Universiteit
Antwerpen

Faculteit Wetenschappen
Departement Fysica

Advances in Unmixing of Hyperspectral Remote Sensing Imagery

Vorderingen in ontmenging van hyperspectrale aardobservatiebeelden

Proefschrift voorgelegd tot het behalen van de graad van

Doctor in de Wetenschappen: Fysica

aan de Universiteit Antwerpen te verdedigen door

Dževdet Burazerović

Promotors:
Prof. Dr. Paul Scheunders
Dr. Rob Heylen

Antwerpen, 2014

Doctoral committee

Chairman

Prof. Dr. Jacques Tempere	(Department of Physics)

Members

Prof. Dr. Paul Scheunders	(Department of Physics, Vision Lab)
Prof. Dr. Nick Schryvers	(Department of Physics, Electron Microscopy for Materials Science – EMAT)
Dr. Rob Heylen	(Department of Physics, Vision Lab)

External members

Dr. Raul Zurita-Milla	(Faculty of Geo-Information Science and Earth Observation, University of Twente)
Dr. Walter Debruyn	(Flemish Institute for Tech. Research – VITO)

Contact information

✉ Dževdet Burazerović
Vision Lab, Dept. of Physics
University of Antwerp
Campus Drie Eiken, D.N.102
Universiteitsplein 1
B-2610 Wilrijk – Antwerp, Belgium

☎ +32 (0) 3 265 24 78

+31 (0) 6 457 66 215

+32 (0) 3 265 22 45

dzevdet.burazerovic@uantwerpen.be

d.burazerovic@chello.nl

http://www.ua.ac.be/dzevdet.burazerovic

To my mother and brother

ACKNOWLEDGMENTS

After fourteen versatile years of working as a researcher, reflecting on the past four-and-a-half years seems difficult without putting too much thought and word, or too little. Certainly, I am most pleased and obliged to acknowledge the support and contributions of which some might otherwise go unnoticed but without all of which this thesis would not have been possible.

First and foremost, I thank my promoter Prof. Paul Scheunders for providing the opportunity to pursue this research at the Vision Lab, for his generous support and continued engagement and perspective. I am indebted to my co-promoter Rob Heylen for his vital contribution and pleasant cooperation throughout this research, as well as for his undemanding attitude and much motivation I received through our broader and nontechnical conversations.

I wish to thank all the researchers and the involved staff from the Vision Lab and the University of Antwerp for making my work and stay there a pleasant experience. This undoubtedly helped to make my commuting between Eindhoven and Antwerp endurable. A special thanks goes to researchers Zahid Mahmood and Guy Thoonen for helpful discussions and sharing of useful information and advice. I also appreciated a good sense of humor from all the above mentioned people.

Certainly, I wish to thank all members of the doctoral committee for critically reviewing this thesis and providing useful remarks that have helped to shape its final form and appearance.

Last but not least, I thank my mother and my brother with his family, as well as my other close relatives and friends (Tarik and Azra in particular), for providing me moral support during this venture.

Dževdet Burazerović,
Eindhoven, June 2014

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

AVIRIS	Airborne visible/infrared imaging spectrometer
CCD	Charged coupled device
CDOM	Colored dissolved organic matter (substance and concentration)
$C(\boldsymbol{E})$	Column space of matrix $\boldsymbol{E}$
CHL	Chlorophyll (substance and concentration)
DN	Digital number – a generic term for pixel value
DR	Dimensionality reduction
$\boldsymbol{E}$	Matrix containing spectra of endmembers as its columns
EE	Endmember extraction
FCLSU	Fully constrained least-squares unmixing
GBM	Generalized bilinear model
IOP	Inherent optical property
kNN	k nearest neighbors
MDS	Multidimensional scaling
ML	Manifold learning
N-findR	EE algorithm exploiting simplex-volume maximization
NLDR	Nonlinear dimensionality reduction
NIR	Near-infrared wavelength range
OLS	Ordinary least squares
PCA	Principal components analysis (transform)
RMSE	Root-mean-square error
ROI	Region of interest in an image
SA	Spectral angle
SIOP	Specific inherent optical property
SNR	Signal-to-noise ratio
SPU	Simplex projection unmixing algorithm
TSM	Totally suspended matter (substance and concentration)
VCA	Vertex component analysis algorithm

CHAPTER 1

INTRODUCTION

Abstract

This chapter explains the domain and scope of our study and introduces in steps a subtopic this study addresses – unmixing of hyperspectral imagery acquired by remote-sensing and Earth observation platforms. The almost exclusive concern of our work is image- and data analysis and processing. However, to offer the reader a broader view, we start by giving a word about the Earth-observation chain and all its elements. The subsequent section then zooms onto a modality of interest, which is the hyperspectral imaging. Hyperspectral image processing differs from standard image processing in that each pixel can be treated individually and used to identify materials in the imaged scene. So we spend few words on explaining the relevant definitions and concepts. The final section is devoted to the mixed-pixel problem and spectral unmixing. This topic has seen a notable concentration of interest and effort from researchers over the years. Here, we try to describe it in a way that will provide an insight into the current state-of-the-art, accentuating those definitions and constructs that are most relevant for understanding the subsequent chapters.

1.1 Remote sensing and Earth observation

To give a crisp definition of remote sensing is not trivial. Perhaps the most integral definition speaks of a science of acquiring information about an object or surface of interest without actually being in contact with it [Lillesand et al., 2008]. Does this mean that standard photography is also remote sensing? Yes and no. Yes, if the focus is on a class of photos captured from aircrafts and satellites. In fact, for a long time *aerial photography* was a synonymous for *remote sensing*, until the latter term was coined in the 1960s in recognition of the proliferation of new modalities (e.g. radar) recording the electromagnetic (EM) energy outside the range of visible light [Campbell, 2006]. This historical legacy explains why remote sensing technology and techniques are mostly discussed in the context of monitoring the Earth's surface- and atmospheric processes, or *Earth Observation* for short [Campbell, 2006; Cracknell and Hayes, 2007]. At the same time, one ought to acknowledge that the same sensory and imaging techniques that are used for Earth observation are also used for various other purposes. Examples include remote sensing (imaging spectrometry) for pharmaceutical and industrial process and quality control, food safety, and biometric and forensic applications (see [Bioucas-Dias et al., 2012] and references therein). There are also examples where remote-sensing modalities are used in a way that is more attributable to other domains, such as *video processing* or *computer vision* (think e.g. of the use of thermal infrared imaging for enhanced vision or surveillance). It is thus clear that the rightfulness of calling some process 'remote sensing' much depends on the context and specifics of all the elements in the processing chain: the object, the sensor(s), the method and the application.

Because our algorithms will work with images intended for Earth observation, we ought to start by introducing the related concepts. A generous amount of detail on this topic is provided by the above mentioned works and on-line sources (see [Natural Resources Canada, 2014]), so we constrain ourselves to the essentials. We start by giving a rough sketch of the Earth-observation process, in Fig. 1.1. There are several elements that need to be distinguished and understood from this picture:

- **Energy source** – emits EM energy that is partly absorbed and partly reflected by objects on the Earth's surface. Two basic source types are recognized: naturally available illumination sources, such as the Sun, and transceivers mounted on board of satellites and aircrafts (or used on the ground), such as the Synthetic Aperture Radar (SAR), Light Detection And Ranging (LIDAR), Sound Navigation And Ranging (SONAR), etc. Note that the Earth itself also acts as a source, as it emits radiation in the form of heat that is detectable in the thermal infrared (IR) portion of the EM spectrum (see also Fig. 1.2).

- **Radiation transfer** – connotes the passing of EM energy through the atmosphere, from its source to the target and backwards. The atmosphere interferes with the radiation via particle absorption (by gas molecules) and scatter (Mie and Rayleigh scattering on water droplets, ice crystals and aerosol). Since these effects depend on the particle size and wavelength of the radiation, some radiation types (e.g. solar) are affected more than others. *Atmospheric correction* models are therefore often used to 'rectify' the images recorded by the sensor [Cracknell and Hayes, 2007].

- **Interaction with the target** – once the radiation reaches the Earth's surface, it interacts with its objects: ground covers, water content, man-made structures, etc. By absorbing and reflecting the incident radiation differently over the entire EM spectrum, each class of objects creates a distinctive signature.

- **Sensor** – picks up and records a percentage of the radiation that is reflected by the imaged surface or target. Two basic types of sensors are distinguished: *passive sensors*, which only record energy that is readily available (solar or thermal) and *active sensors*, which themselves emit and direct radiation towards the target. Recall also what has been said above the energy sources.

- **Transmission, reception and processing** – recordings made by the sensor are transferred as digital data to a receiving and processing station that renders this data into an image specific for each modality. Depending on the sensors used, this process may include *geometric correction* of data acquired by multiple sensors, or by a single sensor during multiple passes or flight lines of satellites and aircrafts, and the aforementioned atmospheric correction.

- **Interpretation and analysis** – refers to visual or computerized processing of the image with the purpose of inferring information about its content. Some of the typical themes are *classification* (grouping of pixels in an image into a discrete number of categories or classes), *fusion* (combining images acquired by multiple sensors at different spatial and/or spectral resolutions), *unmixing*, *target detection*, *change detection*, etc., [Camps-Valls et al., 2011].

- **Application** – the information inferred from the image is used to reveal some property about the imaged scene or target, or to facilitate solving of particular problem. The applications are numerous and can be categorized from various angles. One way is to depart from the object and discern applications to the *atmosphere*, *geosphere*, *biosphere* and *hydrosphere* [Cracknell and Hayes, 2007], or *urban analysis* [Soergel, 2010]. Within the third category, e.g., one may further discern applications to forestry, agriculture and land-cover mapping

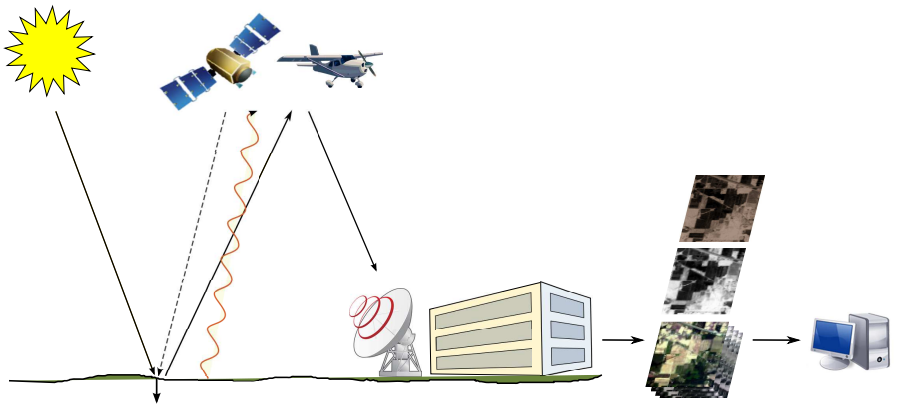

Figure 1.1: Remote sensing of the Earth's surface: The passive and active recording of EM energy yield imagery in multiple modalities: hyperspectral, SAR, thermal, etc.

[Borengasser et al., 2010; Giri, 2012]. Or within the fourth, the oceanography, observation of ground- and surface waters and river flows, water-quality retrieval in inland- and coastal waters, etc. [Robinson, 2010; Su et al., 2011]. Another take is to discern the actual usage scenarios, such as the real-time or seasonal monitoring and assessment, the provision of thematic maps for *geographic information systems* (GIS) [Mesev, 2007], etc.

At this point, it is clear that handling of the entire Earth-observation chain involves multidisciplinary work (we didn't even mention the aspect of conducting field studies and in *situ* measurements needed to calibrate the instruments or construct reference data for the algorithms) that is typically carried out by nationally and internationally governed bodies and institutes (NASA, ESA, CSA, JAXA, etc.) While the separation between many processes in the chain is inherently technical, some is naturally and historically influenced by legal, policy and economic factors [Johnston and Cordes, 2003]. This detachability of some processes from the others has also created much opportunity to develop innovative image- and data processing algorithms, as our later discussion will demonstrate. The downside is that the optimal design approach, as well as validation and commercialization of various solutions, become more difficult without access to all relevant data and information, such as a ground truth or a specification of end-user requirements.

In the sequel, our focus will be exclusively on image analysis. This means that all other aspects explained above will be treated as being out of scope or given.

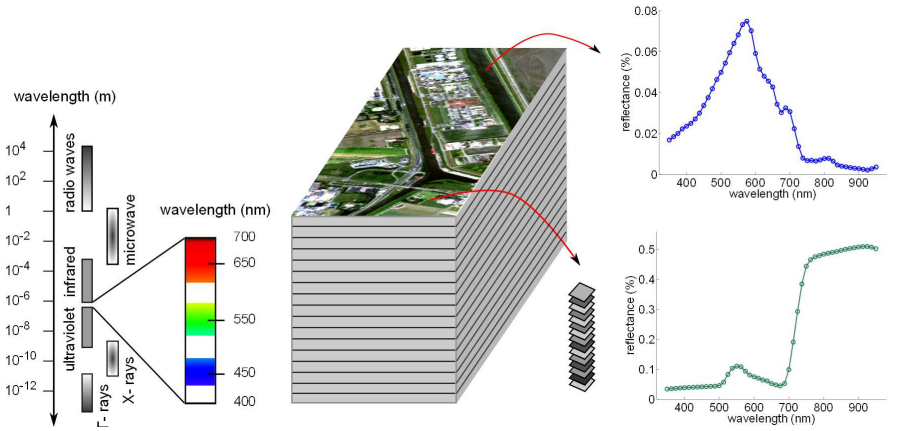

Figure 1.2: Hyperspectral data cube: Each gray-level image corresponds to a different wavelength, making each pixel a spectrum. Here water and vegetation are shown.

1.2 Hyperspectral imaging

A modality that provides the bulk of data for remote sensing applications, partly due to its inherent readability by a human user, is *optical remote sensing* [Prasad et al., 2011]. This modality records radiation in the visible and reflective infrared (IR) part of the EM spectrum: 400-2500nm. An electro-optical sensor collects the light reflected by an illuminated surface in multiple spectral channels or bands, each corresponding to a certain wavelength. By 'stacking' the images recorded at different wavelengths, one forms a data cube in which each pixel becomes a spectrum – the record of how light is reflected by some part of the surface. See Fig. 1.2. Depending on the sensor, one gets a multi-spectral image (MSI) containing a small number of non-uniformly spaced bands, or a hyperspectral image (HSI) having hundreds of contiguous co-registered bands [Prasad et al., 2011]. Using only 3 bands – one for the red, the green and the blue region, yields a plain color composite photo.

