1(t)^2\right] \frac{z_1(t)}{\sigma} - \exp\left[-z_0(t)^2\right] \frac{z_0(t)}{\sigma}}{\operatorname{erf}\left[z_0(t)\right] - \operatorname{erf}\left[z_1(t)\right]} = 0, \tag{5.19}$$

where

$$z_k(t) \equiv \frac{a(t) - \hat{\mu}_k}{\sqrt{2}\sigma}, k = 0, 1. \tag{5.20}$$

Because there is no closed-form solution to (5.19), it must be solved numerically using an optimization approach such as Newton's method.

When the beta distribution is used to describe the mixed-pixel class the parameters α and β must be estimated along with or prior to estimating σ^2. It turns out that obtaining ML estimates for the parameters of a beta distribution requires solving a system of equations with no closed-form solution, typically employing a numerical approach such as Newton's method. An alternative approach that is more straightforward is the method of moments. Not only is method of moments estimation simpler, it is also more accurate than ML estimation for small sample sizes [34]. Because the problem here is even more complicated due to the addition of Gaussian noise, a method of moments approach was implemented to estimate α, β, and σ^2. Because of their complexity, the estimation formulas are shown in Appendix B. Initial testing has shown that the formulas work well when α and β take sufficiently different values and the number of samples is large. However, when the number of samples is small or $\alpha \approx \beta$, the estimation results are questionable.

Because of the complexity of the estimation process and potential unreliability of the estimation results, the decision was made to estimate σ^2 without considering the contribution of the mixed-pixel class. In this case the estimate is obtained from

$$\hat{\sigma}^2 = \frac{\hat{P}_0 \hat{\sigma}_0^2 + \hat{P}_1 \hat{\sigma}_1^2}{\hat{P}_0 + \hat{P}_1}. \tag{5.21}$$

This modification yields an estimate that is not necessarily the ML estimate, which could affect the convergence of the EM algorithm. In practice, however, no negative effects have been observed. This is most likely because only a small number of pixels make up the mixed-pixel class and a reliable estimate of σ^2 can be obtained from (5.21). Although the small number of pixels in the mixed-pixel class enables this simplification, it also precludes application of the SEM algorithm to this problem.

Under normal circumstances, the EM algorithm is guaranteed to converge to a local solution, although not necessarily to the global optimum. As such, it is dependent on initial conditions. Testing of the algorithm has shown that the algorithm is not particularly sensitive to the initial values of μ_0, μ_1, or σ^2, but is a bit more sensitive to the class probabilities, P_0, P_1, and P_{m}. Based on observations from testing of the algorithm, a standard initialization scheme has been developed. In this scheme μ_0 is set to the minimum observed value, μ_1 to the maximum, and σ^2 to the sample variance of the observed data. The class probabilities are initialized to $P_0 = 0.98$, $P_1 = 0.01$, and $P_{\mathrm{m}} = 0.01$, assuming an extremely rare material. Obviously, when the material is rare in the scene this is a good assumption, but when the material is common in the scene the algorithm is still able to converge to an acceptable solution. The converse of this is not true.

One other observation that has been made regarding the class probabilities is that in certain situations when the observed data contain outliers on both sides of the large Gaussian mode corresponding to the empty-pixel class, the algorithm may converge to a solution with an overly large proportion of mixed-pixels. This behavior is restricted in the algorithm by checking the estimated value of P_{m} at each iteration. If the value is greater than 0.5 it is reset to 0.01 and the other class probabilities are normalized accordingly.

5.3.2 Nonlinear Mapping

Once the distribution parameters have been determined, all that remains is to map the observed data back to the ideal. The first step is to determine to which class each independent component value should belong. This is done by calculating the class responsibilities, $\widetilde{\Pi}_k$, defined in (5.11), (5.12), and (5.13). Each independent component value is assigned to the class for which $\widetilde{\Pi}_k$ is largest.

The next step is to determine which pure-pixel class each of the two Gaussian classes should be associated with. This is done in either of two ways. First, since c_0 is expected to be small, the Gaussian class with estimated mean closest to zero can be assigned to the empty class. Second, the estimated class probabilities, $\hat{P}_0$ and $\hat{P}_1$, can be inspected and the empty-pixel class associated with the larger of these two values. This assignment could be wrong if there is a material in the scene which completely fills more than half of the pixels in the image. Although it depends on the type of scene, this condition does not occur frequently and can only happen at most once per image. Also, when a material fills a large area of the image it is fairly common for ICA to split that material into multiple independent components, each of which would fill less than half of the image [see Chapter 4].

The final step is to reassign independent component values to conform to the ideal model. Pixels belonging to the empty-pixel class are set to zero, and pixels assigned to the filled-pixel class are set to one. Finally pixels belonging to the mixed class are mapped linearly to the range $[0,1]$, i.e.,

$$\hat{a}(t) = \frac{a(t) - a_{\min}}{a_{\max} - a_{\min}}, \tag{5.22}$$

where $a_{\max}$ and $a_{\min}$ are the maximum and minimum values, respectively, of the pixels assigned to the mixed pixel class.

5.4 Experimental Results

A variety of experiments have been performed to assess the behavior and performance of the proposed algorithm. Each of these experiments used synthetic hyperspectral data generated using DIRSIG. Synthetic data are especially useful for quantitative assessments of new algorithms because they provide accurate ground truth. For the experiments described below a hyperspectral radiance cube was generated using an AVIRIS-like spectral response with 224 bands. The spatial dimensions of the cube are 128 pixels by 128 pixels. A broadband image of the cube is shown in Fig. 5.3.

From this one synthetic cube a number of test cubes were produced by adding various realizations of Gaussian noise. Each of these cubes was processed using the FastICA algorithm [35], varying the processing options to produce a variety of independent components. These independent components were then rescaled to produce estimated abundances.

In each experiment the estimated abundances are compared to the true abundances associated with the DIRSIG-generated cube. Two metrics are used to quantify this comparison. The first is mean-square error, defined as

$$MSE(\mathbf{x},\mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T}\left(x(t)-y(t)\right)^2, \tag{5.23}$$

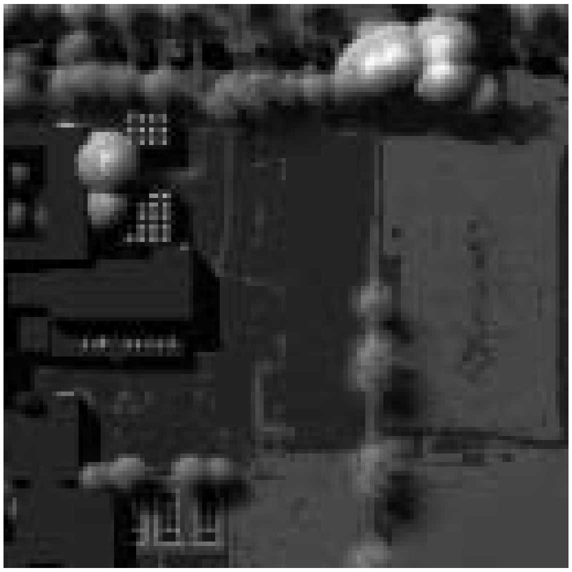


Fig. 5.3: Broadband image of synthetic data used in algorithm assessment. The scene contains large buildings to the left, a parking lot in the center, grassy areas on the right, and trees intermingled with residential roofs at the top.

where $\mathbf{x}$ and $\mathbf{y}$ are T-element vectors. The other proposed metric is the correlation coefficient, given by

$$r(\mathbf{x},\mathbf{y}) = \frac{\sum_{t=1}^{T}\left(x(t)-\overline{x}\right)\left(y(t)-\overline{y}\right)}{\sqrt{\sum_{t=1}^{T}\left(x(t)-\overline{x}\right)^2}\sqrt{\sum_{t=1}^{T}\left(y(t)-\overline{y}\right)^2}}. \tag{5.24}$$

The correlation coefficient is useful in assessing ICA performance since it is invariant to scale, and thus unaffected by the scale ambiguity inherent in ICA outputs. Mean-square error is intuitive and widely used, but is not useful in examining (non-rescaled) ICA outputs. It can also be somewhat misleading when comparing sparse abundance maps.