A spectrum is a plot of a signal versus some property of photons, such as energy, wavelength, frequency, etc. This signal is produced by the absorption or emission of a photon from an atom or molecule, which changes their state according to wave-functions Ball [2001]. One way of recording a spectrum is therefore to observe the ratio of the power of incident light illuminating the sample and outgoing light reflected by that sample. Other types of spectrometers rather observe the polarimetric properties of light [Schott, 2009]. The acquisition of spectra in all different ranges of

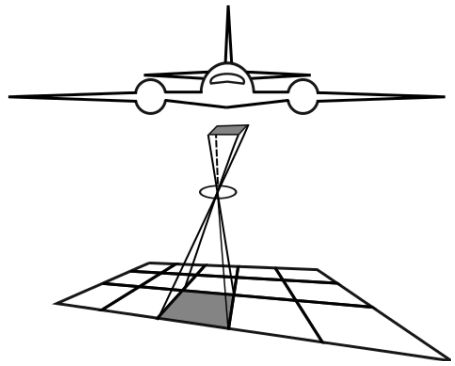

Figure 1.3: Illustration of the instantaneous field of view (IFOV).

the EM spectrum is studied within the field of *spectroscopy* Ball [2001].

When a spectroscopic device records multiple spectral bands and provides two-dimensional coverage, it is called an *imaging spectrometer*, or in the image-processing jargon simply a multi- or hyperspectral camera [Rees, 1999]. Such devices typically do not form the image instantaneously but build it up by scanning. This is achieved by using rotating scan mirrors and other optics and electronic detectors (e.g. CCD arrays), whose operation is outside the scope of this text. More detail is found in e.g. [Cracknell and Hayes, 2007].

Several aspects and parameters are important for qualifying the operation and capabilities of imaging spectrometers used for Earth observation:

- **Spatial resolution** is principally determined by the *instantaneous field of view* (IFOV). This is the area of the imaged surface that is projected through the sensor's optics onto a single detector element [Rees, 1999]. It is also defined as an angular cone of visibility of the sensor, which is converted to a spatial area via multiplication with the distance from the ground to the sensor [Abid, 2005]. The IFOV normally equals the size of a pixel, although the pixel may be smaller if the image is over-sampled. Refer also to Fig. 1.3.

- **Spectral resolution** concerns the ability of a sensor to discern wavelengths in the recorded signal. The related *spectral sensitivity* specifies the interval between two wavelengths within which the sensor's response drops to one half of its maximum. This enables to define the spectral resolution as the narrowest interval or *bandwidth* resolvable by the sensor [Campbell, 2006].

- **Signal to noise ratio (SNR)** measures the change of the recorded brightness or *radiance* (angular-area flux density) relative to the noise power in the sensor. This change is proportional to both the spectral resolution and the square of spatial resolution [Landgrebe, 2005]. So, for a given SNR, one resolution cannot be increased without decreasing the other. Another formula proves the same by showing that the *area coverage rate* is proportional to the square of the ground sample area, and inversely proportional to the squares of the SNR and the number of spectral bands [Shaw and hua K. Burke, 2003].

- **Radiometric resolution** describes the sensitivity of the sensor to changes in the magnitude of the reflected radiation (so it is measured in decibels). Because the recorded power can be increased by letting the radiation be received from a larger spatial area or a wider spectral bandwidth, radiometric resolution can be increased at the expense of spatial or spectral resolution.

- **Temporal resolution** is defined by the time interval between consecutive observations of a same location. For a spaceborne system, this is determined by the satellite orbit, the swath width of the instrument, and whether the instrument has a fixed or variable viewing geometry [Rees, 1999].

One process that is particularly relevant for optically recoded images concerns their projection onto geographic latitude/longitude or a chosen coordinate grid, which is known as *geometric correction* [Toutin, 2004]. This involves removal of both systematic and random distortions, where the first group is inherent to the imaging process and includes, e.g., a tilt of the focal plane relative to the aircraft position, panoramic distortion and misalignment of images acquired from multiple flight lines, etc. Some of these errors require the use of photo control points connecting the image pixels with their true geographic location [Morgan and Falkner, 2001].

Yet another important concept is the aforementioned *atmospheric correction*. The transmission of EM radiation through the atmosphere is described by the radiative transfer equation, but its parameters are not always precisely estimable, so empirical methods are often developed that exploit simulations and knowledge about the type of radiation involved (microwave, thermal, visible light, etc.) [Rees, 1999]. A distinction is also made as to whether a method calibrates with in *situ* measurements, uses historical and meteorological data in atmospheric models, or treats the atmospheric effects on a pixel-by-pixel basis [Cracknell and Hayes, 2007].

Finally, it is worth pointing out that all optical sensors in principle record the radiance as *digital numbers* (DN). Since DN obtained from different scenes are not comparable in the absolute sense, they are converted to radiance in a way that requires knowledge of calibration specific to each instrument [Campbell, 2006]. A

general formula is: $L_\lambda = gain \cdot \mathrm{DN} + bias$. The recorded radiance, in turn, can be converted to *reflectance*, which is the relative expression of how incident radiation is reflected by the target (see Fig. 1.2). This is done by taking into account the solar irradiance, the solar zenith angle, the bidirectional reflectance distribution function of a Lambertian surface, etc. [Liou, 2002]. From now onwards, all our mention of 'spectra' will assume the use of reflectance, unless stated otherwise.

1.3 The spectral mixing problem

Spectral mixing refers to the situation where multiple constituent spectra make up the spectrum of a single target. While signal mixtures are studied in various scientific fields and problems, spectral mixing is an underlying limitation of remote sensing. Because hyperspectral cameras record the radiance in numerous spectral bands, they do so with constrained spatial resolution (see above). A consequence is that the recorded imagery will often include mixed pixels displaying a plurality of disparate materials, each covering the pixel area by a certain fraction.

The unraveling of mixed pixels has been widely studied as spectral unmixing, for which various linear and nonlinear approaches have been adopted and developed over the years [Keshava and Mustard, 2002; Bioucas-Dias et al., 2012]. The main assumption of the prevalent linear unmixing paradigm is that all the materials in a mixture will contribute to the total reflectance independently, as the incident photons will interact with each material once before reaching the sensor. This is viable when the mixing scale is macroscopic and the materials appear in a pixel mostly as spatially segregated regions (e.g. [Fisher, 1997] distinguishes four such configurations). On the other hand, the nonlinear models presume non-negligible multiple reflections of photons between the materials, which may be provoked by their three-dimensional topography or the intimacy of their mixture. See Fig. 1.4.

1.3.1 Linear unmixing

The linear spectral mixing in remote sensing is commonly presented via the linear mixing model (LMM). This model expresses the mixed spectrum of a pixel as a linear sum of spectra of intrinsic or 'pure' materials, the so-called *endmembers*, having weights or *abundances* that are positive and sum to one [Boardman, 1994]. The two constraints on the abundances tell that only the whole pixel area and no negative sub-pixels are taken into account, which is sensible to demand from the physical perspective. The accompanying mathematical formula is given by Eq. (1.1), where $\boldsymbol{x}$ is a $d \times 1$ vector representing a pixel spectrum, $\{\boldsymbol{e}_i\}$ are the equally sized

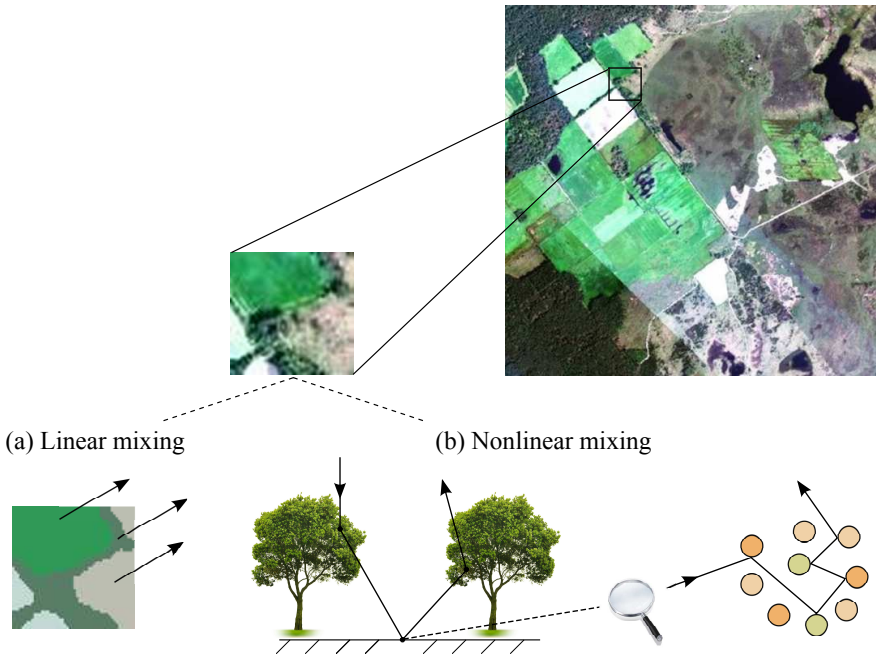

Figure 1.4: A view of spectral mixing: Depending on the viewing scale, different ground covers (e.g. vegetation and soils) add to the pixel's reflectance independently (linearly) as separate regions, or interactively (nonlinearly) via multiple reflections.

spectra of endmembers, and $\{a_i\}$ are the corresponding abundance coefficients. Furthermore, p is the known or estimated number of endmembers, while $\boldsymbol{n}$ stands for additive model errors (e.g. due to mis-estimation of $\{\boldsymbol{e}_i\}$) and noise. In most practical situations $\boldsymbol{n} \neq 0$, so one normally hopes to minimize the *reconstruction error*, i.e. the difference between an original data point $\boldsymbol{x}$ and its approximation $\widehat{\boldsymbol{x}} = \sum_i a_i \boldsymbol{e}_i$, as measured by its Euclidean norm: $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}\|_2$ or some other out of several applicable criteria and scores [Christophe et al., 2005].

More broadly, Eq. (1.1) has been known as the definition of a *p–simplex*, or a Euclidean geometric spatial element having the minimum number of boundary points [Elte, 1912]. This corresponds to a line segment in one-dimensional space ($p = 2$), a triangle in two-dim. space ($p = 3$), a tetrahedron in three-dim. space ($p = 4$), etc. See Fig. 1.5. Another general definition destines the $\{a_i\}$ from Eq. (1.1) as the normalized *barycentric coordinates* of $\boldsymbol{x}$ with respect to $\{\boldsymbol{e}_i\}$ [Coxeter, 1963].

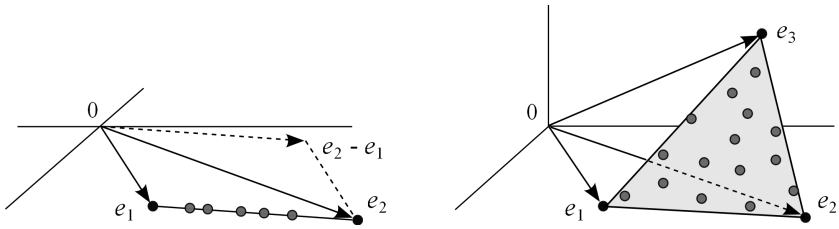

Figure 1.5: Geometry of Eq. (1.1) for $p = 2$ and $p = 3$. When both constraints on $\{a_i\}$ are obeyed (and $\boldsymbol{n} = 0$), all data points $\boldsymbol{x}$ reside within a p–simplex defined by $\{\boldsymbol{e}_i\}$.

$$\boldsymbol{x} = \sum_{i=1}^{p} a_i \boldsymbol{e}_i + \boldsymbol{n}, \quad \forall i : \; a_i \geq 0 \,, \sum_{i=1}^{p} a_i = 1 \tag{1.1}$$

Recently, an exhaustive survey of approaches to solving Eq. (1.1) was given by [Bioucas-Dias et al., 2012], discerning geometrical, statistical, sparsity-based and spatial-contextual methods. The first group basically take advantage of the above explained convex geometry of the LMM equating the endmembers to vertices of a simplex. Hence, most methods from this group estimate the $\{\boldsymbol{e}_i\}$ from a data through an independent process of *endmember extraction* [Veganzones and Graña, 2008], by searching for purest pixels or the minimum-volume embedded or enclosing simplex in the data [Plaza et al., 2011]. Once the $\{\boldsymbol{e}_i\}$ are known, the inversion of Eq. (1.1) needed to determine the $\{a_i\}$ is typically realized through optimization techniques. We discuss more about this topic in Chapter 3.

The other three mentioned groups all involve approaches that use priors or impose extra constraints on the solution of Eq. (1.1). For example, statistical approaches use parametric techniques and estimation frameworks (e.g. Bayesian) to address the variability in the data implied by its high degree of mixing or the uncertainty of endmembers. These methods often jointly estimate the endmembers and abundances [Arngren et al., 2011; Dobigeon et al., 2009]. (Note that the variability of endmembers has also been widely treated outside the statistical framework, by using the spectral feature extraction, signal transformations, etc. [Somers et al., 2011].) Another take considers the situation where a mixture is expected to be composed of few endmembers. Some methods from this group use sparsity-promoting priors to automatically determine the endmembers, including their number [Zare and Gader, 2007]. Others perform sparse regression, where some endmembers are expected to be known beforehand (e.g. from in *situ* or lab measurements), which translates the problem to that of finding the optimal (small) subset of endmembers in a (large)

spectral library to be used in each pixel [Iordache et al., 2011, 2014]. Finally, the group of spatial-contextual methods essentially include additional constraints derived from the assumption about local homogeneity of pixels (or abundances) into the other frameworks. Many of these works follow the idea originally used for classification, where pixels were no longer treated individually but by taking into account the correlations between spatial and spectral neighbors [Tarabalka et al., 2010]. Much of the applied methodology also coincides with that used for classification, including the use of convolution filters, Markov random fields, mathematical morphology, or spatial transforms [Bioucas-Dias et al., 2012; Camps-Valls et al., 2011].

An important aspect for any of these approaches concerns the determination of the number of endmembers, p in Eq. (1.1). Many studies fix this number empirically using prior knowledge about the imaged scene. Others try to estimate it from the data, which is the same as gauging its intrinsic dimensionality. We have seen that, if a data is a mixture of p endmembers, then it will reside within a p–simplex, so the data may be projected onto a $(p-1)$–dimensional subspace of the original space. This is why the problem has also been addressed as *subspace identification*, often by devising specific criteria or treatments for the eigenvalues of the data-derived correlation and noise matrices [Chang and Du, 2004; Bioucas-Dias and Nascimento, 2008]. We explain more about dimensionality reduction in Sec. 1.3.3.

Finally, it should be mentioned that solving Eq. (1.1) while in fact disregarding one of the constraints has also been considered in unmixing-related applications, such as the sub-pixel target detection [Duran and Petrou, 2009]. Notice that omitting the positivity $a_i > 0$ positions the data on a hyperplane containing the simplex (in Fig. 1.5 a line extending the segment $\overline{\boldsymbol{e}_1\boldsymbol{e}_2}$ and the entire plane of $\triangle \boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3$).

1.3.2 Nonlinear unmixing

The use of nonlinear unmixing in the remote sensing context has been acknowledged in taxonomies, surveys and feasibility studies stemming from more than a decade ago [Liangrocapart and Petrou, 1999; Keshava et al., 2000]. However, since recently the subject has received an expanding amount of interest from many authors [Bioucas-Dias et al., 2012]. This is also recognized in the fact that elaborate surveys are now starting to appear that provide an update on the use of exclusively nonlinear unmixing techniques [Dobigeon et al., 2014; Heylen et al., 2014].

One way to assort nonlinear unmixing is to discern model-based and data-driven approaches. While the first group typically tries to understand and describe the mechanism producing a nonlinearly structured data, the second opts to transform or learn this structure from the data itself. Typical examples from the first category are studies that regard the secondary illumination prompted by topography of

the ground covers such as tree canopy, orchards, low vegetation, etc. [Chen and Vierling, 2006; Somers et al., 2009]. A model that has been found suitable for this type of mixing is the *bilinear model*, which extends the LMM by adding the cross terms accounting for the (first-order) reflections between the endmembers. This is described by Eq. (1.2), where '$\odot$' takes the Hadamard or entry-wise product of two vectors. There are several variants of this equation (some are listed in [Halimi et al., 2011a]), and its extension with quadratic and higher-order powers is in theory also justified (observe a double reflection on the *tree* class in Fig. 1.4). However, for the reflectance data that is by definition confined to interval [0,1], those powers would have a diminishing contribution. Moreover, the bilinear abundances $\{b_{ij}\}$ do not even have a real physical meaning if the sum-to-unity for $\{a_i\}$ is maintained. This has been addressed by reformulating Eq. (1.2) such to express the $\{a_i\}$ and $\{b_{ij}\}$ in a weighted form [Nascimento and Bioucas-Dias, 2009].

$$\boldsymbol{x} = \sum_{i=1}^{p} a_i \boldsymbol{e}_i + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} b_{ij} \boldsymbol{e}_i \odot \boldsymbol{e}_j + \boldsymbol{n} \tag{1.2}$$

Another class of mixtures allowing a model-based treatment are the so-called intimate mixtures occurring on a more microscopic scale, e.g. between the mineral grains in rocks, sands and soils. Refer once more to Fig. 1.4. Here, the *Hapke model* from spectroscopy provides a standard equation relating the bidirectional reflectance of a particle in particulate medium to its *single scattering albedo* [Hapke, 1981, 2012]. The model has been adapted for use in Earth observation in many works; however, it is observed that many of its parameters remain dependent on the experiment [Dobigeon et al., 2014]. Another factor that limits the utility of this model is the extended spatial scale at which the mixtures are normally viewed in Earth observation. This has been mitigated by combining the Hapke model with the LMM in several ways [Broadwater and Banerjee, 2010; Close et al., 2012].

Aside from these modeled cases, situations are conceivable where nonlinear mixing is likely but perhaps a bit more elusive. One example of this are water-related applications, since mixing of water masses (bodies, constituents) is more 'intangible', but nonetheless makes the radiation transfer nonlinear [Mobley, 1994]. Here, data-driven approaches such as *manifold learning* were found to be useful, though computationally expensive [Bachmann et al., 2005, 2009]. The crux of manifold learning is to construct a mapping that preserves some global or local relationship from the (nonlinear) data manifold while projecting it to a lower-dimensional space (see also Fig. 1.7). In this new space, linear classification or unmixing algorithms can then be applied. This form of subspace identification has recently received a lot of interest in different fields of research [Ma et al., 2011]. A counterpart to this

type of processing are the *kernel methods* that map the input data to a much higher dimensional space, in which the data becomes linearly separable. The 'kernelization' of linear algorithms has been explored in both unmixing Broadwater and Banerjee [2009] and wider signal and pattern analysis [Müller et al., 2013].

We ought to mention that efforts have also been devoted to the logical 'middle route', which rather examines (usually via statistical testing) if a hyperspectral data is structured linearly or nonlinearly [Han and Goodenough, 2008; Altmann et al., 2013]. For more details about these and several other ways of dealing with nonlinear unmixing, interested reader is referred to the above mentioned surveys.

1.3.3 Related concepts

Many concepts deserve to be linked to spectral unmixing, as the authors researching this topic have made good use of the many applicable techniques and angles (see the aforementioned surveys and references therein). Our aim here is to clarify few concepts that we will use or refer to more often in the following chapters.

Dimensionality reduction

High dimensionality is inevitably induced in many types of data (multimedia, medical and remote-sensing data, etc.) recorded with high spectral-, spatial- or temporal resolution. The goal of dimensionality reduction (DR), ideally, is to trim back the data to its true or intrinsic dimension. This is defined as the smallest number of independent parameters that is needed to generate the data [Bennett, 1969].

Traditionally, DR has been realized through linear techniques, such as factor analysis, principal component analysis (PCA) or multidimensional scaling (MDS) [Maaten van der et al., 2008]. These techniques all assume that the data occupies a space where each point (vector) is a linear combinations of a number of basis vectors. This makes the intrinsic dimension equal to the smallest number of vectors in a basis that spans the space [Strang, 2009]. For example, Fig. 1.5 depicts a group of points residing on a 2-dimensional plane embedded in $\mathbb{R}^3$. The job of DR in this case would be to recognize any two vectors spanning the plane, thereby allowing each point to be described with those two instead of three components. Most linear DR techniques derive a basis for such reduced space (a subspace) by performing the *eigen-decomposition* of some positive-definite matrix derived from the data. For such a matrix $\boldsymbol{A}$, this entails finding the *eigenvalues* $\{\lambda_i\}$ and the corresponding *eigenvectors* $\{\boldsymbol{q}_i\}$ that satisfy: $\boldsymbol{A}\boldsymbol{q} = \lambda\boldsymbol{q}$. For some $n \times l$ matrix $\boldsymbol{X}$ holding zero-mean data (n data points with l components), the MDS derives eigenvalues $\{\lambda_i\}$ and eigenvectors $\{\boldsymbol{t}_i\}$ by decomposing the Gramian matrix: $\boldsymbol{G} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$. The PCA gets

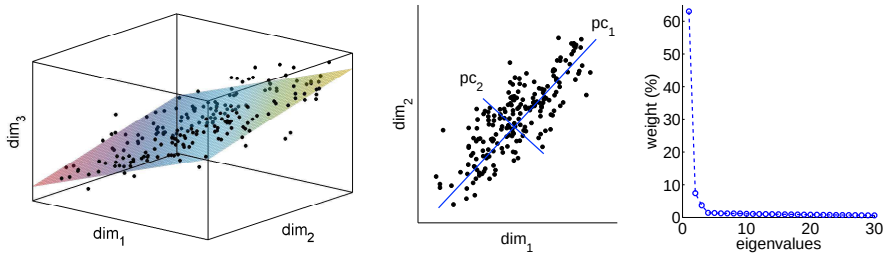

Figure 1.6: Illustration of PCA: Most of the data variability is described by few PCs.

its eigenvectors $\{\boldsymbol{u}_i\}$ from the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. Since $\sqrt{\lambda_i}\boldsymbol{t}_i = \boldsymbol{X}\boldsymbol{u}_i$ [Chatfield and Collins, 2000], the two techniques are essentially equal. Moreover, the PCA has this interpretation: Let $\{\boldsymbol{u}_i\}$ $(i = \overline{1,l})$ be sorted in descending order – then each $\boldsymbol{u}_{j+1}$ represents the direction accounting for as much of the variability left in the data relative to its projection onto a subspace spanned by $\{\boldsymbol{u}_k\}$ $(k = \overline{1,j})$. To demonstrate this, let $\boldsymbol{X}$ be a (zero-mean) data and $\boldsymbol{\Sigma} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ its covariance matrix, and let $\boldsymbol{u}$ be a *principal component*, i.e. vector derived by solving: $\boldsymbol{\Sigma}\boldsymbol{u} = \lambda\boldsymbol{u}$. The projection of $\boldsymbol{X}$ onto $\boldsymbol{u}$ (containing the weight of each data point in $\boldsymbol{X}$ along $\boldsymbol{u}$) is $\boldsymbol{X}\boldsymbol{u}$, making $\boldsymbol{\Sigma}_u = (\boldsymbol{X}\boldsymbol{u})^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{u}) \equiv \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u}$ the covariance of this projection. Moreover, we know that for any symmetric matrix (which $\boldsymbol{\Sigma}$ is), the corresponding eigenvectors $\{\boldsymbol{u}_i\}$ can be chosen orthonormal, i.e., such that $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = 0$, if $i \neq j$ and $\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j = 1$, if $i = j$ [Strang, 2009]. This allows defining a *Lagrangian* and deriving the condition for its maximization via partial derivatives, as:

$$f(\boldsymbol{u}) = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{u} - \lambda(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} - 1) \tag{1.3}$$

$$\frac{\partial f}{\partial \lambda} = 0 \Rightarrow \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1\,; \quad \frac{\partial f}{\partial \boldsymbol{u}} = 0 \Rightarrow \boldsymbol{\Sigma}\boldsymbol{u} = \lambda\boldsymbol{u} \tag{1.4}$$

The latter equation tells that $\boldsymbol{u}$ that maximizes f (and $\boldsymbol{\Sigma}_u$) is indeed the same $\boldsymbol{u}$ that is derived by the PCA. Figure 1.6 gives an example: A plot of a high-dimensional data in its selected three and two dimensions hints at its predominantly directional structure, while the comparison of eigenvalues confirms that most of the variation is contained within the initial 2-3 eigenvectors, i.e. principal components.

Outside this classical view, one reckons that data can also display more complex dependencies that cannot be preserved by projecting the data linearly. In such cases the task of DR is entrusted to nonlinear techniques. One of those techniques (see a taxonomy and survey by [Maaten van der et al., 2008]) is the popular *manifold*

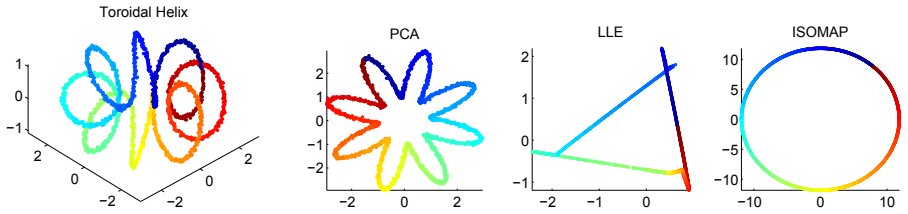

Figure 1.7: A nonlinear 2-dimensional manifold embedded in $\mathbb{R}^3$ (*left*) and its disentangling produced by one linear (PCA) and two nonlinear DR methods.

learning (ML), which assumes that the data lies on a nonlinear manifold embedded in the higher-dimensional space. For example, Fig. 1.7 shows a sampling from a helix curve in $\mathbb{R}^3$ perturbed with Gaussian noise [Coifman and Lafon, 2006]. Each data point is specified by its position on this curve (the value of t in its parametric equation $g(t) = (x(t), y(t), z(t))$) and the noise parameter, so the data is two-dimensional. Hence, the figure shows reduction of the data to two dimensions by applying the earlier explained PCA and two ML techniques: locally linear embedding (LLE) [Roweis and Saul, 2000] and ISOMAP [Tenenbaum et al., 2000]. We see that only ISOMAP was able to fully reveal the data structure (imagine untangling a rubber band).

The exact way these ML techniques operate is for now of no interest, as we will revisit some of them later. We know that each is designed to preserve specific local or global properties of the original manifold, which is clearly data dependent. Nevertheless, all ML techniques also share some common characteristics. For one, they do not provide a parametric mapping between the high- and the low-dimensional space. Unlike the linear mappings that are fully determined by their projection matrices (eigenvectors), the ML techniques must specify a mapping for each data point. Also, they all must optimize one or more free parameters, like e.g. the connectedness of a data point with its neighbors. These and other properties have several implications, of which one of the most noteworthy is the much increased computational and memory complexity. This has sometimes been mitigated by deriving the mapping from only a small subset of the data – the so-called landmark points [Silva de and Tenenbaum, 2002; Chen et al., 2006])

Classification

Another concept that is closely related to unmixing is pattern recognition or classification. Classification connotes assigning an input vector to one of a finite number of discrete categories or classes. Each such vector will ideally comprise measurements

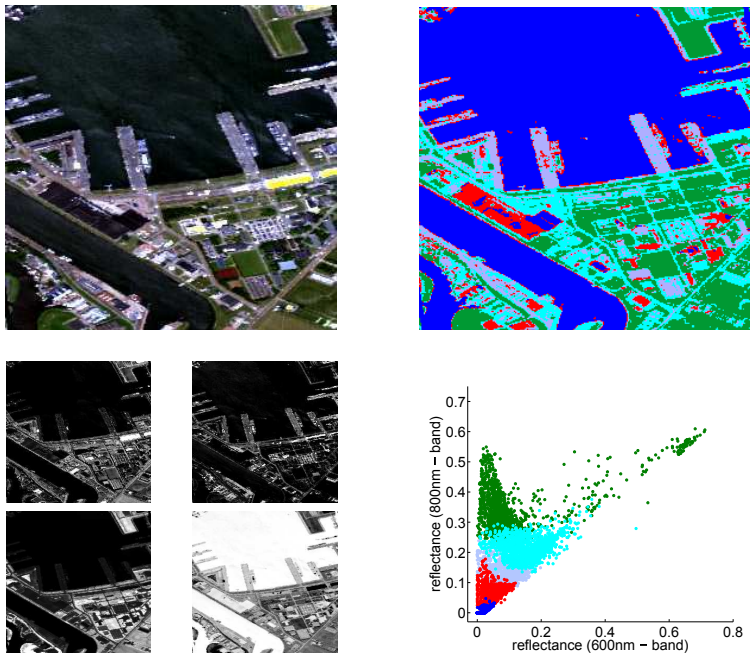

Figure 1.8: An RGB view of a hyperspectral image and the output from its clustering and unmixing, including a scatter plot discerning the pixels from each cluster.

of some distinguishing *features* that are very similar for objects from the same class, and very dissimilar for objects from different classes [Duda et al., 2001]. In the case of hyperspectral imagery, pixels are normally classified based on (dis)similarity of their spectra [Tso and Mather, 2009]. Still, the maturation of high-spatial-resolution optical sensing by several platforms (e.g. low-cost Unmanned Aerial Vehicles or UAVs) has prompted many works that extract additional spatial and contextual features from individual or grouped pixels (see a taxonomy of some of those works in [Camps-Valls et al., 2011]). Another relevant distinction concerns supervised and unsupervised classification. While the former uses training examples of input vectors along with their class labels, the second discovers groups of similar examples in the data, which is also referred to as *clustering* [Bishop, 2007].