The proposed method, referred to hereafter as class-based abundance rescaling (CBAR) is compared to two other rescaling approaches. The first is an intuitive linear mapping of the independent component to the range $[0,1]$. This approach is referred to as linear abundance rescaling (LAR). The formula for this mapping is given by (5.22) with $a_{\max}$ and $a_{\min}$ corresponding to the maximum and minimum values, respectively of the independent component. A final rescaling approach is the abundance quantification algorithm (AQA) [22] given by

$$\hat{a}(t) = \frac{|a(t)| - \min_t |a(t)|}{\max_t |a(t)| - \min_t |a(t)|}. \tag{5.25}$$

This approach is the LAR approach applied the absolute value of the independent component.

The remainder of this section demonstrates the performance of the CBAR algorithm. In the next two subsections, CBAR behavior is demonstrated by showing rescaling results for two specific independent components. One of these components is is fairly dense, while the other is relatively sparse. In the final subsection the average performance of the algorithm across a variety of independent components is shown. In all of these scenarios CBAR performance is compared to the LAR and AQA approaches.

5.4.1 Dense Material Example

The independent component shown in Fig. 5.2 and corresponding to the roofing material shown in Fig. 5.1 is an example of an independent component that conforms well to the model in (5.10). This independent component was rescaled using the CBAR algorithm. A histogram of the independent component with the estimated pdf is shown in Fig. 5.4. The figure shows good agreement between the data and the model. The true class probabilities can be obtained from the truth data by determining the number of empty, filled, or mixed pixels in the truth map and dividing by the number of pixels in the image. Comparing these results to the values obtained by the parameter estimation step of the CBAR algorithm we have: $P_0 = 0.8552$, $\hat{P}_0 = 0.8580$, $P_1 = 0.1126$, $\hat{P}_1 = 0.0942$, and $P_{\mathrm{m}} = 0.0322$, $\hat{P}_{\mathrm{m}} = 0.0478$. The estimation error for all of these values is less than 0.02. The estimated mean values are $\hat{\mu}_0 = -3.0359$ and $\hat{\mu}_1 = 0.1066$ and the estimated variance is $\hat{\sigma}^2 = 0.0638$. The threshold values obtained by the nonlinear mapping step of the CBAR algorithm are $t_0 = -2.4947$ and $t_1 = -0.6504$. Because this component is inverted, any value less than t_0 is mapped to the filled-pixel class and any value greater than t_1 is mapped to the empty-pixel class. The image resulting from the rescaling is shown in Fig. 5.5 along with the truth map, the original independent component, and the AQA-rescaled image. Also shown are the histograms associated with each image. The histograms and figures show the denoising effect that the CBAR algorithm has, compared to other approaches. The mean-square error and correlation coefficient associated with each of these are shown in Table 5.1. These results show that the CBAR approach has the lowest MSE and largest correlation coefficient of any of the approaches. Note that the negative correlation coefficients for some of the estimates indicates that the estimate is inverted.

Table 5.1: Mean-square error and correlation coefficient for gray roofing gravel abundance estimates.

	MSE	r
Independent component	1.8652	-0.9097
LAR abundance	0.4961	-0.9097
AQA abundance	0.0244	0.9090
CBAR abundance	0.0168	0.9222

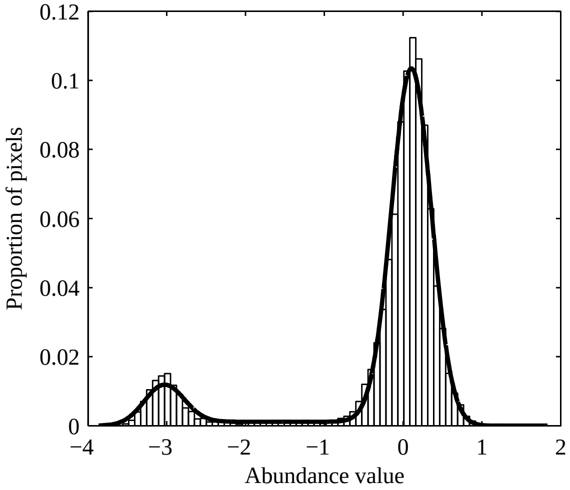
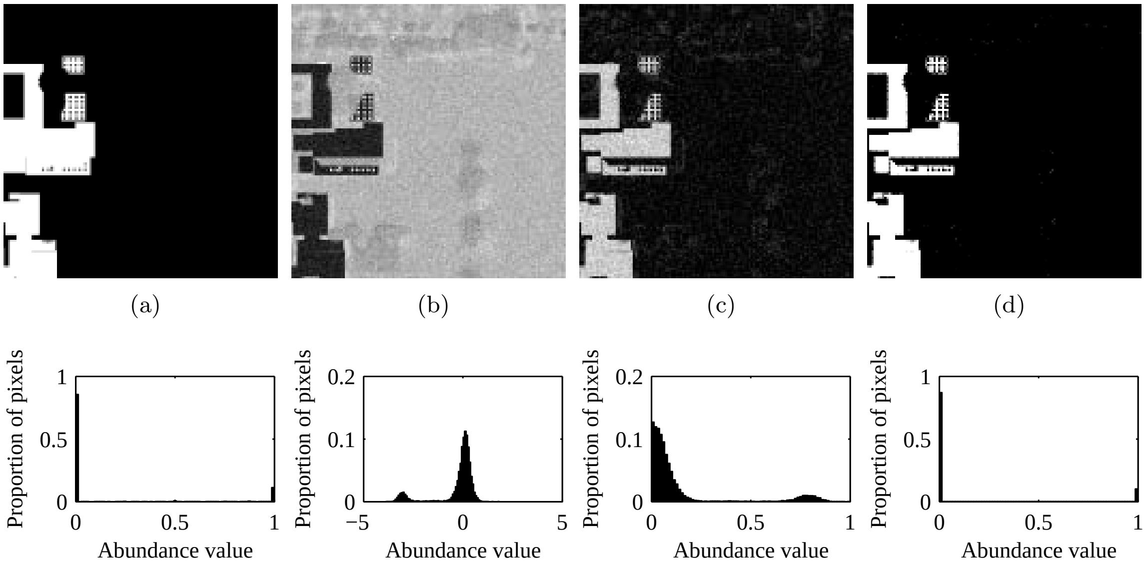


Fig. 5.4: A histogram of the independent component associated with the gray roofing gravel material along with the estimated pdf, generated according to the independent component model in (5.10).



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 5.5: A comparison of abundance maps (first row) and histograms (second row) associated with the gray roofing gravel material. (a) and (e) Abundance truth map. (b) and (f) Independent component. (c) and (g) Estimated abundance obtained using the AQA algorithm. (d) and (h) Estimated abundance obtained using the CBAR algorithm.

5.4.2 Sparse Material Example

In the previous example the material of interest was present in nearly 15% of the image pixels, making it relatively dense. The CBAR algorithm also works to rescale materials that are much more sparse. In this example a material is considered which is present in less than 0.3% of the image pixels. Figure 5.6 shows the truth map for a brown brick siding material, along with associated independent component and rescaling results. The histograms are also shown, although it is difficult to discern the Gaussian mode corresponding to the filled-pixel class. Because only one Gaussian mode is visible, the histogram and estimated pdf are not shown. Comparing the estimated class probabilities to the true values gives $P_0 = 0.9973$, $\hat{P}_0 = 0.9962$, $P_1 = 7.9346 \times 10^{-4}$, $\hat{P}_1 = 1.3739 \times 10^{-4}$, and $P_\mathrm{m} = 0.0019$, $\hat{P}_\mathrm{m} = 0.0037$. The errors in these estimates are all less than 0.002. The other estimated values are $\hat{\mu}_0 = -15.8143$, $\hat{\mu}_1 = 6.4748$, and $\hat{\sigma^2} = 0.5030$. The class thresholds are located at $t_0 = -15.8142$ and $t_1 = -3.6141$.