With these definitions, it acknowledgeable that the similarity between unmixing and classification is both conceptual and methodological. This follows directly

from the translatability between classes and endmembers. Several works have even assumed a full equivalence, by readily assigning the classes by ranking the endmember abundances [Adams et al., 1995; Goodwin et al., 2005]. A more indirect approach has used the abundance comparison to divide the image into segments, and therefrom extract features for the classification [Luo and Chanussot, 2009; Dopido et al., 2011]. Another approach has utilized the abundances as a reference for the sub-pixel mapping, i.e. for predicting the spatial location of the classes within a coarser or mixed pixel (another reference being a classification obtained at a higher spatial resolution, typically from a different modality) [Villa et al., 2011; Mahmood et al., 2013]. Yet another example is the sub-pixel anomaly and target detection, which has been approached from both the classification and unmixing perspective (see [Manolakis et al., 2003; Bioucas-Dias et al., 2013] and references therein) and by combining those two [Glenn et al., 2013]. Also, similar strategies and ingredients have often been used in composing the classification- and unmixing algorithms. In Sec. 1.3.1 we already mentioned the coinciding treatments of the spatial and contextual priors and constraints. Other instances include the use of DR (which in classification is used to avoid the 'curse of dimensionality' [Duda et al., 2001]), the use of kernel methods, Bayesian estimation for dealing with the class- or endmember variability, etc.

To make some of the above said more tangible, consider an example of unsupervised unmixing and classification depicted in Fig. 1.8. The unmixing was done with 4 automatically extracted endmembers from the image, while the classification is essentially a clustering of the image into 5 clusters. The spatially sub-sampled abundance maps show a linear mapping of $\{a_i\}$ ($i = \overline{1,4}$) from Eq. (1.1) to interval $[0,255]$, so that brighter pixels indicate larger abundance. We see that the classes coincide pretty well with the endmembers (water, vegetation, etc.), but the scatter plot also hints that some pixels from the 'green' cluster reside on the outskirt of the data, thereby qualifying as a separate endmember.

1.4 Objectives

The work described in the thesis addresses spectral unmixing from two main objectives: advancement of unmixing methodology and introduction of spectral unmixing in new applications. The first objective, in particular, is concerned with the development of data-driven algorithms for linear and non-linear spectral unmixing that do not suffer from the dependency on physical parameters and models, and can mitigate high computational complexity induced by the popular use of optimization techniques. The second objective is concerned with formulating a

spectral-unmixing approach that can address practical problems and applications of remote sensing, in particular where unmixing is not necessarily a method of choice. The latter is to be done through theoretical and empirical exploration using progressively released new source data and associated results from available algorithms that we can use as our references.

In order to achieve these objectives, the following research questions are asked:

- Can we reformulate or extend the existing unmixing algorithms and approaches, in particular the geometric framework induced by the linear mixing model, to enable alternative ways of processing?
- Can we sensibly exploit, introduce or combine other data-processing techniques with unmixing algorithms to get superior performance in our context?
- Can we theoretically motivate the use of unmixing in studied applications and formulate it such that we can maintain or reach translatability of the unmixing result to the inquired (physical) quantities?

1.5 Outline of the thesis

In accordance with the dual objective of our work described in the previous section, the thesis has been divided into two parts bearing the same distinction in their focus.

PART I

The first part comprises two chapters describing algorithms that address particular aspects of spectral unmixing. While Chapter 2 is concerned with both the extraction of unknown endmembers and the computation of their abundances, Chapter 3 focuses on the latter task when the endmembers are fully known. In fact, an imperfection of the abundance determination in the first algorithm has partly inspired the development of the second. Common to both algorithms is that they realize a notable computational gain compared to their direct alternatives.

Chapter 2 describes a distinctive, data-driven approach to nonlinear unmixing. The distinctiveness resides in the integration aspect, as the approach adopts the classical linear, i.e. geometrical framework (see Sec. 1.3.1) but reformulates its elements using distance geometry, in particular incorporating a nonlinear distance measure inspired by manifold learning (see Sec. 1.3.3).

Chapter 3 presents a fresh approach and solution for computing the abundances, i.e. performing the inversion step, in a fully constrained linear unmixing. The

approach exploits the equivalence of solving the fully constrained least-squares problem and geometric projection of a point onto a simplex. It then introduces several concepts to accomplish the latter. The result is an analytical solution, which equally permits to be reformulated using (nonlinear) distance measures.

PART II

The second part comprises two independent chapters, each introducing or elaborating an unmixing approach in an atypical (for unmixing) application. Accordingly, each chapter follows a same structure whereby the use of particular mixing model(s) is motivated, possibly complemented with additional constructs, and compared with an available reference method.

N.B. Our use of particular methods as references has been prescribed by practical more than fundamental considerations, which has to do with the (in)accessibility of tools and algorithms and their results on real datasets that became progressively available to us during our work and writing. At the same time, all the methods we adopt as references have been documented and in cases compared against some of the credited alternatives in recent literature that is also cited in this thesis.

Chapter 4 examines the utility of linear and nonlinear unmixing for the detection of adjacency effect. This effect is caused by atmospheric scatter inducing path interference between the solar radiation reflected from different ground surfaces. Hence, what is investigated is the suitability of the unmixing techniques to separate the true signature of a pixel from the scatter imparted by its adjacent neighborhood.

Chapter 5 explores the use of unmixing for estimating the water quality in inland and coastal waters. The starting point here is a prevalent analytical model relating the reflectance of water to the inquired concentrations of its impurities or constituents. The goal, thus, is to invert this model by employing the constructs from spectral unmixing.

PART I

UNMIXING TECHNIQUES AND ALGORITHMS

CHAPTER 2

UNMIXING BY GEODESIC SIMPLEX-VOLUME MAXIMIZATION

Abstract

Spectral mixtures observed in hyperspectral images often display nonlinear mixing effects. Since most traditional unmixing methods are based upon the linear mixing model, they perform poorly in retrieving the correct endmembers and their abundances from nonlinear mixtures. This chapter presents an algorithm that is capable of computing both quantities under nonlinear mixing assumptions. The algorithm adopts the classical simplex-volume maximization, but reformulates it using distance geometry and in particular geodesic distances. Besides realizing a data-driven treatment of nonlinear mixtures, this approach offers notable computational advantage relative to a mere precedence of linear unmixing by nonlinear dimensionality reduction. The algorithm is evaluated through comparison with the reference linear method, on both synthetic data and real hyperspectral images acquired from distinct scenes.

N.B. Most of the content from this chapter coincides with the work described in [Heylen et al., 2011b]. The author of the thesis contributed to that work in relation to its implementation and experimental part. This also inspired a spin-off, streaming formulation of (linear) endmember extraction, whose details are omitted in this text but can be found in [Burazerović et al., 2011].

2.1 Introduction

Spectral unmixing is prevalently done under the assumption of linear mixing; however, one often encounters intimate mixtures or topographic multi-layering of materials that induce their non-negligible secondary and higher-order reflections. Unmixing in such cases has often been handled by explicit modeling of these effects [Halimi et al., 2011a; Somers et al., 2009], or by employing more data-driven methods for dealing with nonlinearity (e.g. kernel-based processing Broadwater and Banerjee [2009] and artificial neural networks Nascimento and Bioucas-Dias [2009]; Plaza et al. [2009]; Liu and Wu [2005]; Guilfoyle et al. [2001]). Refer also to Sec. 1.3.2.

Another methodology for treating nonlinearly mixed pixels consists of performing a nonlinear dimensionality reduction (NLDR), yielding a linear space or reduced dimension in which traditional (linear) methods can then be applied. Many NLDR techniques are data-driven and exploit some form of (unsupervised) manifold learning Gashler et al. [2008]. This connotes construction of a mapping that preserves some global or local relationship from the manifold constituted by the source data while projecting it to a lower-dimensional space [Maaten van der et al., 2008]. The subsequent linear operations may relate to any of the unmixing, classification or compression techniques. For example, several NLDR methods (Local Linear Embedding [Roweis and Saul, 2000], Laplacian Eigenmaps [Belkin and Niyogi, 2001] and Local Tangent Space Alignment [Zhang and Zha, 2004]) were compared as a precursor to a k-nearest neighbor classifier by [Wu et al., 2009]. The Isomap algorithm [Tenenbaum et al., 2000] was used as a preprocessing for classification [Bachmann et al., 2005; Yangchi et al., 2005], while its specific implementation using landmark points was recently also used for subsequent unmixing [Chi and Crawford, 2013].

An impacting disadvantage of most NLDR techniques is their high computational cost and memory requirements, making then rather impractical for use with sizable hyperspectral data. This problem was acknowledged in [Bachmann et al., 2006], where some strategies were proposed for realizing a scalable operation of Isomap. The scalability was there mostly achieved by aligning parallel executions of Isomap on image tiles, while streamlining the computation of geodesic distances and definition of local neighborhoods on the manifold. The idea of tile alignment was also followed in [He et al., 2009], where Isomap was replaced by a coordinate representation derived from diffusion maps. Other methods have resorted to explicit use of supervision by allowing the manifold structure to be used as input in off-line learning of a classification model or deriving a new distance metric for a classifier [Wang et al., 2006; Ma et al., 2010]. Despite these improvements, it can be fairly said that NLDR of large hyperspectral data has usually been considered for applications where ample nonlinear effects can be expected beforehand (e.g. in bathymetry

[Bachmann et al., 2009]).

In this work, we propose a different approach to spectral unmixing of nonlinearly mixed pixels. The central idea is to reformulate a prevalent methodology for linear unmixing inspired by the original N-findR [Winter, 1999], by expressing its steps in terms of distance geometry. Under certain curvature conditions, we then introduce geodesic distances, so that the unmixing is done taking into account the (nonlinear) structure of the data manifold. This approach enables dealing with nonlinearly mixed data, while at the same time cutting down the complexity of a two-step approach where some linear unmixing method is simply preceded by NLDR.

The chapter is organized as follows: Section 2.2 gives a more formal description linear unmixing and explains our methodology for introducing nonlinearity in all its stages. These stages are explained in detail in separate subsections, including a word about complexity. Section 2.3 describes the related experiments. Here, two subsections are devoted to describing a comparison with the reference linear approach, first using synthetic data, and then also real hyperspectral images. Section 2.4 closes with conclusions and discussion.

2.2 Methodology

We start from the prevalent *linear mixing model* (LMM) that we discussed earlier in Sec. 1.3.1. According to this model, the spectrum of a mixture is represented by a linear sum of the spectra of its constituents, which is put mathematically as:

$$\boldsymbol{x} = \sum_{i=1}^{p} a_i \boldsymbol{e}_i + \boldsymbol{n}, \quad \forall i : \, a_i \geq 0 \, , \sum_{i=1}^{p} a_i = 1 \tag{2.1}$$

where $\boldsymbol{x} = (r_1, \ldots, r_l)$ is an l-dimensional spectrum and $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{il})$ are the spectra of pure or intrinsic materials, the so-called *endmembers*, constituting the mixture. Furthermore, p denotes the known or estimated number of endmembers, a_i is the fractional representation or *abundance* of endmember $\boldsymbol{e}_i$ in the mixture, while $\boldsymbol{n}$ stands for additive modeling errors and noise. Considering the system of equations (2.1) in a normal case where $l > p$ and $\{\boldsymbol{e}_i\}$ are independent makes it solvable with ordinary least-squares; however, physically more meaningful solution is obtained by forcing the also indicated positivity- and sum-to-one constraints. In the context of hyperspectral imagery this is simply saying that only the whole pixel area and no negative sub-pixel contributions are taken into account. A consequence is that any $\boldsymbol{x}$ expressed by Eq. (2.1), when $\boldsymbol{n} = 0$, will now reside within a $(p-1)$–simplex of which $\{\boldsymbol{e}_i\}$ ($i = \overline{1, p}$) are the vertices or extreme points. Refer also to Fig. 2.1. This geometrical reading has inspired many methods and algorithms that

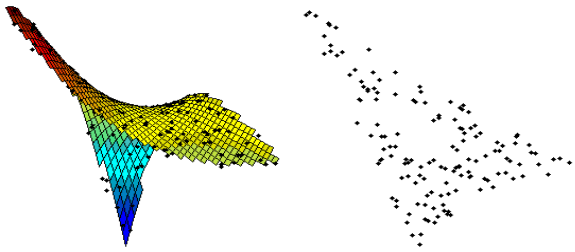

Figure 2.1: A data residing on a nonlinear manifold embedded in $\mathbb{R}^3$ (*left*) and its linear projection on $\mathbb{R}^2$ (*right*) yielding a data enclosed by a 3–simplex, i.e. a triangle.

effectively search for embedded or enclosing simplices in the data, including the popular N-findR [Winter, 1999], [Dowler and Andrews, 2011], pixel purity index (PPI) [Chaudhry, 2005], simplex growing algorithm [Chang et al., 2006], convex cone analysis [Ifarraguerri and Chang, 1999], vertex component analysis (VCA) [Nascimento and Bioucas-Dias, 2005b], etc. See also a survey from [Parente and Plaza, 2010] and other references mentioned in Sec. 1.3.1.

When the endmembers $\{\boldsymbol{e}_i\}$ in Eq. (2.1) are known, the inversion needed to determine the $\{a_i\}$ is typically attained through constrained least-squares optimization (see Sec. 3.1). This makes the search for $\{\boldsymbol{e}_i\}$ treatable as a separate problem, where different approaches are conceivable depending on the assumptions about decomposability or sparsity of those endmembers [Bioucas-Dias et al., 2012]. The assumption that each endmember $\boldsymbol{e}_i$ will designate a *pure pixel* representing a single surface material has been prevalent (see e.g. [Plaza et al., 2004] for a comparative study of 6 popular endmember-extraction algorithms where this assumption is made), but is often not valid in practice. Rather, pixels close to the true endmember in the spectral space will often be retrieved by the endmember extraction process.

A more impacting limitation of the LMM that is addressed by our work concerns those situations when the spectral mixing happens non-linearly. This means that every pixel spectrum in principle becomes a non-linear function of endmembers and abundances. Without fully knowing this function, we can make assumptions about continuity. First, a pixel with very large abundance for a given endmember $\boldsymbol{e}_i$, and almost zero abundances for all $\{\boldsymbol{e}_j\}, j \neq i$, will have a spectrum that lies close to $\boldsymbol{e}_i$ in the spectral space. Furthermore, when the abundances $\{a_i\}$ ($i = \overline{1,p}$) vary gradually from one set to another, we can assume that the corresponding pixel spectrum will vary smoothly from some initial state to a final one. If this were not

the case, one would observe discrete jumps in the observed spectra at certain sets of abundances, which does not seem physically plausible. One way to model such non-linearity is thus to assume a non-linear but continuous bijective mapping F between the linear space of abundance coefficients and the spectral space:

$$\boldsymbol{x}_k = F\left(\sum_{i=1}^{p} a_{ki}\boldsymbol{e}_i\right) \tag{2.2}$$

where k is the pixel index. This mapping induces a manifold in the spectral space composed of the continuous projection of a linear simplex. In practice, we consider this manifold to resemble a non-linearly transformed low-dimensional simplex embedded in the high-dimensional spectral space, such that the endmembers will still correspond to vertices of the data manifold. Figure 2.1 illustrates this for artificial data. In the sequel, we describe an algorithm that realizes data-driven endmember extraction and abundance estimation taking into account this manifold structure. The central idea is to express a linear unmixing method in terms of distances in the spectral space, so that its parameters can be computed using non-Euclidean distance measures.

2.2.1 Distance based formulations

Many endmember-extraction algorithms exploit the same notion of a minimum-volume enclosing simplex. A typical way of computing the volume of a simplex having as vertices $\{\boldsymbol{e}_i\}$ $(i = \overline{1,p})$ is to apply the formula:

$$V(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_p) = \frac{1}{(p-1)!}\cdot\left|\det\begin{bmatrix}\boldsymbol{e}_1^{\mathrm{T}} & 1\\ \vdots & \vdots\\ \boldsymbol{e}_p^{\mathrm{T}} & 1\end{bmatrix}\right| \tag{2.3}$$

where $|\cdot|$ takes absolute value of the matrix determinant. One way to understand this formula is by knowing that the volume of an n-dimensional cube is given by the absolute value of the determinant of $n \times n$ matrix whose column- or row vectors represent the cube's edges [Strang, 2009]. For example, imagine translating an arbitrary 2-simplex or triangle until one of its 3 vertices, say $\boldsymbol{e}_1$, coincides with the origin. This will make the area of the triangle expressible as one half of the

2-dimensional cube (parallelogram) with edges $\boldsymbol{e}_2-\boldsymbol{e}_1$ and $\boldsymbol{e}_3-\boldsymbol{e}_1$, implying:

$$\frac{1}{2}|V(\boldsymbol{e}_2-\boldsymbol{e}_1,\boldsymbol{e}_3-\boldsymbol{e}_1)| = \begin{vmatrix} x_2-x_1 & y_2-y_1 \\ x_3-x_1 & y_3-y_1 \end{vmatrix} = \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - \begin{vmatrix} x_2 & y_2 \\ x_1 & y_1 \end{vmatrix} - \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + \begin{vmatrix} x_1 & y_1 \\ x_1 & y_1 \end{vmatrix} = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = |V(\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3)| \quad (2.4)$$

where we have made use of some basic properties and the co-factor formula for determinants. The reasoning is easily extended to any $(p-1)$ dimensions, leading to Eq. (2.3). Because the matrix $[\boldsymbol{E}|\boldsymbol{1}]$ in (2.3) is typically non-square ($l > p$), dimensionality reduction (DR) techniques have commonly been used in the above mentioned algorithms to reduce the data to dimension $p-1$. Such DR is justified, because, if $\{\boldsymbol{e}_i\}$ $(i=\overline{1,p})$ are true endmembers, then all their mixtures will reside within the $(p-1)$-simplex even if it is embedded in a higher than $(p-1)$-dimensional space.

However, the volume V of a simplex also permits to be expressed via inter-vertex distances that do not depend on the dimensionality of any vertex. Let d_{ij} be the distance between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$; we can reformulate Eq. (2.3) by exploiting the Cayley-Menger determinant:

$$(-1)^p 2^{p-1}((p-1)!)^2 V^2 = \det(\boldsymbol{C}_{1,2,\ldots,p}) \quad (2.5)$$

$$\boldsymbol{C}_{1,2,\ldots,p} = \begin{bmatrix} 0 & \boldsymbol{1}^{\mathrm{T}} \\ \boldsymbol{1} & \boldsymbol{D}^2_{1,2,\ldots,p} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{C}_{1,2,\ldots,p-1} & \boldsymbol{d}_p \\ \boldsymbol{d}_p^{\mathrm{T}} & 0 \end{bmatrix} \quad (2.6)$$

where $\boldsymbol{D}^2_{1,2,\ldots,p} = \left[d^2_{ij}\right](i,j=\overline{1,p})$ is a $p\times p$ matrix, $\boldsymbol{d}^2_p = (1, d^2_{1p},\ldots,d^2_{(p-1)p})$ and $\boldsymbol{1}$ is a $p\times 1$ vector of ones. With these definitions proving the matrix identity from Eq. (2.6) is straightforward. Deriving Eq. (2.5) requires more steps, but a good sense can again be obtained by starting from Eq. (2.3) when $p=3$. The crux is write out $(1/4)\cdot V^2 = \det(\boldsymbol{E}|\boldsymbol{1})\cdot\det((\boldsymbol{E}|\boldsymbol{1})^{\mathrm{T}})$ and recognize the invariance of placing this product as co-factor of the first element in the larger 4×4 determinant, $|p_{ij}|$, where $p_{1n}=1$ for $n=1,2,3,4$ and $p_{m1}=0$ for $m=2,3,4$. This is then followed up by suitable scaling and addition of columns of this determinant to yield Eq. (2.5). More theory about Cayley-Menger determinants is found in [Blumenthal, 1970].

A worthwhile implication of the above formulas is also that we can apply the well-known *Schur complement* to the right side of Eq. (2.6), to derive:

$$\det(\boldsymbol{C}_{1,2,\ldots,p}) = -\left(\boldsymbol{d}_p^{\mathrm{T}}\boldsymbol{C}^{-1}_{1,2,\ldots,p-1}\boldsymbol{d}_p\right)\det(\boldsymbol{C}_{1,2,\ldots,p-1}) \quad (2.7)$$

Comparing this to Eq. 2.5, we see that the volume of a p–simplex can be expressed in terms of the volume of a $(p-1)$-simplex and the projective distance of a new point

$\boldsymbol{e}_p$ from the latter simplex. This property was exploited in [Burazerović et al., 2011] to extract the endmembers by estimating the largest simplex in a streaming fashion, i.e. by progressively evaluating one pixel at a time.

2.2.2 Introducing nonlinearity

The distances used in the previous simplex-volume calculations are all Euclidean distances. However, when replacing them by geodesic distances calculated on the data manifold, one can think of Eq. (2.5) as providing an estimate for a volume measured along the manifold itself. This interpretation is justified as long as the data manifold can be covered by Euclidean space. This is the case when the manifold is completely flat, i.e., has zero intrinsic or Riemannian curvature in each point. Intuitively, this corresponds to a situation where the data manifold is created by embedding a (subset of) low-dimensional Euclidean space into a higher dimensional space by folding or deforming this low-dimensional space, but without 'stretching'. In non-artificial data this assumption about continuity, convexity and local linearity need not always be met. For hyperspectral imagery, this is clearly something that will depend on the structure of the imaged scene. Generally, it seems reasonable to assume that data manifolds induced by natural scenes will permit only small non-zero curvatures and thus the said assumption to hold.

A well-known procedure for approximating geodesic distances on manifolds consists of constructing a k nearest-neighbor (kNN) graph from the data points and measuring the shortest-path distances along this graph [Tenenbaum et al., 2000]. The initial graph is constructed by taking the Euclidean distance between any two points $\boldsymbol{x}_m$ and $\boldsymbol{x}_n$, and connecting every point to its k nearest neighbors, with k fixed or varying according to specific radius ε. The nearest neighbor property is not reciprocal, meaning that $\boldsymbol{x}_m$ being closest to $\boldsymbol{x}_n$ does not guarantee that the opposite is also true. Hence, a symmetrization applied to the graph to make it undirected. This step can be formulated as: $(\boldsymbol{G}, \boldsymbol{W}) \longmapsto (\overline{\boldsymbol{G}}, \overline{\boldsymbol{W}})$, where $\boldsymbol{G}$ and $\boldsymbol{W}$ are the graph's incidence- and weight matrix. Specifically, $\boldsymbol{W} = [w_{mn}]$ with $w_{mn} \equiv d_{mn} = d(\boldsymbol{x}_m, \boldsymbol{x}_n)$ the distance between two points and $\boldsymbol{G} = [g_{mn}]$, where $g_{mn} = 1$ if $d_{mn} (m \neq n)$ satisfies the criterion and $g_{mn} = 0$ otherwise. The criterion in this case tests if d_{mn} is one of k smallest elements among $\{d_{m1}, \ldots, d_{mn}, \ldots d_{mN}\}$, or alternatively if $d_{mn} \leq \varepsilon$ for given m, where N is the number of data points. The matrices $\boldsymbol{W}$ and $\boldsymbol{G}$ defined in this way are square and the symmetrization is thus achieved by taking $\overline{\boldsymbol{G}} = [\overline{g}_{mn}], \overline{g}_{mn} = \max(g_{mn}, g_{nm})$ and $\overline{\boldsymbol{W}} = \boldsymbol{W} \circ \overline{\boldsymbol{G}}$, which stands for the Haddamard or entry-wise product. After this step, the well-known algorithms for calculating the shortest paths on a graph [Dijkstra, 1959; Floyd, 1962] can be used to find the shortest distance between any pair of points. The distances computed in this way

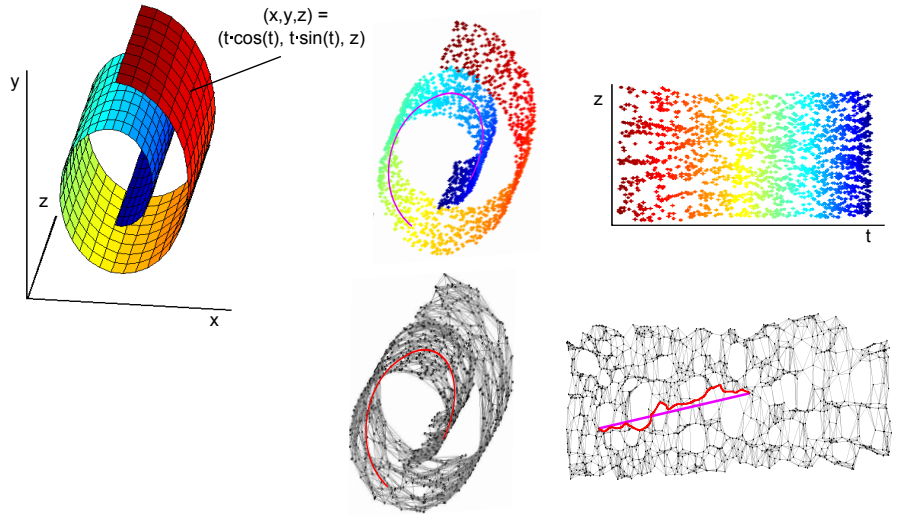

Figure 2.2: *Top*: "Swiss roll", its 2000-point sampling and embedding of the latter found by Isomap. *Bottom*: A kNN graph for the sampling, rolled up and in a plane, evincing the alikeness of the graph-based and geodesic distances in matching colors.

will approximate the true geodesic distances as would be measured along the surface of the data manifold. Figure 2.2 gives an illustration.

2.2.3 Endmember extraction

The algorithm for endmember extraction is inspired by the original N-findR but modifies it by involving the above explained concepts. To find the largest simplex along the data manifold, the following algorithm is proposed:

1. Construct a weighted, symmetrical and connected kNN graph on the data set.
2. Randomly select $\{\boldsymbol{x}_1,\ldots\boldsymbol{x}_p\}$ as the initial vertices. Calculate the shortest-path distances from each of these points to all other points using the Dijkstra's algorithm. Use Eq. (2.5) to calculate the simplex volume.
3. Pick a random point $\boldsymbol{x}$ and calculate the simplex volume when each element from the set $\{\boldsymbol{x}_1,\ldots\boldsymbol{x}_p\}$ is successively replaced by $\boldsymbol{x}$. If a larger simplex is found, keep the new vertex $\boldsymbol{x}$, calculate its distance to all other points (again with the Dijkstra's algorithm) and recalculate the matrix identities in (2.6).

4. Perform the previous step until no larger simplex is found for any point. The final vertices $\{\boldsymbol{e}_1, \ldots \boldsymbol{e}_p\}$, designating the endmembers, are now known.

At this point two parameters still need to be decided: the number of endmembers, p and the connectivity of the graph, which is governed by k. Choosing p is essentially equivalent to gauging the intrinsic dimensionality of the data (see in Sec. 2.2.1), which is known to be a highly non-trivial task. As previously noted, many seminal algorithms for endmember extraction used linear DR techniques, such as the principal component analysis (PCA) [Jolliffe, 2002] or minimum noise fraction (MNF) [Green et al., 1988], to reduce the dimensionality of the data. The problem of determining the number of endmembers then becomes the problem of setting an appropriate threshold for selecting p principal eigenvalues obtained from decomposing some positive-definite matrix derived from the data (see Sec. 1.3.3). Automatic techniques such as virtual dimensionality (VD) [Harsanyi et al., 1993; Chang and Du, 2004] and HySime [Bioucas-Dias and Nascimento, 2008] have been reported to produce more accurate estimates of the number of endmembers needed to span a given dataset. Still, most of these methods cannot be used straightforwardly with distance geometry.

One method that can be easily adapted to use distances is to exploit the following notion about the volumes: If a data lies on a $(p-1)$ dimensional hyperplane (or any p–simplex residing on it), then any volume computed in one dimension higher is zero. This entails that plotting the maximum simplex volumes for increasing values of p should display a sharp drop at some p, indicating the dimensionality of the data. This technique has several drawbacks. First, it is computationally intensive. Second, due to noise and numerical sensitivity of the distance matrix, the simplex volumes will practically be non-zero for any dimension, which obscures the expected discontinuity. For all these reasons, we go along with a common practice of choosing a suitable value for p empirically for each dataset. This is justified minding the objectives of our method and the reference algorithms with which it is to be compared, which use the same notion.

The second unknown parameter is the number of nearest neighbors, k. This constant should not be too small, or one ends up with disconnected clusters and large 'holes' in the graph. Artificially connecting such clusters usually destroys information about the manifold structure due to the creation of shortcuts. On the other hand, a value for k that is too large will smoothen possible fine structures on manifold, equally causing a loss of information. We observe that setting $k = 20$ usually yields good results and stability with datasets we used in practice (e.g. space- and airborne images of natural sites and landscapes).