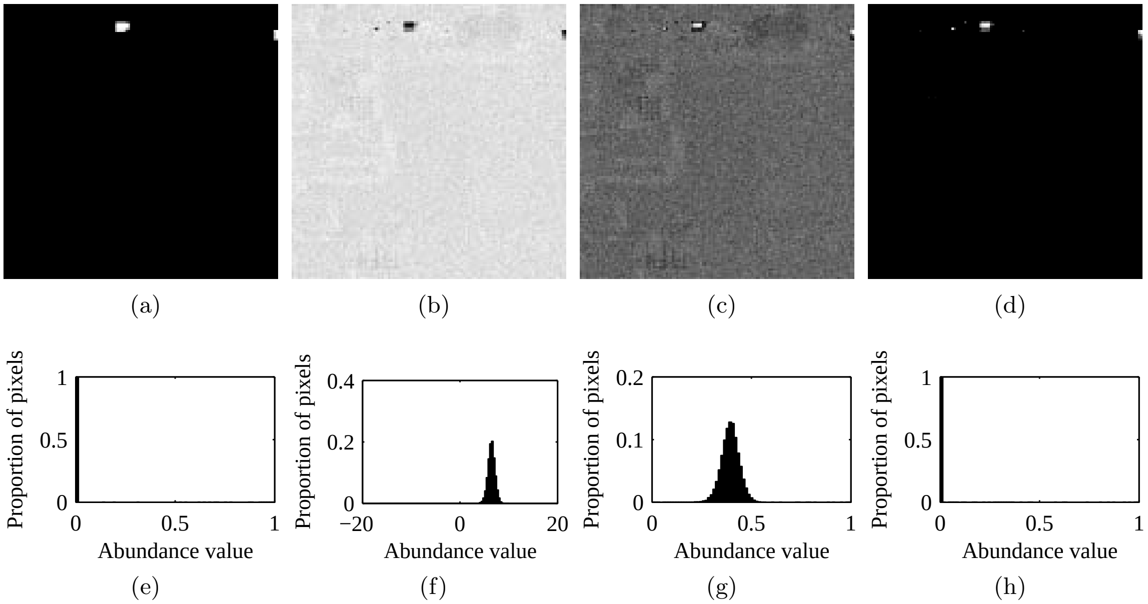
Based on the way FastICA is supposed to behave, the value of c_0 in (5.3) is expected to be close to zero. This independent component violates that assumption. In this case, the mean corresponding to the empty-pixel class is at 6.4748. The CBAR and LAR algorithms are tolerant of this violation, however the AQA approach is not. Because the values corresponding to empty pixels are not close to zero, the AQA approach actually reduces the correlation coefficient. These results are shown in Fig. 5.6 and Table 5.2.

5.4.3 Average Performance

In order to quantify the performance of the CBAR algorithm more generally, a large

Table 5.2: Mean-square error and correlation coefficient for brown brick siding abundance estimates.

	MSE	r
Independent component	42.4931	-0.6231
LAR abundance	0.7509	-0.6231
AQA abundance	0.1575	0.0384
CBAR abundance	5.2022×10^{-4}	0.8492



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 5.6: A comparison of abundance maps (first row) and histograms (second row) associated with the brown brick siding material. (a) and (e) Abundance truth map. (b) and (f) Independent component. (c) and (g) Estimated abundance obtained using the AQA algorithm. (d) and (h) Estimated abundance obtained using the CBAR algorithm.

number of independent components were generated by running FastICA with a variety of parameters on the test cube shown in Fig. 5.3 with a number of different noise levels and realizations. Each independent component was then processed by the LAR, AQA, and CBAR rescaling approaches. When any result, including the independent component, had a correlation magnitude greater than 0.6, it was retained, along with the corresponding results from the other processing approaches. This step allows for erroneous independent components to be excluded from the characterization. After this exclusion step, 904 independent components were retained. The average mean-square error and correlation coefficient magnitude for each processing approach was then calculated. The averages were computed separately for sparse materials and dense materials. Dense materials are loosely defined as those that appear in at least hundreds of image pixels (greater than 1.5%) and appear as a "pure" material in at least a handful of pixels. Sparse materials constitute the remaining materials.

The average results are shown in Table 5.3. Notice that for these computations the magnitude of the correlation coefficient was considered, ignoring sign ambiguity. All three rescaling approaches were able to reduce the MSE, with AQA performing better than LAR, and CBAR better than the both of them. For sparsse materials the improvement over AQA was an order of magnitude. Also, the CBAR approach was the only approach that was able to increase the correlation coefficient. The improvement was moderate for dense materials, but significant for sparse materials.

5.5 Conclusion

The problem of hyperspectral unmixing has been approached in a variety of ways, utilizing a vast assortment of algorithms. One approach has been to use independent component analysis to perform the unmixing. Although promising, using ICA is hindered because the output abundance estimates exhibit a scale ambiguity. In practice, a constant offset is also often present. Additionally, independent components also tend to be noisy. The better the scale ambiguity, offset, and noise can be eliminated, the more useful the independent components will be as abundance maps.

A statistical model for abundances has been developed which assumes three classes of abundance pixels: empty, filled, and mixed. This model can be extended to produced a corresponding statistical model for independent component values. The independent component model takes into account the effects of scale ambiguity, offset, and noise. Using the relationship between the ideal model and the independent component model, an algorithm has been developed which estimates model parameters to fit an independent component

Table 5.3: Average mean-square error and correlation coefficient magnitude for rescaling approaches.

	Sparse Materials		Dense Materials		All Materials	
	MSE	$\lvert r\rvert$	MSE	$\lvert r\rvert$	MSE	$\lvert r\rvert$
Independent component	16.2966	0.5060	3.9703	0.7503	10.4062	0.6228
LAR abundance	0.3809	0.5060	0.3426	0.7503	0.3615	0.6228
AQA abundance	0.0246	0.4672	0.0314	0.7096	0.0280	0.5884
CBAR abundance	0.0021	0.8198	0.0242	0.7700	0.0127	0.7960

to the independent component model. These estimated parameters can then be used to map independent component values to ideal values, effectively reducing the impact of scale ambiguity, offset, and noise.

The proposed rescaling algorithm was compared to two other approaches, using synthetically generated data to provide a quantitative assessment. The proposed algorithm was the only approach that was able to both reduce the mean-square error and increase the correlation correlation coefficient of the estimates. For sparse materials, the increase was significant.

References

[1] M. S. Kim, A. Lefcourt, K. Chao, Y. Chen, I. Kim, and D. Chan, "Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part I. Application of visible and near-infrared reflectance imaging," *Transactions of the American Society of Agricultrual and Biological Engineering*, vol. 45, no. 6, pp. 2027–2037, 2002.

[2] M. Moran, Y. Inoue, and E. Barnes, "Opportunities and limitations for image-based remote sensing in precision crop management," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 319–346, 1997.

[3] F. Kruse, J. Boardman, and J. Huntington, "Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1388–1400, June 2003.

[4] T. Schmid, M. Koch, J. Gumuzzio, and P. M. Mather, "A spectral library for a semi-arid wetland and its application to studies of wetland degradation using hyperspectral and multispectral data," *International Journal of Remote Sensing*, vol. 25, pp. 2485–2496, July 2004.

[5] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.

[6] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, pp. 23–26, 1995.

[7] R. Neville, K. Staenz, T. Szeredi, J. Lefebvre, and P. Hauff, "Automatic endmember extraction from hyperspectral data for mineral exploration," *Proceedings of 21st Canadian Symposium on Remote Sensing*, pp. 891–897, 1999.

[8] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sept. 2002.

[9] M. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

[10] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE*, vol. 3753, pp. 266–275, 1999.

[11] 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, Mar. 1999.

[12] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[13] D. R. Fuhrmann, "A simplex shrink-wrap algorithm," *Proceedings of SPIE*, vol. 3718, pp. 501–511, 1999.

[14] J. Nascimento and J. Bioucas-Dias, "Hyperspectral unmixing algorithm via dependent component analysis," *IEEE Geoscience and Remote Sensing Symposium*, pp. 4033–4036, July 2007.

[15] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomputing*, vol. 71, no. 10-12, pp. 2194–2208, June 2008.

[16] A. Cichocki and S. ichi Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. West Sussex, UK: John Wiley & Sons, 2002.

[17] J. Bayliss, J. A. Gualtieri, and R. F. Cromp, "Analyzing hyperspectral data with independent component analysis," *Proceedings of SPIE*, vol. 3240, pp. 133–143, 1997.

[18] X. Zhang and C. H. Chen, "New independent component analysis method using higher order statistics with application to remote sensing images," *Optical Engineering*, vol. 41, pp. 1717–1728, 2002.

[19] J. Nascimento and J. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[20] T. Tu, "Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis approach," *Optical Engineering*, vol. 39, no. 4, pp. 897–906, Apr. 2000.

[21] B. R. Foy and J. Theiler, "Scene analysis and detection in thermal infrared remote sensing using independent component analysis," *Proceedings of SPIE*, vol. 5439, pp. 131–139, 2004.

[22] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 260–2616, Sept. 2006.