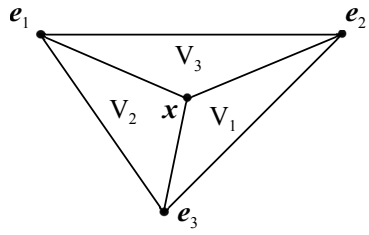

Figure 2.3: A point $\boldsymbol{x}$ lies inside a simplex S_3 spanned by $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$. The abundance coefficient a_i equals the volume ratio V_i / V, with $V = \sum_i V_i$.

2.2.4 Abundance estimation

Looking at Eq. (2.1), first without the constraints, it is clear that unique solution for a_i can only be found if $\boldsymbol{n} \neq 0$ and the vector $\boldsymbol{x}$ lies in the row space of a $p \times l$ matrix whose rows are (independent) $\boldsymbol{e}_i$, that is, if $\boldsymbol{x}$ is an exact combination of $\{\boldsymbol{e}_i\}$ ($i = \overline{1,p}$). This is normally not the case with every data point, so one tries to minimize the error between a data point $\boldsymbol{x}$ and some $\widehat{\boldsymbol{x}}$ reconstructed from $\boldsymbol{e}_i$ and a_i ($i = \overline{1,p}$). This leads to the ordinary least-squares formulation, or equivalently the projection onto the row space of the matrix whose rows are $\{\boldsymbol{e}_i\}$, yielding values for $\{a_i\}$ that are not necessarily positive or sum up to one. In many unmixing algorithms this limitation has been addressed by applying constrained least-squares techniques regarding one or both constraints from Eq. (2.1), depending on the implementation. If both constraints are taken into account, one finds a fully constrained least-squares solution (FCLSU) for the unmixing problem [Heinz and Chang, 2001].

However, in the non-linear case, every data point should be viewed in terms of its distances to the endmembers rather than its coordinates in the spectral space. Because we assumed that the data manifold is flat, the inter-point distances measured in a local neighborhood still obey all Euclidean geometric properties. Thus, to estimate the $\{a_i\}$, we require a procedure that uses distance geometry, and preferably maintains both unmixing constraints. One procedure that fulfills this can be formulated by exploiting the observation that a_i from Eq. (2.1) can be written as:

$$a_i = \frac{V(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{i-1}, \boldsymbol{x}, \boldsymbol{e}_{i+1}, \ldots, \boldsymbol{e}_p)}{V(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p)} \tag{2.8}$$

where the nominator designates the volume of a simplex obtained by replacing an i-th vertex of a simplex defined by $\{\boldsymbol{e}_i\}$ ($i = \overline{1,p}$) with $\boldsymbol{x}$. See Fig. 2.3. This property follows directly from the fact that $\{a_i\}$ play the role of homogeneous barycentric coordinates in the coordinate system of $\{\boldsymbol{e}_i\}$ and the equivalence between barycentric

and areal coordinates in simplices [Coxeter, 1963]. The utility of this property in the unmixing context was also observed in [Honeine and Richard, 2012]. The volumes $V(\cdot)$ in Eq. (2.8) can be calculated in terms of inter-vertex distances using Eq. (2.5).

2.2.5 Complexity

In several approaches dealing with nonlinearly structured hyperspectral data, one uses Isomap as a preprocessing step before applying some linear algorithm for classification or unmixing (see Sec. 2.1). Considering the latter use, if the geodesic distances in the original spectral space would exactly match the Euclidean distances in the projected lower-dimensional space, than this two-step approach would be virtually equivalent to our proposed algorithm.

One large difference, still, concerns the computational complexity: To calculate the low-dimensional embedding of some data using Isomap, the complete geodesic distance matrix is required. For N data points this entails N executions of the Dijkstra's algorithm (which is a $O(N\log(N))$ algorithm in itself) and a $O(N^2)$ memory requirement. In addition, one should also count in the last step of Isomap that is the actual projection onto a low-dimensional space via multi-dimensional scaling (MDS) [Tenenbaum et al., 2000], or equivalently the PCA (see Sec. 1.3.3). On the other hand, our algorithm only calculates the distances from the initial p endmember candidates to all other points, followed by another such calculation every time a larger simplex is found. Since the number of vertex updates is only a small fraction of N, the number of executions of the Dijkstra algorithm before terminating the algorithm is much smaller than N. Moreover, only the distances from the p endmembers to all other points need to be stored at any point, reducing the memory requirements by a factor N/p. These observations allow executing the proposed algorithm on large images without the need for introducing landmarks and divide-conquer-merge strategies.

2.3 Experiments

The experiments are structured in two parts. First, the algorithm is validated on synthetic data. Next, the same is done using two real hyperspectral images. In all cases the nonlinear algorithm is compared to a linear approach combining the original N-findR (extracting the endmembers) and FCLSU (computing the abundances).

2.3.1 Synthetic data

We define an artificial dataset by taking random samples from a linear two-dimensional simplex, transformed into a three-dimensional nonlinear manifold by

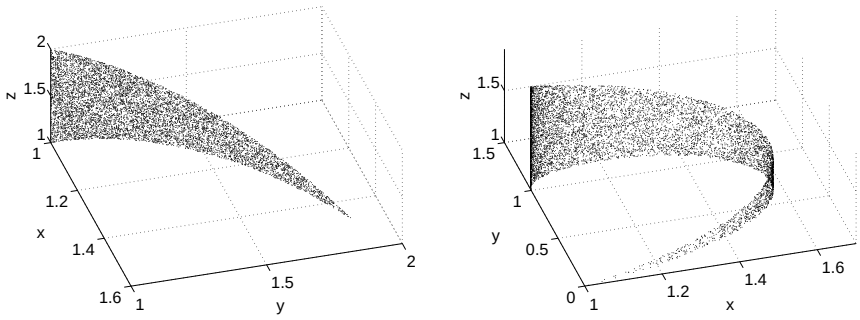

Figure 2.4: The artificial data set for $\sigma = 0.5$ (*left*) and $\sigma = \pi$ (*right*) for 10^4 randomly generated abundances $a_{i1,2}$.

projection onto a bended surface. Equations (2.10)–(2.11) describe this, where a_{ij} denote the abundances and σ is a parameter controlling the bending. Figure 2.4 shows an example of this data when a_{ij} are chosen randomly for two different σ.

$$x_i = a_{i1}\sin(\sigma a_{i1}) + 1 \tag{2.9}$$
$$y_i = a_{i1}\cos(\sigma a_{i1}) + 1 \tag{2.10}$$
$$z_i = a_{i2} + 1 \tag{2.11}$$

The endmembers can be determined analytically from these equations by setting a single abundance a_{ij} equal to one and all others equal to zero, which yields the following expressions:

$$\boldsymbol{e}_1 = (\sin(\sigma)+1, \cos(\sigma)+1, 1) \tag{2.12}$$
$$\boldsymbol{e}_2 = (1,1,2) \tag{2.13}$$
$$\boldsymbol{e}_3 = (1,1,1) \tag{2.14}$$

We unmix this dataset with two algorithms: The proposed nonlinear algorithm, and the reference linear approach using N-findR to extract the endmembers and the FCLSU (see in Sec. 2.2.4) to determine the abundances. This combination will henceforward be simply referred to as "N-findR".

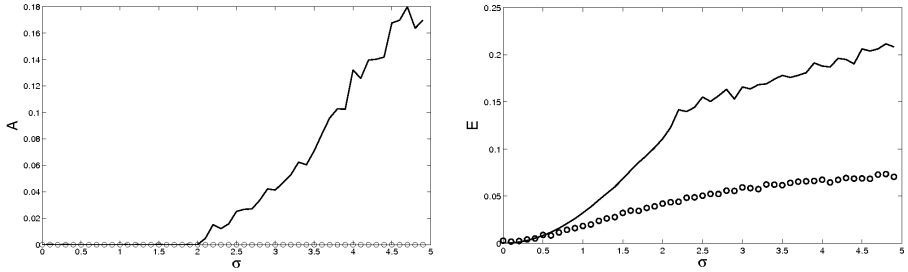

Figure 2.5: The averaged minimum spectral angle (*left*) and absolute error on the abundances (*right*) as functions of σ. The solid lines show the results of N-findR and the circles correspond to the nonlinear algorithm, using $k = 20$.

In the first experiment, we search for 3 endmembers ($p = 3$) and for each found endmember $\widehat{\boldsymbol{e}}_i$ we calculate the minimum of the *spectral angles* with the true endmembers from Eq. (2.12) – (2.13). This quantity is then averaged over all $\widehat{\boldsymbol{e}}_i$ to yield a measure for the deviation of the extracted endmembers from the true ones. Equation (2.15) formally expresses this. This measure is plotted as a function of σ on the left-hand side in Fig. 2.5. One can see that the non-linear algorithm retrieves all endmembers exactly for the entire range of σ. The N-findR fails to do this once σ becomes larger than 2, and the errors quickly increases afterwards.

$$A = \frac{1}{p}\sum_{i=1}^{p} \min_j \left(\arccos\left(\frac{\widehat{\boldsymbol{e}}_i \cdot \boldsymbol{e}_j}{\|\widehat{\boldsymbol{e}}_i\| \|\boldsymbol{e}_j\|} \right) \right) \tag{2.15}$$

As a second experiment, we use the retrieved endmembers to compute the abundances for all N data points, using FCLSU for the linear case and simplex-volume ratios for the nonlinear case. We then compare the resulting abundance maps $\{\widehat{\boldsymbol{a}}_i\}$ to the true abundance maps $\{\boldsymbol{a}_i\}$ using the error measure formulated in Eq. (2.16). This measure is plotted on the right-hand side of Fig. 2.5, again as a function of σ. For low σ, the data manifold does not deviate much from a linear simplex and both algorithms perform well. For higher σ, the abundance maps obtained with either method start deviating. The non-linear algorithm however still performs clearly better. For $\sigma > 2$, the linear algorithm fails to correctly retrieve all endmembers and this further exacerbates the deviation of its abundance maps.

$$E = \frac{1}{p}\sum_{i=1}^{p} \min_j \left(\frac{1}{N}\sum_{k=1}^{N} |\widehat{a}_{ik} - a_{jk}| \right) \tag{2.16}$$

We have rerun the experiments by varying k in the range $3-20$. For $k>5$, this variation did not have much impact on either the endmember search or the abundance estimation. The lower values did worsen the performance, as the connectivity of the kNN graph was apparently too low to approximate the true geodesic distances.

2.3.2 Real images

The experiments from Sec. 2.3.1 demonstrated that the proposed algorithm outperforms linear unmixing in case of nonlinearly mixed data, but the used data gives only an idealized representation of what can be encountered in practice. In this section, we conduct similar experiments using real hyperspectral images. One of these images is the well-known AVIRIS Cuprite dataset of a mining region in Nevada, USA (see e.g. [Kruse et al., 2003] for a detailed description), which has been widely used in studies of spectral unmixing. The second example is a proprietary reflectance image of a heathland area in Belgium.

Cuprite dataset

From the full AVIRIS Cuprite image we take a $350\times400\times50$ excerpt, with the 50 spectral bands in the range $1.99-2.48$ μm that is the most relevant for discerning mineral signatures [Clark, 1999; Parente and Zymnis, 2005]. A similar dataset was used in the original N-findR [Winter, 1999]. This dataset is also available in the ENVI software under the name "cup95eff" and is illustrated in Fig. (2.6) on the left. A ground photo from the Cuprite region shown on the right side of the figure gives some more idea about the terrain in the vicinity of the site.

As a first experiment, we extract endmembers with N-findR and the nonlinear algorithm, and compare those to the mineral spectra from the freely available USGS "splib04c" spectral library. For every extracted endmember we assign a unique match by minimizing the *spectral angle*. Both algorithms were run with $p=8,\dots16$. The most abundant minerals in the scene [Winter, 1999; Parente and Zymnis, 2005], such as *alunite*, *kaolinite* and *calcite*, were properly identified for all values of p by both algorithms. The more scarce minerals, such as *buddingtonite*, *muscovite* and *montmorrilonite*, where detected for some p but not all. No specific value of p lead to clearly optimal performance with either of the two algorithms.

Figure 2.7 depicts 3 out of 14 endmembers extracted with N-findR. The others are not shown for brevity, but include two *calcite* endmembers, *butlerite*, *gaylussite*, *augite*, *pyrite*, *ammoniummillsmec*, *lepidolite*, *dumotierite* and a shade endmember, although the spectral match with several of the latter is not so good due to noise present in some endmember pixels. Several of these minerals are known to be

Figure 2.6: Enhanced false-color image of the AVIRIS Cuprite dataset (R=2.34, G=2.20, B=2.10μm) and photo of "Kaolinite hill" in Cuprite, Nevada – one of the premier calibration sites for imaging spectrometers used in mineral mapping (source: *www.usgs.gov*). AVIRIS is flown at an altitude of 20km, resulting in approximately 20-m pixels ([Kruse et al., 2003]).

present in the scene and similar results can also be found in [Winter, 1999]. A much resembling picture to the one shown in Fig. 2.7 was obtained with the nonlinear algorithm using $k = 20$ and $p = 16$. By this we mean that the deviation of the spectral angle between each of these nonlinear results and the corresponding reference plot in Fig. 2.7 was less than 3 degrees. Some of the other endmembers found with good spectral match were *calcite*, *diaspore*, *lepidolite*, *vermiculite*, *rutile* and *andradite*.

Next, we calculated the abundance maps using FCLSU and the simplex-volume ratios for the geodesic approach. Figure 2.8 gives an illustrative and somewhat typical comparison for one of the extracted endmembers. The linear abundance map can be compared to those found in e.g. [Winter, 1999; Parente and Zymnis, 2005] and shows exact agreement. The nonlinear map however deviates by having more bright pixels and some salt-and-pepper noise. This deviation could be ascribed to genuine nonlinear mixing effects, but it could also be due to the violation of the sum-to-one constraint by the nonlinear unmixing, which produces non-zero abundances where they should ideally be zero. The lack of ground reference information restricts us from formulating a more quantitative comparison.