[23] C. Shah, M. Arora, S. Robila, and P. Varshney, "ICA mixture model based unsupervised classification of hyperspectral imagery," *Proceedings of Applied Imagery Pattern Recognition Workshop*, pp. 29–35, Oct. 2002.

[24] S. Chiang, C.-I. Chang, and I. Ginsberg, "Unsupervised hyperspectral image analysis using independent component analysis," *IEEE Geoscience and Remote Sensing Symposium*, vol. 7, pp. 3136–3138, 2000.

[25] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[26] *The DIRSIG User's Manual*, Rochester Institute of Technology, Oct. 2006. [www.dirsig.org/documentation]

[27] E. J. Ientilucci and S. D. Brown, "Advances in wide-area hyperspectral image simulation," *Proceedings of SPIE*, pp. 110–121, 2003.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[29] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 3, pp. 618–633, May 1993.

[30] M. T. Eismann and R. C. Hardie, "Stochastic spectral unmixing with enhanced endmember class separation," *Applied Optics*, vol. 43, no. 36, pp. 6596–6608, Dec. 2004.

[31] A. D. Stocker and A. P. Schaum, "Application of stochastic mixing models to hyperspectral detection problems," *Proceedings of SPIE*, vol. 3071, pp. 47–60, 1997.

[32] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[33] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, ser. Springer series in statistics. New York: Springer, 2009.

[34] C. B. Owen, "Parameter estimation for the beta distribution," Master's thesis, Brigham Young University, Provo, UT, 2008.

[35] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, May 1999.

Chapter 6

An Extended Algorithm to Rescale Independent Components for Abundance Quantification[1]

6.1 Introduction

Many approaches have been applied to unmixing hyperspectral data including algorithms that exploit the geometry of the linear mixing model [1–7], extract "extreme" spectra [8–11], minimize signal complexity [12], estimate statistical model parameters [13–17], and many others. The goal of the unmixing process is to separate the individual spectra of materials in a pixel—allowing for material identification based on spectral matching or pattern analysis, and to estimate the amount of the material present in the pixel—material (abundance) quantification.

Independent component analysis (ICA) has been explored as a spectral unmixing method. ICA encompasses a variety of algorithms that unmix data by finding results—typically abundances—that are statistically independent. ICA algorithms used to unmix hyperspectral data include joint cumulant-based ICA [18], joint approximate diagonalization of eigen-matrices (JADE) [19], FastICA [19–22], and a FastICA modification called linear spectral random mixture analysis (LSRMA) [23].

ICA outputs, called independent components (ICs), have a scale ambiguity that arises because scalar multiplication does not affect the independence of two signals. If the ambiguity is not resolved, the practical application of ICA for abundance estimation is difficult, since abundance estimates should range from zero to one [24]. Various approaches have been proposed to rescale independent components for abundance quantification. These include the abundance quantification algorithm (AQA) [22], which finds the minimum and

---

[1] M.R. Stites, J.H. Gunther, T.K. Moon, and G.P. Williams (to be submitted for publication)

maximum values of the absolute value of an IC and performs a linear mapping on the range $[0, 1]$ and the class-based abundance rescaling (CBAR) [see Chapter 5] algorithm, which uses statistical models to describe both ideal abundances and ICs and uses the theoretical relationship between the two models to perform a nonlinear mapping on the range $[0, 1]$. Experiments comparing rescaled abundances to truth data showed that AQA typically produces significant reductions in mean-square error (MSE) compared to the unscaled data, but at the cost of decreasing the correlation coefficient, while CBAR consistently lowers MSE and simultaneously increases the correlation coefficient [see Chapter 5].

In this chapter, an extended CBAR algorithm is proposed called CBAR-X. Generally, ICs obtained using ICA represent single materials. However, when material spectra are negatively correlated, a single IC can represent two distinct materials—a condition that occurred slightly over 25% of the time in experiments. The CBAR-X algorithm extends the statistical models used to describe abundance values to address two-material cases where the original models do not describe observed data. This extension enables accurate extraction of two materials from a single IC when that component represents two negatively-correlated materials. Additionally, when the data conform to the original CBAR model, (i.e., when the IC represents a single material), there is no reduction in performance or penalty incurred by using CBAR-X.

6.2 Abundance Model and Rescaling Algorithm

The CBAR algorithm assumes that for a given material, an abundance pixel belongs to one of three classes: empty, filled, or mixed. Empty pixels have a value of zero, filled pixels a value of one, and mixed pixels have a value between zero and one. This leads to a statistical model for ideal abundances given by

$$p_{\mathrm{ideal}}(a) = P_{\mathrm{e}}\delta(a) + P_{\mathrm{f}}\delta(1-a) + P_{\mathrm{m}}f_{\mathrm{m}}(a), \tag{6.1}$$

where P_{e}, P_{f}, and P_{m} are the proportions of empty, filled, and mixed pixels, respectively, and $f_{\mathrm{m}}(a)$ is a probability density function (pdf) that describes the distribution of the mixed-pixel abundance values. Both the uniform and beta distributions were considered to describe the mixed-pixel abundance distribution.

An abundance map for gray roofing gravel generated by DIRSIG [25] is shown in Fig. 6.1(a) with Fig. 6.1(b) showing the simplicity of the three-class model for ideal abundances. Although synthetically-generated, DIRSIG data are generated using physical modeling of real-world scenes and spectral measurements, so the ground truth images provide an accurate representation of abundance data in real scenes.

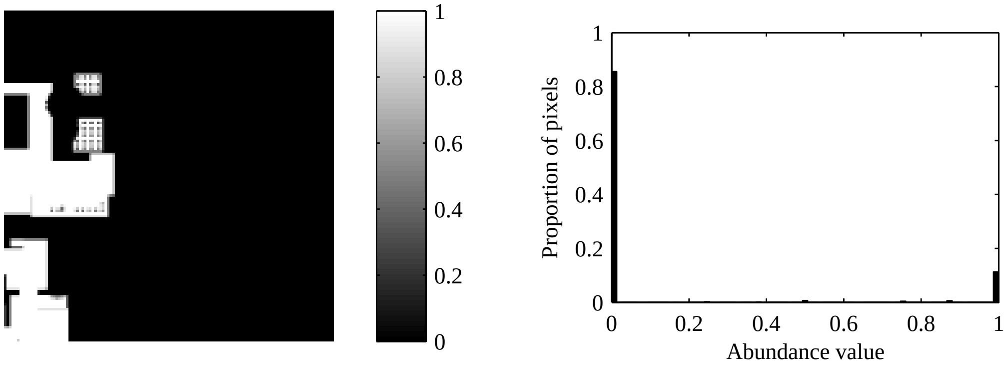
A statistical model for observed independent components is produced by scaling (by c_1), shifting (by c_0), and adding zero-mean Gaussian noise to the ideal data. This leads to a pdf given by

$$p_{\mathrm{ic}}(a) = P_{\mathrm{e}} p_{\mathrm{e}}(a) + P_{\mathrm{f}} p_{\mathrm{f}}(a) + P_{\mathrm{m}} p_{\mathrm{m}}(a), \tag{6.2}$$

where $p_{\mathrm{e}}(a) = \phi(a|c_0, \sigma^2)$, $p_{\mathrm{f}}(a) = \phi(a|c_1 + c_0, \sigma^2)$, and

$$\phi(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \tag{6.3}$$

The mixed-pixel class distribution of the IC, $p_{\mathrm{m}}(a)$, is obtained by convolving $f_{\mathrm{m}}(a)$ with (6.3). An IC obtained using FastICA that is strongly correlated to the abundance map in Fig. 6.1(a) is shown in Fig. 6.2(a) with corresponding histogram in Fig. 6.2(b). The



(a) (b)

Fig. 6.1: DIRSIG-generated abundance truth data for gray roofing gravel. (a) Abundance map image. (b) Histogram of abundance values. This histogram conforms well to the ideal abundance model given by (6.1).

curve overlaying the histogram is generated by (6.2), and shows good agreement between the model and data.

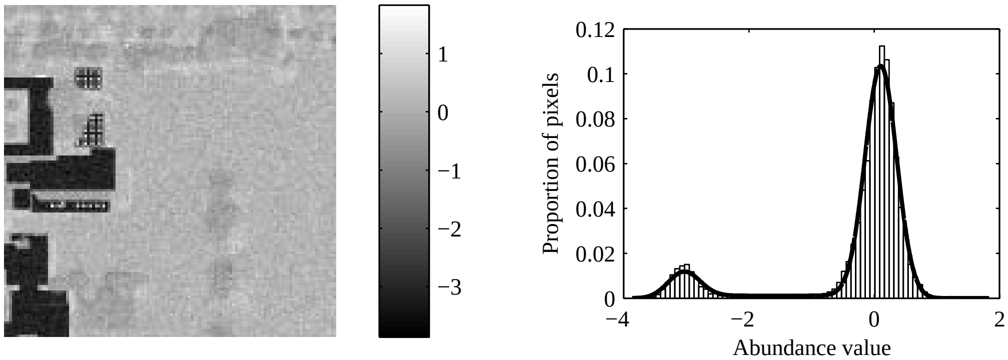
The CBAR rescaling algorithm uses expectation-maximization (EM) [26] to estimate model parameters in (6.2) to best fit the IC values. For each IC value, the algorithm calculates the probability of belonging to each class and the most probable class for each pixel. Empty and filled pixels are mapped to zero and one, respectively. Mixed-class pixels are mapped linearly between zero and one. Figure 6.3 shows an abundance map generated using CBAR to rescale the IC in Fig. 6.2(a). The rescaling reduced the MSE from 1.8652 to 0.0168 and increased the correlation coefficient magnitude from 0.9097 to 0.9222.

The CBAR algorithm performs well for data that fit the three-class model where each IC represents one material. However, there can be ICs that do not conform to this model. Figure 6.4 shows an example. The histogram shows that this component has a dominant mode near zero corresponding to empty pixels, but also shows, unexpectedly, that there are a significant number of pixels to the left and the right of this mode. Figure 6.4 shows the values to the right of zero as bright pixels and values to the left as dark pixels in the image. Here the bright pixels correspond to white sheet metal and the dark pixels correspond to brown siding. A number of similar examples were observed, indicating that two-sided independent components are not rare. The two-sided ICs occur when two materials are strongly negatively correlated which makes sense given ICA is a linear unmixing approach. Figure 6.5 presents the reflectance spectra for the white sheet metal and brown siding materials and shows their strong negative correlation. These two material spectra have a correlation coefficient of -0.7914.

To address two-sided ICs that represent abundances for two distinct materials, the CBAR model was extended to contain five classes:

$$p_{\mathrm{ic}}(a) = P_{\mathrm{e}} p_{\mathrm{e}}(a) + P_{\mathrm{f}^+} p_{\mathrm{f}^+}(a) + P_{\mathrm{m}^+} p_{\mathrm{m}^+}(a) + P_{\mathrm{f}^-} p_{\mathrm{f}^-}(a) + P_{\mathrm{m}^-} p_{\mathrm{m}^-}(a). \tag{6.4}$$

CBAR-X retains an empty class with mixed and filled pixels on either side. The rescaling algorithm is the same two steps, parameter estimation followed by nonlinear mapping.



(a) (b)

Fig. 6.2: Independent component obtained using FastICA corresponding to gray roofing gravel. (a) Independent component image. (b) Histogram of independent component values. This histogram conforms well to the observed abundance model given by (6.2), showing an obvious scaling error and blurring of the sharp peaks of the ideal abundance.

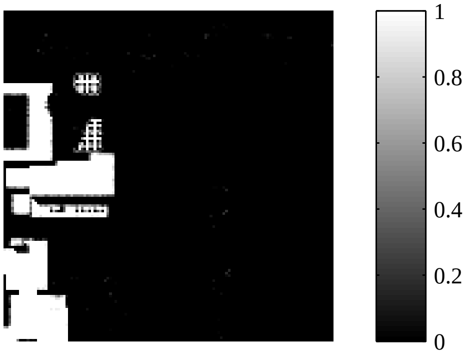
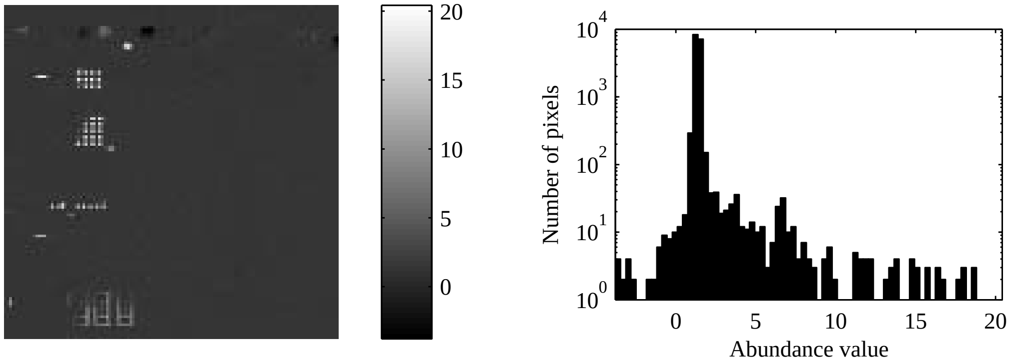


Fig. 6.3: Abundance map produced by using CBAR to rescale the independent component corresponding to the gray roofing gravel material.



(a) (b)

Fig. 6.4: Independent component corresponding to both white sheet metal and brown siding. (a) Independent component image. (b) Histogram of abundance values showing a non-negligible number of pixels on both sides of the empty pixels. To more clearly illustrate the two-sided nature of the data, this histogram shows—on a logarithmic scale—the number of pixels, rather than the proportion of pixels.

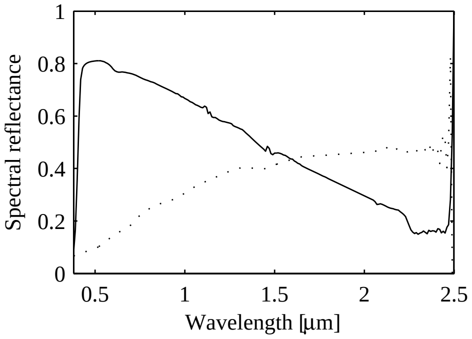


Fig. 6.5: Reflectance spectra for white sheet metal (solid) and brown siding (dotted). The correlation coefficient between these two spectra is -0.7914.

The parameter estimation remains the same as in the original CBAR algorithm, but with additional parameters. The nonlinear mapping for the five-class model is similar to the three-class model, but is be performed twice, once for each material. This is illustrated in Fig. 6.6, which shows a distribution generated according to (6.4) and the threshold values between classes. For the negative (left) material, the values greater than b are assigned to the empty class, and values less than a are assigned to the filled class. For the positive (right) material, values less than c are assigned to the empty class and values greater than d are assigned to the filled class.

6.3 Experimental Results

An experiment was performed to compare the performance of CBAR-X (five-class) with CBAR (three-class). DIRISG was used to generate radiance data cubes with an AVIRIS-like spectral response [27]. Different Gaussian noise realizations were added to produce 106 test cubes. Each test cube was processed using FastICA, producing a large number of ICs. Each IC was rescaled with CBAR and CBAR-X which resulted in four abundance maps for each IC: the IC itself, the CBAR rescaled abundance, and two CBAR-X rescaled abundances. For each of these four results, the mean-square error and correlation coefficient was computed for every abundance truth map. If the correlation coefficient magnitude for any of these results was greater than 0.6, all four results were retained. This was done to exclude ICs that do not correspond to actual materials. Ultimately 923 results were retained. The

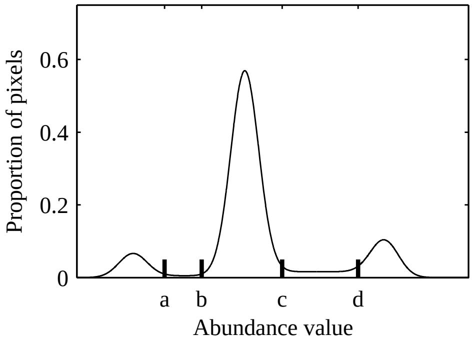


Fig. 6.6: An example of a distribution following the five-class model given in (6.4) and the threshold between the five classes.

second CBAR-X result was retained only if there was a material for which the correlation coefficient magnitude was greater than 0.6 to ignore extraneous second material results for ICs with only one material. There were 236 results (25.57%) where two distinct materials were extracted from a single IC.