In conclusion, both linear and non-linear unmixing of the Cuprite dataset yield comparable endmembers, many of which are known to be present in the scene. The

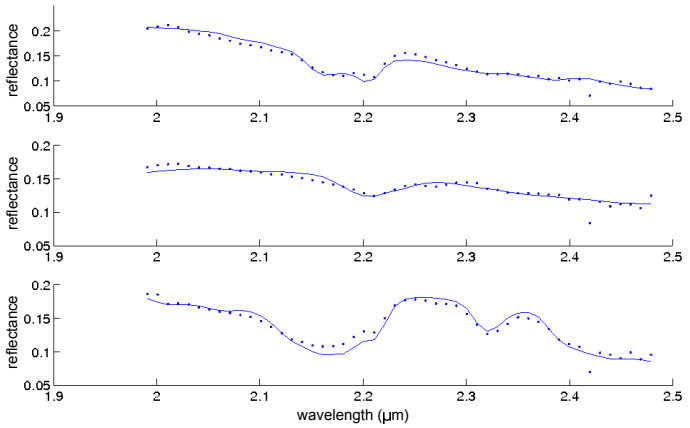

Figure 2.7: Three (out of $p = 14$) extracted endmembers with N-findR (*dots*) and the best matching library spectra (*solid lines*), where the quality of a match was measured as reciprocal of the spectral angle (see Eq. (4.9))). From top downwards: *kaolinite*, *montmorrilonite* and *alunite*.

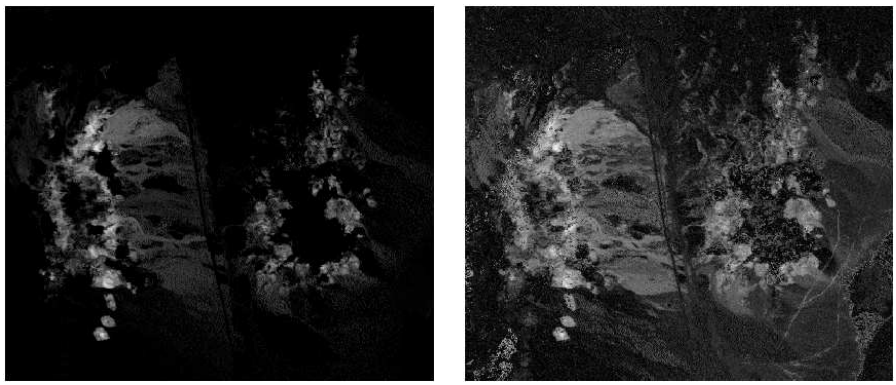

Figure 2.8: From left to right: abundance maps for *alunite* obtained from FCLSU and the nonlinear unmixing algorithm. The abundances have been mapped to the intensity value range $[0,255]$ by ordinary linear scaling.

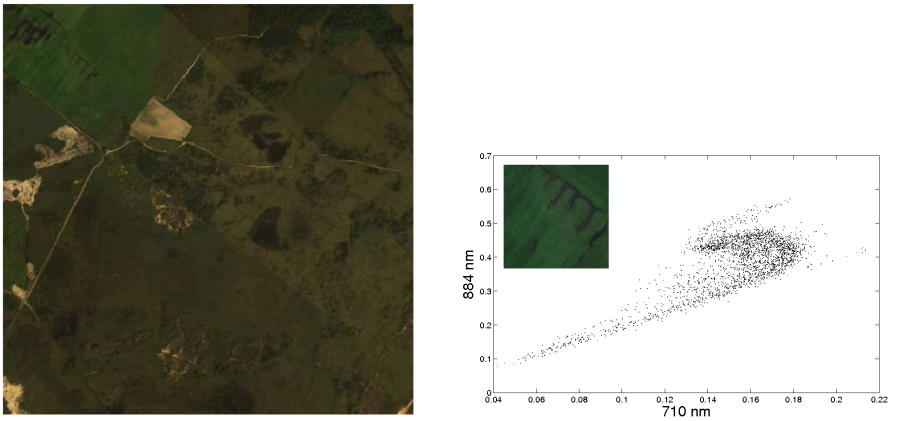

Figure 2.9: *Left*: RGB view of the ROI. *Right*: Scatter plot of band 10 (710 nm) and band 16 (884 nm) of a part of this ROI showing partly submerged grassland.

abundance maps obtained with the linear unmixing are similar to those reported in the literature. The abundance maps obtained from nonlinear unmixing have distinct features, but we cannot properly quantify them due to the lack of ground truth. It is conceivable that the Cuprite data set can be unmixed well linearly, and that the nonlinear unmixing algorithm will not bring a worthwhile improvement.

Heathland dataset

The second dataset available to us is a reflectance image acquired over a heathland area in Belgium called "Kalmthoutse heide". The original spectral information is given in 63 bands covering the range 0.456 - 2.55 μm, of which we exclude the noisy ones to retain 52 usable bands. From the original 4509 × 4359 image we select a more manageable region of interest (ROI) counting 500 × 500 pixels. This ROI is plotted in Fig. 2.9, where one may recognize partly submerged grassland, forest, heathland, arable land, sand dunes and water bodies. The scatter plot shown in the same figure belongs to a zoomed part from the ROI image and hints at the presence of nonlinearity in the data.

There is a ground reference for this dataset: about 2000 pixels had been classified by field measurements, resulting in 24 classes. From this ground truth, classification maps were estimated for the entire image [Thoonen et al., 2010]. These maps can be helpful in identifying some of the endmembers and their abundance maps, but

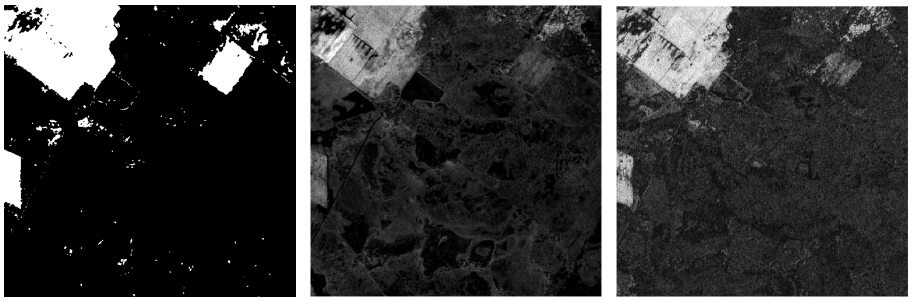

Figure 2.10: Classification map for the "grassland" class (*left*) and abundance maps obtained from the linear- (*center*) and nonlinear (*right*) unmixing.

exact correspondence between endmembers and classes cannot be expected. For this reason, we do not attempt to identify endmembers by examining their spectra. Instead, we search for the endmembers in an unsupervised manner (without identification), generate their abundance maps by unmixing, and study these. We set the number of endmembers empirically to $p = 16$ and use this setting with both the linear and the nonlinear unmixing algorithm.

A typical result is shown in Fig. 2.10: Out of the abundance maps generated for all 16 endmembers using the two algorithms, we have selected two maps that are visually similar but also have differences. Both abundance maps show respectable agreement with the classification map. The map produced by the nonlinear unmixing algorithm agrees the most, but once more contains salt-and-pepper noise due to the violation of the sum-to-one constraint in some pixels. Especially in the upper-left corner, where mixing of water and grassland is present, one notices that the nonlinear unmixing produced consistently larger abundance values (brighter pixels) relative to those obtained from the linear unmixing. This is in accordance with the suspected presence of nonlinearly mixed pixels in this regions (see the scatter plot in Fig. 2.9).

2.4 Conclusion

We have described and evaluated a distinctive data-driven and unsupervised approach that is capable of unmixing sizable hyperspectral data under the assumption of nonlinear mixing. The approach adopts a common methodology based on

simplex-volume maximization, but reformulates its steps using distance geometry, in particular incorporating a graph-based approximation of the geodesic distance.

Both aspects of the algorithm – endmember extraction and abundance estimation have been tested using synthetic data and real hyperspectral images. By examining the spectra and visualizing the abundances, we have witnessed both compatibility and deviation from the linear reference where these should be theoretically expected. A more rigorous assessment would require a dataset with a ground truth at the pixel level, preferably also discerning the type of mixing (linear or nonlinear), but such dataset was not available to us at the time of writing.

Besides realizing an extension beyond the prevalent linear unmixing paradigm, the proposed approach offers practical benefits relative to its direct nonlinear alternatives. By merging linear and nonlinear methodology, instead of cascading them, our approach notably reduces the computational complexity and makes execution on large data feasible with regular hardware. Similar or larger computational gain is also observed relative to methods that use explicit (bilinear) model fitting. The downside of this integration is that it makes the unmixing equation implicit, thereby disallowing reconstruction of the pixel spectra and enforcement of sum-to-unity on the abundances. Still, similar limitations are inherent to other methods. One example is the 'pre-image problem' in kernel methods [Kwok and Tsang, 2004], which has do with the reverse mapping from feature space back to input space, or in this case the reconstruction of a pixel spectrum. Another example is the ambiguity of interpreting the abundances associated with products of endmembers in bilinear mixing equations (see Sec. 1.3.1). Clearly, the hindrance implied by these limitations will depend on the intended qualitative or more quantitative use for the abundances, as in e.g. visualization, classification or sub-pixel mapping.

Since recently, there has been a growing interest among researchers to apply nonlinear techniques for unmixing of hyperspectral imagery. This was recognized in an extensive survey of unmixing algorithms from [Bioucas-Dias et al., 2012] and is also exemplified by the recent appearance of surveys that discuss exclusively nonlinear unmixing techniques [Dobigeon et al., 2014; Heylen et al., 2014]. Comparing our algorithm against all these alternatives would be impractical and pretentious, but perhaps few words should be added when considering this renewed context. Conceptually, our algorithm is perhaps best compared with a two-step approach that lets linear unmixing be preceded by nonlinear dimensionality reduction (DR). Both the unmixing algorithms and DR techniques are many and have their strengths and weaknesses, which leaves many worthwhile combinations. Yet, inspired by the 'No Free Lunch Theorem' from pattern classification [Duda et al., 2001], we can safely say that no combination will be optimal outside a specific context, i.e. dataset or application, that matches its assumptions. The same qualification could be made

for any other data-driven approach. The model-based approaches, on the other hand, can be more adequate when specific type of nonlinear mixing is studied or expected, but they are usually less flexible and prone to over-fitting.

Summarizing, we can say that the described algorithm offers a rather distinctive and practical compromise between linear and nonlinear unmixing. In the absence of nonlinearity, the algorithm is expected to perform similar as its linear basis, and in the presence of nonlinearity to enrich the linear results in a meaningful way, all without adding objectionable computational complexity.

CHAPTER 3

FULLY CONSTRAINED UNMIXING BY SIMPLEX PROJECTION

Abstract

A distinctive algorithm and approach to fully-constrained linear spectral unmixing with known endmembers is presented. The algorithm is based upon the equivalence of solving the fully constrained least-squares problem and geometric projection of a point onto a simplex. First, several geometrical properties of simplices are introduced, and then combined to yield a recursive algorithm realizing the simplex projection. An implementation for arbitrarily sized datasets is provided, and the algorithm is benchmarked against the conventional method employing constrained least-squares optimization. This is done on both artificial data and a real hyperspectral image. Unlike the preexisting solutions, the presented algorithm does not employ any optimization steps and is fully analytical, which enables it to complete the inversion task of unmixing while severely cutting down the computational load.

N.B. Most of the content from this chapter coincides with the work published in [Heylen et al., 2011a]. The contribution of the author of this thesis to that work is reserved to implementation and elaboration of some building blocks. A worthwhile update is provided in this chapter via inclusion of supplemental and clarifying passages, illustrations and examples.

3.1 Introduction

The algorithms for spectral unmixing perform at least one of these two tasks: estimating the unknown spectra of endmembers, and determining their respective fractions or abundances in each pixel. Most algorithms do this under the assumption of fully-constrained linear mixing, for which a renewed motivation was given in a recent survey of unmixing methodology by [Bioucas-Dias et al., 2012]. In our study covered by this chapter, we are in fact concerned with the second mentioned task, which is the estimation of abundances. Our main goal is to find an analytical solution for this problem.

We give a chronological overview of algorithms that preceded ours, starting off with what we explained in Sec. 2.2.4. Due to the presence of noise in the data and possible mis-estimation of endmembers, the LMM need not have an exact solution for each pixel. This problem was originally addressed via the ordinary least-squares, i.e. linear regression by [Chang et al., 1998], but this could not guarantee abundances that are positive or sum up to one. Later it proved easy to incorporate the sum-to-unity constraint, leaving to be solved a non-negatively constrained least-squares (LS) problem [Heinz and Chang, 2001]. Another take was formulated in [Chang et al., 1998] making use of orthogonal subspace projection. In [Shimabukuro and Smith, 1991], a weighted least-squares method was described and it was observed that the fully-constrained unmixing problem would require quadratic programming. None of these techniques was able to find the optimal solution.

Then, Heinz and Chang [2001] turned the fully-constrained unmixing into a non-negatively constrained least-squares problem by introducing a Lagrangian, after which the solution was found using the non-negatively constrained LS techniques (NNLS) described in Lawson and Hanson [1974]. This was the first algorithm to find the LS solution that obeys both the positivity and sum-to-one constraint on the abundances, which became known as the fully-constrained least-squares unmixing (FCLSU). Soon after, other methods followed that found the solution via other routes. For example, [Bajorski, 2004] noted the equivalence between the FCLSU and simplex projection and used a quadratic programming to solve the latter in the context of spectral unmixing. In [Velez-Reyes and Rosario, 2004], the authors transformed the least-squares into a least-distance formulation and subsequently solved the latter via the aforementioned NNLS techniques. In [Heinz and Chang, 2001; Chang et al., 2004], a combined method was proposed to simultaneously perform the FCLSU and endmember extraction by minimizing reconstruction errors. More recently, [Dobigeon et al., 2008, 2009] described a hierarchical Bayesian model that similarly realizes a 'joint' solution. Another take described by [S. and Qian, 2009] treats the FCLSU as a blind source separation problem, using non-negative matrix

factorization to find the endmembers and abundances. (We should say that blind unmixing of hyperspectral data via basic independent component analysis (ICA) – which seeks to linearly decompose data into statistically independent sources, was largely refuted in [Nascimento and Bioucas-Dias, 2005a] due to the fact the abundances of endmembers are not independent but cohere via the mixing constraints.) Yet another distinctive approach by [Miao et al., 2007] uses a maximum-entropy principle to improve the LMM inversion when encountering strong noise or mutual proximity of endmembers.

Most approaches that provide the exact FCLSU solution use an optimization or maximization procedure, gradient or steepest-descent methods, or require a random sampler from some distribution. For example, the Lawson-Hanson NNLS algorithm and quadratic programming are both optimization techniques that require appropriate initial vectors and step sizes. The Bayesian techniques depend on random samplers. Although the results obtained with these algorithms can be made arbitrarily accurate by modifying the algorithmic parameters, they do not provide an exact analytical solution.