To demonstrate the ability of CBAR-X to extract two materials from a single IC, the IC shown in Fig. 6.4(a) was used. Figure 6.7 shows the results and corresponding truth maps. Table 6.1 shows the MSE and correlation coefficient for this IC, processed by CBAR and CBAR-X. For the primary white sheet metal material the improvements from CBAR and CBAR-X over the unprocessed IC are significant and similar. For the secondary brown siding material, the abundance from CBAR-X is significantly improved over the original IC and is as good as the result obtained for the primary material.

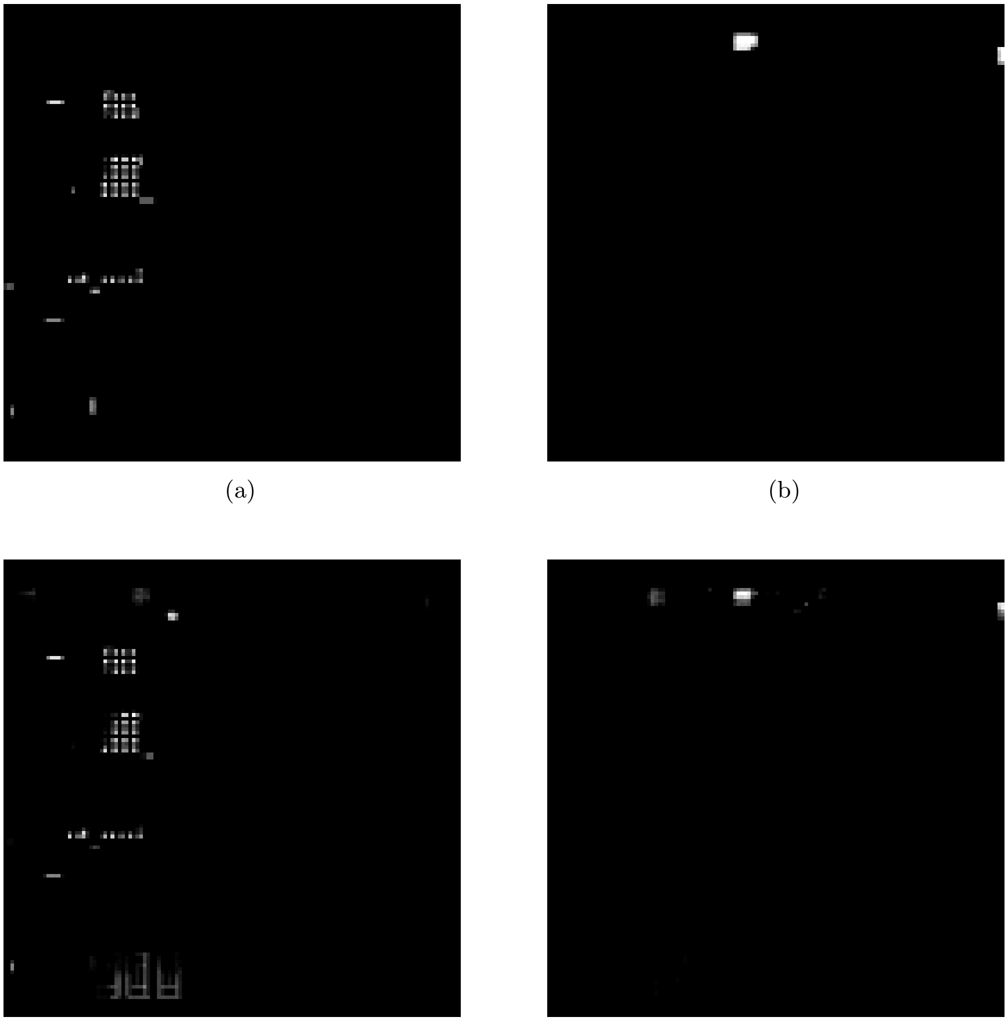
Table 6.2 shows the average performance for the 923 primary and 236 secondary results. Both CBAR approaches significantly reduce the MSE and increase the correlation coefficient, compared to the unscaled IC data. This improvement is especially pronounced for sparse materials. CBAR-X with five-classes, can extract a second material in over 25% of the ICs with performance metrics on par with those obtained for the primary materials and no negative impact on the accuracy of the primary material extraction.

6.4 Conclusion

The CBAR algorithm (three classes) was extended to use five-classes and called CBAR-X. CBAR has been shown to be an effective approach to rescale ICs for use in abundance

Table 6.1: Mean-square error and correlation coefficient magnitude for rescaling approaches applied to the independent component in Fig. 6.4(a).

	MSE	r
Independent component, white sheet metal	3.175	0.8026
CBAR, white sheet metal	0.0010	0.8218
CBAR-X, white sheet metal	8.5553×10^{-4}	0.8222
Independent component, brown siding	3.2027	-0.1455
CBAR, brown siding	0.0047	-0.0057
CBAR-X, brown siding	5.0969×10^{-4}	0.8540



(a) (b)

(c) (d)

Fig. 6.7: CBAR-5 rescaling results and corresponding truth maps. (a) Abundance truth map for white sheet metal material. (b) Abundance truth map for brown siding material. (c) Positive (right) CBAR-5 rescaling result. (d) Negative (left) CBAR-5 rescaling results.

Table 6.2: Average mean-square error and correlation coefficient magnitude for rescaling approaches.

	Sparse Materials		Dense Materials		All Materials	
	MSE	$\|r\|$	MSE	$\|r\|$	MSE	$\|r\|$
Independent component, primary material	16.3802	0.5028	4.2433	0.7387	10.5156	0.6168
CBAR, primary material	0.0048	0.8151	0.0272	0.7568	0.0156	0.7869
CBAR-X, primary material	0.0201	0.8241	0.0179	0.7645	0.0190	0.7953
Independent component, secondary material	6.9357	0.1346	12.6495	0.4802	7.7831	0.1858
CBAR, secondary material	0.0279	0.0290	0.0359	0.0114	0.0291	0.0264
CBAR-X, secondary material	0.0003	0.7917	0.0052	0.7596	0.0011	0.7870

quantification. However—as demonstrated in this chapter—when material spectra are negatively correlated CBAR does not retain all of the material information contained in the IC. By extending CBAR to use five, rather than three classes, this information is retained while still achieving the desired (and significant) reduction of mean-square error and increase of correlation coefficient. To make CBAR-X more practical, a better method should be developed to determine when a secondary material is present in an independent component. This may be possible by analyzing the results of the parameter estimation step of the rescaling algorithm. The relative distances between the means of the filled classes and mean of the empty class may be an indicator for the presence or absence of a second material. The CBAR-X results are promising, both in cases with a single material as well as cases where two materials are present in an IC. Experimental results show that two-material cases occur frequently. CBAR-X has the potential to enable unmixing approaches with scale ambiguity such as ICA to be used more effectively for abundance quantification.

References

[1] M. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, May 1994.

[2] J. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS '09*, pp. 1–4, Aug. 2009.

[3] D. R. Fuhrmann, "A simplex shrink-wrap algorithm," *Proceedings of SPIE*, vol. 3718, pp. 501–511, 1999.

[4] C. Bateson, G. Asner, and C. Wessman, "Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1083–1094, Mar. 2000.

[5] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE*, vol. 3753, pp. 266–275, 1999.

[6] J. Nascimento and J. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[7] 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, Mar. 1999.

[8] J. W. Boardman, F. A. Kruse, and R. O. Green, “Mapping target signatures via partial unmixing of AVIRIS data,” *Summaries of the Fifth Annual JPL Airborne Earth Science Workshop*, pp. 23–26, 1995.

[9] C.-I. Chang and A. Plaza, “A fast iterative algorithm for implementation of pixel purity index,” *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 63–67, Jan. 2006.

[10] D. 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.

[11] A. Plaza, P. Martinez, R. Perez, and J. Plaza, “Spatial/spectral endmember extraction by multidimensional morphological operations,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, Sept. 2002.

[12] S. Jia and Y. Qian, “Spectral and spatial complexity-based hyperspectral unmixing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.

[13] A. D. Stocker and A. P. Schaum, “Application of stochastic mixing models to hyperspectral detection problems,” *Proceedings of SPIE*, vol. 3071, pp. 47–60, 1997.

[14] J. Nascimento and J. Bioucas-Dias, “Hyperspectral unmixing algorithm via dependent component analysis,” *IEEE Geoscience and Remote Sensing Symposium*, pp. 4033–4036, July 2007.

[15] J. Nascimento and J. Bioucas-Dias, “Learning dependent sources using mixtures of dirichlet: Applications on hyperspectral unmixing,” *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS ’09*, pp. 1–5, Aug. 2009.

[16] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, “Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

[17] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. Hero, “Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery,” *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[18] X. Zhang and C. H. Chen, “New independent component analysis method using higher order statistics with application to remote sensing images,” *Optical Engineering*, vol. 41, pp. 1717–1728, 2002.

[19] J. Nascimento and J. Dias, “Does independent component analysis play a role in unmixing hyperspectral data?” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[20] T. Tu, "Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis approach," *Optical Engineering*, vol. 39, no. 4, pp. 897–906, Apr. 2000.

[21] B. R. Foy and J. Theiler, "Scene analysis and detection in thermal infrared remote sensing using independent component analysis," *Proceedings of SPIE*, vol. 5439, pp. 131–139, 2004.

[22] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 260–2616, Sept. 2006.

[23] C.-I. Chang, S. S. Chiang, J. A. Smith, and I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 2, pp. 375–392, 2002.

[24] S. Chiang, C.-I. Chang, and I. Ginsberg, "Unsupervised hyperspectral image analysis using independent component analysis," *IEEE Geoscience and Remote Sensing Symposium*, vol. 7, pp. 3136–3138, 2000.

[25] *The DIRSIG User's Manual*, Rochester Institute of Technology, Oct. 2006. [www.dirsig.org/documentation]

[26] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[27] E. J. Ientilucci and S. D. Brown, "Advances in wide-area hyperspectral image simulation," *Proceedings of SPIE*, pp. 110–121, 2003.

Chapter 7

Summary and Future Work

In this dissertation, the problem of utilizing independent component analysis for hyperspectral unmixing was explored. The material in Chapters 1–3 provided background, motivation, and context for the problem. In Chapters 4–6 various aspects of the problem were examined in more detail. Throughout, a number of informative observations and useful results were presented. Additionally, open questions and potential directions for future research were identified.

In Chapter 4, a quantitative assessment was made of spectral unmixing results obtained using the FastICA algorithm. Such an assessment was possible through the use of synthetically-generated but physically-realistic hyperspectral imagery. This assessment showed that FastICA produced results that were strongly correlated to ground truth abundance maps. This was especially true for sparse and super-sparse materials. Two of the fundamental steps of the FastICA algorithm, dimension reduction (using PCA) and orthogonalization were also assessed. The assessment of PCA demonstrated the importance of retaining a large number of dimensions in order to accurately unmix sparse materials. The assessment of orthogonalization showed that forcing unmixed results to be orthogonal was only detrimental when endmembers were spatially correlated. These results, combined with others in the published literature demonstrate the potential ICA has for unmixing sparse materials, complementing existing unmixing approaches.

Some observations from Chapter 4 point toward areas for future research. First, it was noted that ICA results seem to improve as materials become more sparse. This is in contrast to most existing spectral unmixing approaches which tend to focus only on dense materials. Thus, an approach that combines ICA and another approach such as N-FINDR or VCA may improve results over either algorithm individually. A mechanism for such a combination has yet to be explored. A second observation was the tendency for ICA to split a single endmember into two or more ICs. An additional step during or after the unmixing process to identify such splitting and merge independent components as appropriate has strong potential to improve ICA unmixing performance, especially for dense materials.

With the potential for ICA established, Chapters 5–6 focused on making independent components physically meaningful as abundances. In Chapter 5, a new algorithm, called class-based abundance rescaling (CBAR), was introduced. A quantitative comparison—again using synthetic data—was made between CBAR and existing abundance rescaling methods. This comparison showed CBAR performance to be superior both in reducing abundance error and increasing correlation with ground truth. An extended CBAR algorithm, CBAR-X, was developed in Chapter 6 which allows two negatively correlated materials appearing in a single independent component to be simultaneously separated and rescaled. It was shown that this can be done with no adverse impact on the quality of the results when compared to the CBAR algorithm.

Although the CBAR and CBAR-X results are superior to other rescaling algorithms, there are research areas which could enable further improvement. First, an automated method of determining whether an independent component contains one or two materials would prevent CBAR-X from producing superfluous results. Second, and somewhat more ambitious, is the integration of CBAR (or CBAR-X) with FastICA or a similar ICA algorithm. Rescaling the independent components prior to orthogonalization would be expected to improve the accuracy of subsequent results. Whether the nonlinear rescaling step and the linear orthogonalization step could be successfully combined is an open question.

In summary, the application of ICA to the spectral unmixing problem has been considered from a number of perspectives. The hurdle of rescaling independent components for use as abundance estimates has been addressed. The various experimental results considered herein have been generally positive, warranting continued utilization of ICA for hyperspectral unmixing while simultaneously illuminating areas for future research and potential improvement.

Appendices

Appendix A

Material Lists for Synthetic Test Images

Table A.1: Megascene 1, Tile 1 Test Image Materials.

ID	**Material Name**	**Total Pixels Present**	**Total Pure Pixels**	**Fractional Area**
Super–Sparse Materials (indicated by $\circ$ in plots)				
1	Siding, Mineral, Painted, Dark Green	1	0	0.016
2	Siding, Wood, Painted Off White, Fair	1	0	0.078
3	Tree, Black Oak, Bark	2	0	0.031
4	Siding, Cedar, Stained Dark Brown, Fair	2	0	0.078
5	Siding, Wood, Painted White, New, Rough	2	0	0.094
6	Brick, Old Carolina Brick Company, Charlestowne	2	0	0.453
7	Glass	3	0	0.047
8	Brick, Brampton Brick, Old School, Red	4	0	0.313
9	Siding, Vinyl, Off White, Fair	4	0	0.594
10	Roadway Surfaces, Sidewalk, Brick, Sealed, Mixed Color	4	0	0.813
11	Vinyl, Vision Pro Sample Board, Blue D-4	7	0	0.719
12	Roof Shingle, Asphalt, Mix Brown, Good	7	0	0.781
Sparse Materials ($\times$)				
13	Stone Siding, Apple Ridge, Buckingham Fieldstone	9	0	1.156
14	Sheet Metal, Gray, Shiny, Dusty	11	0	1.078
15	Swimming Pool (Lining and Water)	12	0	5.375
16	Siding, Wood, Planks, Brown	13	0	2.359
17	Siding, Wood, Painted Tan, Fair	15	0	2.313
18	Roof Shingle, Asphalt, Harmony Sample Board, Cove Gray	26	5	16.281
19	Roof Shingle, Asphalt, Eclipse Sample Board, Twilight Gray	27	0	1.859
20	Roof Shingle, Asphalt, Black, Weathered	29	16	20.516
21	Roof Shingle, Asphalt, Black, Fair	30	6	17.453
22	Roof Shingle, Asphalt, Eclipse Sample Board, Shadow Black	30	12	20.328
23	Roof Shingle, Asphalt, Dark Light, Fair	30	12	20.813
24	Roof Shingle, Asphalt, Brown and Red Blend, Fair	31	0	9.984
25	Roof Shingle, Asphalt, Eclipse Sample Board, Forest Green	35	15	24.281
26	Roof Shingle, Asphalt, Brown, Black, New	35	10	24.719
27	Brick, Siding, Mix Brown, Fair	44	13	33.750
28	Roof Shingle, Asphalt, Harmony Sample Board, Sequoia Tile	64	16	40.953
29	Brick, Brampton Brick, Old School, Brown	76	0	13.672
30	Tree, Dogwood, Leaf	77	3	30.797
31	Brick, KF Plymouth Blend, Red Brick	84	0	14.563
32	Tree, Maple, Trunk	140	0	5.406
33	Tennis Court, Playing Surface, White Line	194	0	37.625
Intermediate Materials ($\Diamond$)				
34	Tree, Black Oak, Leaf	212	14	77.469
35	Sheet Metal, White, Fair	222	0	58.188
36	Tennis Court, Playing Surface, Red	250	67	155.688
37	Tennis Court, Playing Surface, Green	262	59	175.625
38	Tree, Norway Maple, Leaf	1005	196	632.422
39	Tree, Silver Maple, Leaf	1299	717	1013.297
40	Tree, Red Maple, Leaf	1360	7	611.625
41	Roof, Gravel, Gray	2373	1845	2176.047
Dense Materials ($\Box$)				
42	Asphalt, Black, New	8198	2928	4975.422
43	Grass, Brown and Green w/ Dirt	9275	3124	6158.922

Table A.2: Megascene 1, Tile 4 Test Image Materials.

ID	Material Name	Total Pixels Present	Total Pure Pixels	Fractional Area
Super–Sparse Materials (indicated by ◦ in plots)				
1	Sheet Metal, Maroon, Shiny, Fair	1	0	0.016
2	Tree, Maple, Trunk	1	0	0.016
3	Tree, Red Maple, Leaf	1	0	0.016
4	Tree, Dogwood, Trunk	2	0	0.031
5	Brick, Old Carolina Brick Company, Charlestowne	2	0	0.266
Sparse Materials (×)				
6	Sheet Metal, Black, Shiny, Dirty	21	0	1.969
7	Brick, Hampton Brick, Sandmist	36	0	2.875
8	Concrete, Cinder Blocks, Textured	68	17	40.344
9	Brick, Mixed Tan and Caramel Colors	82	0	10.234
10	Brick, Old Carolina Brick Co., Savannah Gray	102	10	45.578
11	Sheet Metal, White, Fair	183	0	55.734
Intermediate Materials (◊)				
12	Tree, Silver Maple, Leaf	206	117	165.953
13	Sheet Metal, Tan, Shiny, Fair	276	187	229.172
14	Building Roof, Painted Metal, Gray, Weathered	333	129	224.063
15	Tree, Dogwood, Leaf	370	27	175.938
16	Sheet Metal, Gray, Shiny, Dusty	660	9	101.656
17	Tree, Norway Maple, Leaf	667	182	401.734
18	Siding, Vinyl, Tan, Fair	1115	771	938.859
19	Roof, Gravel, Gray	1194	767	998.188
Dense Materials (□)				
20	Grass, Brown and Green w/ Dirt	8718	2534	5864.234
21	Asphalt, Black, New	10880	4726	7127.125

Appendix B

Parameter Estimation for a Gaussian-Corrupted Beta Distribution

A potentially useful model consists of data from a beta distribution which corrupted by additive Gaussian noise. Let $V \sim \mathcal{N}(\mu, \sigma^2)$, $W \sim B(\alpha, \beta)$, and $X = V + W$. Assuming the noise is zero–mean, this model is parameterized by three values, α, β, and σ^2. One approach to estimating these parameters from the data is the method of moments [1]. This approach consists of describing the moments of a distribution in terms of the distribution parameters and equating those to the corresponding sample moments. When the number of samples is small, the method of moments approach has been shown to be superior to maximum-likelihood (ML) estimation of beta distribution parameters [2].

Let x_n denote the nth sample of the random variable X where there are N total samples available. The sample moments are obtained as

$$\bar{m} = \frac{1}{N} \sum_{n=1}^{N} x_n, \tag{B.1}$$

$$\bar{r} = \frac{1}{N} \sum_{n=1}^{N} x_n^2, \tag{B.2}$$

and

$$\bar{t} = \frac{1}{N} \sum_{n=1}^{N} x_n^3. \tag{B.3}$$

Before looking at the moments of X it is useful to examine the moments of V and W individually. Since there are three parameters to be estimated, the first three moments will be needed. The moments of order t of the beta distribution are given by [2]

$$M_W^{(t)} = E\left[W^t\right] = \frac{\Gamma(\alpha + t)\,\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + t)\,\Gamma(\alpha)}. \tag{B.4}$$

The moments of the Gaussian distribution can be obtained from the moment generating function as

$$M_V^{(t)} = E\left[V^t\right] = \frac{d^t}{ds^t} \exp\left\{s\mu + \frac{1}{2}\sigma^2 s^2\right\}\Bigg|_{s=0}. \tag{B.5}$$

The moments needed for this problem, assuming a zero-mean Gaussian, are shown in Table B.1. The moments of X can be obtained from $E\left[X^t\right] = E\left[(V+W)^t\right]$. Expanding the products, exploiting the independence of V and W, and using the values in Table B.1 yields

$$M_X^{(1)} = E[W] = \frac{\alpha}{\alpha+\beta}, \tag{B.6}$$

$$M_X^{(2)} = E\left[W^2\right] + E\left[V^2\right] = \frac{(\alpha+1)\,\alpha}{(\alpha+\beta+1)\,(\alpha+\beta)} + \sigma^2, \tag{B.7}$$

and

$$M_X^{(3)} = E\left[W^3\right] + 3E\left[W\right]E\left[V^2\right] = \frac{(\alpha+2)\,(\alpha+1)\,\alpha}{(\alpha+\beta+2)\,(\alpha+\beta+1)\,(\alpha+\beta)} + \frac{3\alpha\sigma^2}{\alpha+\beta}. \tag{B.8}$$

The sample moments in (B.1), (B.2), and (B.3) can then be equated to the moment formulas in (B.6), (B.7), and (B.8), respectively. Solving these equations for the parameters of interest yields

$$\hat{\sigma}^2 = \frac{-b \pm \sqrt{4ac + b^2}}{2a}, \tag{B.9}$$

Table B.1: The first three moments of the beta and zero-mean Gaussian distributions.

Order	Beta	Gaussian
$M^{(1)}$	$\frac{\alpha}{\alpha+\beta}$	0
$M^{(2)}$	$\frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}$	σ^2
$M^{(3)}$	$\frac{(\alpha+2)(\alpha+1)\alpha}{(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)}$	0

where

$$a = 4\bar{m} - 2, \tag{B.10}$$

$$b = 6\bar{m}^3 - 4\bar{m}^2 - 5\bar{m}\bar{r} + 4\bar{r} - \bar{t}, \tag{B.11}$$

and

$$c = \bar{m}^2\bar{r} - 2\bar{m}^2\bar{t} + \bar{m}\bar{r}^2 + \bar{m}t - 2\bar{r}^2 + \bar{r}\bar{t}. \tag{B.12}$$

Once σ^2 has been estimated, α and β can be obtained as

$$\hat{\alpha} = \frac{\bar{m}\left(\bar{r} - \hat{\sigma}^2 - \bar{m}\right)}{\hat{\sigma}^2 + \bar{m}^2 - \bar{r}}, \tag{B.13}$$

and

$$\hat{\beta} = \frac{(1 - \bar{m})\left(\bar{r} - \hat{\sigma}^2 - \bar{m}\right)}{\hat{\sigma}^2 + \bar{m}^2 - \bar{r}}. \tag{B.14}$$

References

[1] L. Bain and M. Engelhardt, *Introduction to probability and mathematical statistics*, ser. Classic Series. Boston: Duxbury, 1992.

[2] C. B. Owen, "Parameter estimation for the beta distribution," Master's thesis, Brigham Young University, Provo, UT, 2008.

Vita

Matthew R. Stites

(April 2012)

Education

- Ph.D., Electrical Engineering, Utah State University, May 2012 (expected)
  Dissertation Topic: Statistical Approaches for Blind Hyperspectral Unmixing
- M.E., Electrical Engineering, Utah State University, December 2005
- B.S., Electrical Engineering, Utah State University, May 2002

Professional Experience

- Scientist, Space Dynamics Laboratory, North Logan, UT, July 2010 to present
- Student Affiliate, Los Alamos National Laboratory, Los Alamos, NM, August 2008 to November 2008
- Electrical Engineer, Space Dynamics Laboratory, North Logan, UT, May 2002 to August 2006

Published & Prepared Journal Articles

- Using Physically-Modeled Synthetic Data to Assess Hyperspectral Unmixing Approaches, M.R. Stites, J.H. Gunther, T.K. Moon, and G.P. Williams, (to be submitted for publication).
- Rescaling Independent Components for Abundance Quantification, M.R. Stites, J.H. Gunther, T.K. Moon, and G.P. Williams, (to be submitted for publication).

- An Extended Algorithm to Rescale Independent Components for Abundance Quantification, M.R. Stites, J.H. Gunther, T.K. Moon, and G.P. Williams, (to be submitted for publication).
- Detection of Amorphously Shaped Objects Using Spatial Information Detection and Enhancement (SIDE), C.S. Grant, T.K. Moon, J.H. Gunther, M.R. Stites, and G.P. Williams, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (accepted for publication).

Published Conference Papers

- A Bayesian Framework for Abundance Estimation in Hyperspectral Data Using Markov Random Fields, M.R. Stites, T.K. Moon, J.H. Gunther, and G.P. Williams, *IEEE Forty-First Asilomar Conference on Signals, Systems, and Computers*, 2007.

Awards

- Space Dynamics Laboratory Tomorrow Fellowship, 2006-2009