In our approach, we take the linear algebra view of the FCLSU problem. The central idea is to exploit the equivalence between the FCLSU and simplex projection. This equality was also observed in Bajorski [2004]; Gillis et al. [2002], but to our knowledge all preexisting algorithms for simplex projection either employ quadratic programming Michelot [1986]; Bajorski [2004] or project only on canonical simplices Tuenter [2001]. Hence, we introduce some properties inspired by convex geometry in lower dimensions and extend these to any p–simplex. This yields a recursive algorithm that reduces the dimensionality of the problem by one with each step, until a proper abundance vector is found. The algorithm achieves this by successively performing orthogonal projections onto sub-simplices in lower dimensions. Such operation makes the algorithm fully analytical and gives it a high computationally efficiency. What's more, this analytical nature of the algorithm should in principle allow a reformulation of its steps in terms of (nonlinear) distance geometry, making it a proper alternative for the nonlinear abundance computation from Sec. 2.2.4.

The remainder of this chapter is structured as follows: In Sec. 3.2, we explain the equivalence between the FCLSU and the simplex projection. Section 3.3 describes the reasoning that led to the algorithm, the required mathematical concepts and the algorithm's steps. Separate subsections are devoted to implementation for large data, complexity and exception handling. Section 3.4 verifies the algorithm via a comparison with the established FCLSU. This is done, again in separate subsections, on synthetic data and a real hyperspectral image. Finally, Section 3.5 closes with conclusions and remarks.

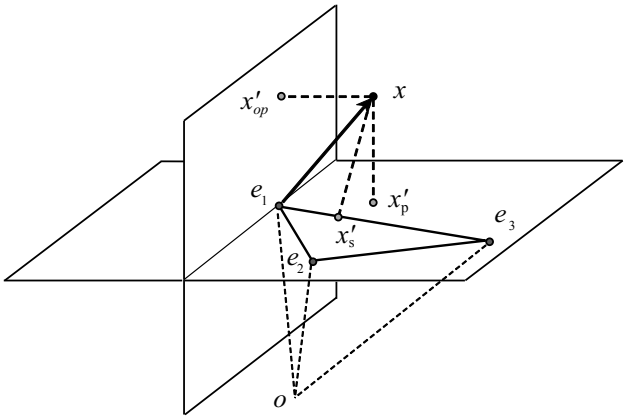

Figure 3.1: Geometry of linear unmixing for $p = 3$. Without the unmixing constraints, a point $\boldsymbol{x}$ is orthogonally projected onto the simplex plane, yielding $\boldsymbol{x}'_p$. With the constraints, the projection is onto the simplex (triangle) defined by $\{\boldsymbol{e}_i\}$, yielding $\boldsymbol{x}'_s$.

3.2 Methodology

Once again our starting point is the fully-constrained LMM, which we now also write in its equivalent matrix form:

$$\boldsymbol{x} = \sum_{i=1}^{p} a_i \boldsymbol{e}_i + \boldsymbol{n} \Leftrightarrow \boldsymbol{E}\boldsymbol{a} = \widehat{\boldsymbol{x}}, \quad \forall i:\ a_i \geq 0\,,\ \sum_{i=1}^{p} a_i = 1 \tag{3.1}$$

where $\boldsymbol{x}$ is a $d \times 1$ vector (spectrum of a pixel), $\boldsymbol{a}$ is a $p \times 1$ vector of *abundances* and $\boldsymbol{E}$ is a $d \times p$ matrix containing *endmembers* $\{\boldsymbol{e}_i\}$ as its columns, i.e. $\boldsymbol{E} = [\boldsymbol{e}_1\ \boldsymbol{e}_2\ \ldots\ \boldsymbol{e}_p]$. Furthermore, $\boldsymbol{n}$ stands for additive model errors and noise. From now on, we take $\widehat{\boldsymbol{x}} = \boldsymbol{x}$ and tacitly understand that any $\boldsymbol{x}$ may not be exactly reconstructible from given $\{\boldsymbol{e}_i\}$. The unmixing, or the inversion of the LMM, is about solving this problem: Given some $\{\boldsymbol{e}_i\}$, find $\{\boldsymbol{a}_{ij}\}$ in each pixel j so that Eq. (3.1) holds. We start by leaving out the constraints from Eq. (3.1) and then gradually explain the fully-constrained situation as an equivalent problem of projecting a point onto a simplex.

Consider the system $\boldsymbol{E}\boldsymbol{a} = \boldsymbol{x}$ for given $\boldsymbol{E}$ and $\boldsymbol{x}$. In the hyperspectral context it is typical that $d > p$ and that $\{\boldsymbol{e}_i\}$ are independent, which translates to a 'full-column-rank case' of solving a system of linear equations [Strang, 2009]. Accordingly, $\boldsymbol{E}\boldsymbol{a} = \boldsymbol{x}$ can have exactly one solution $\boldsymbol{a}$, if and only if $\boldsymbol{x} \in C(\boldsymbol{E})$, where $C(\boldsymbol{E})$ is the *column space* of matrix $\boldsymbol{E}$. This is another way of saying that $\boldsymbol{x}$ must be an exact combination

of the columns of $\boldsymbol{E}$, $\{\boldsymbol{e}_i\}(i=\overline{1,p})$. Since this is not met for any $\boldsymbol{x}$, such $\boldsymbol{x}$ needs to be projected onto the $C(\boldsymbol{E})$, yielding a projected point $\boldsymbol{x}' \in C(\boldsymbol{E})$ that does become a combination of $\{\boldsymbol{e}_i\}(i=\overline{1,p})$. The best choice is clearly to perform orthogonal projection that minimizes the $\mathscr{L}_2$-distance from $\boldsymbol{x}$ to $\boldsymbol{x}'$, which is described as:

$$\boldsymbol{x}' = \boldsymbol{E}\widehat{\boldsymbol{a}}, \quad \widehat{\boldsymbol{a}} = (\boldsymbol{E}^{\mathrm{T}}\boldsymbol{E})^{-1}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{x} \tag{3.2}$$

$$\widehat{\boldsymbol{a}} = \underset{\{\widehat{\boldsymbol{a}}\}_k}{\operatorname{argmin}}\Delta, \quad \Delta = \left\|\boldsymbol{x}' - \boldsymbol{x}\right\|_2 \tag{3.3}$$

where Δ is the distance, or the *reconstruction error*. Thus, the solution to the linear unmixing without constraints is given by the projection matrix $\boldsymbol{P}_{\mathrm{LSU}} = \boldsymbol{E}(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{E})^{-1}\boldsymbol{E}^{\mathrm{T}}$ and we can write $\boldsymbol{x}' = \boldsymbol{P}_{\mathrm{LSU}}\boldsymbol{x}$, where LSU stands for *least-squares unmixing*. It is interesting that the complement of LSU dubbed *orthogonal subspace projection* (OSP), which is realized via the matrix $\boldsymbol{P}_{\perp\mathrm{LSU}} = \boldsymbol{I} - \boldsymbol{P}_{\mathrm{LSU}}$, has also been explored in studies of unmixing and related problems (e.g. target detection) [Chang, 2005].

For the following explanation we refer to the sketch from Fig. 3.1. As a start, consider the plane of the 3–simplex, which is the plane going through all its defining points, $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$. A *vector-space* equivalent for the plane is obtained if the plane will also contain the origin ($\boldsymbol{o}$). For example, having the origin at $\boldsymbol{e}_1$ makes $\{\boldsymbol{e}_2 - \boldsymbol{e}_1, \boldsymbol{e}_3 - \boldsymbol{e}_1\}$ a *basis*, as all vectors in the plane become expressible as combinations of those two independent vectors. This by definition turns the plane into the column space $C(\widehat{\boldsymbol{E}})$ of matrix $\widehat{\boldsymbol{E}} = [(\boldsymbol{e}_2 - \boldsymbol{e}_1)\ (\boldsymbol{e}_3 - \boldsymbol{e}_1)]$. From this and what we have explained above, we see that the solution to $\widehat{\boldsymbol{E}}\boldsymbol{a} = \boldsymbol{x}$ is essentially the projection of $\boldsymbol{x}$ onto the simplex plane. In Fig. 3.1 this projection is depicted as $\boldsymbol{x}'_p$ (and $\boldsymbol{x}'_{op}$ is its orthogonal complement). By analogy with Eq. (3.2), the projection onto the plane of an arbitrary p–simplex is given by:

$$\boldsymbol{x}' = \widehat{\boldsymbol{E}}(\widehat{\boldsymbol{E}}^{\mathrm{T}}\widehat{\boldsymbol{E}})^{-1}\widehat{\boldsymbol{E}}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{e}_1) + \boldsymbol{e}_1 \tag{3.4}$$

$$\widehat{\boldsymbol{E}} = [(\boldsymbol{e}_2 - \boldsymbol{e}_1)\ (\boldsymbol{e}_3 - \boldsymbol{e}_1)\ \ldots\ (\boldsymbol{e}_p - \boldsymbol{e}_1)] \tag{3.5}$$

where the addition of $\boldsymbol{e}_1$ accounts for the plane translation, i.e. making $\boldsymbol{e}_1$ the origin.

With this interpretation, it is now clear that placing extra constraints on $\boldsymbol{a}$ in $\boldsymbol{E}\boldsymbol{a} = \boldsymbol{x}$ means restricting the projection of $\boldsymbol{x}$ (i.e. $\boldsymbol{x}'$) to some portion of the simplex plane. If both constraints from Eq. (3.1) are met, this portion will be the convex hull of $\{\boldsymbol{e}_i\}(i=\overline{1,p})$, or equivalently the p–simplex defined by those points. This follows directly from the fact that the two constraints by definition render the $\{\boldsymbol{a}_i\}$ as normalized *barycentric coordinates* with respect to the $\boldsymbol{e}_i$ (see Sec. 1.3.1, yielding a simplex lying on a $(p-1)$-dimensional plane in the d–dimensional space (since $\boldsymbol{e}_i \in \mathbb{R}^d$). The sketch from Fig. 3.1 shows an example of a 2-dimensional simplex in '$\mathbb{R}^3$', where $\boldsymbol{x}'_{\mathrm{s}}$ is the projection of $\boldsymbol{x}$ onto the simplex.

If we define a p–simplex, S as:

$$\boldsymbol{x} \in S \iff \exists\, a_1, \ldots, a_p \in [0,1] : \begin{cases} \boldsymbol{x} = \sum_{i=1}^{p} a_i \boldsymbol{e}_i\,, & \boldsymbol{e}_i \in \mathbb{R}^d \\ \sum_{i=1}^{p} a_i = 1 \end{cases} \tag{3.6}$$

then we can introduce an operator P projecting a point $\boldsymbol{x}$ onto the closest point $\boldsymbol{x}'$ on this simplex, as:

$$\boldsymbol{x}' \in S : \boldsymbol{x}' = P(\boldsymbol{x}) \iff \forall\, \boldsymbol{y} \in S : \|\boldsymbol{x} - \boldsymbol{y}\|_2 \geq \|\boldsymbol{x} - \boldsymbol{x}'\|_2 \tag{3.7}$$

The operator P is idempotent ($\forall \boldsymbol{x} : P(P(\boldsymbol{x})) = P(\boldsymbol{x})$), so it is a projection. By its construction, the projected point $\boldsymbol{x}'$ belongs to S and by solving $\boldsymbol{E}\widehat{\boldsymbol{a}} = \boldsymbol{x}'$, which is now exactly solvable since $\boldsymbol{x}' \in C(\widehat{\boldsymbol{E}})$, we can find the abundances $\widehat{\boldsymbol{a}}$. This shows the equivalence between approaching the FCLSU problem via optimization (which is about minimizing the reconstruction error) and by projecting a point onto a simplex. In the sequel, we shall refer to the latter as *simplex projection unmixing* (SPU).

3.3 Simplex projection algorithm

The SPU algorithm for projecting a point onto a simplex and finding the corresponding abundances rests on several observations. We first list these observations, together with a proof, which will clarify the reasoning that led to the algorithm. This will include some exception handling. We then describe the algorithm's steps, devoting separate sections to implementation for large datasets and complexity.

3.3.1 Prerequisites

Our first observation is inspired by Fig. 3.1 and states that one can arrive from $\boldsymbol{x}$ to $\boldsymbol{x}'_s$ by first going from $\boldsymbol{x}$ to $\boldsymbol{x}'_p$ and then from $\boldsymbol{x}'_p$ to $\boldsymbol{x}'_s$. This effectively means that we can consider a $(p-1)$–dimensional simplex plane, instead of the full d-dimensional spectral space, by orthogonally projecting all data onto this plane. The projected points (vectors) however retain d components (unless we choose to reduce them to $p-1$ components via the PCA or some other DR technique).

Lemma 1. *Orthogonal projection onto the hyperplane supported by the endmembers leaves the simplex projection invariant.*

Proof. Consider a p–simplex S defined by endmembers $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p\}$. Let $\boldsymbol{y}$ be the orthogonal projection of a point $\boldsymbol{x}$ onto the plane going through the endmember points $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p\}$. Choose any $\boldsymbol{v} \in S$. We then have that $\|\boldsymbol{x} - \boldsymbol{v}\|_2^2 = \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \|\boldsymbol{y} - \boldsymbol{v}\|_2^2$

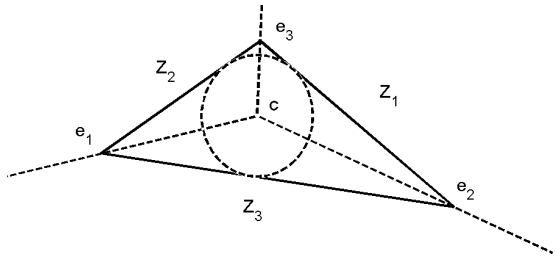

Figure 3.2: An arbitrary 2-dim. simplex, with the incenter and the three bisective cones indicated. A point in a cone Z_i will have a simplex projection with $\widehat{a}_i = 0$.

(*Pythagoras*), because $\boldsymbol{x}-\boldsymbol{y}$ and $\boldsymbol{y}-\boldsymbol{v}$ are orthogonal. Since $\|\boldsymbol{x}-\boldsymbol{y}\|_2$ is constant, the $\boldsymbol{v}$ that minimizes $\|\boldsymbol{x}-\boldsymbol{v}\|_2$ will also minimize $\|\boldsymbol{y}-\boldsymbol{v}\|_2$. Hence, $P(\boldsymbol{x}) = P(\boldsymbol{y})$ for the simplex projection operator P. □

The next observation, again referring to Fig. 3.1, is that the minimum-distance projection of $\boldsymbol{x}'_p$ onto a simplex renders a point $\boldsymbol{x}'_s$ on some face of this simplex (rather than its interior). This point must therefore have at least one abundance coefficient zero.

Lemma 2. *Consider a p–simplex S determined by endmembers* $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_p\}$*, and a point* $\boldsymbol{x}$ *that lies in the plane of this simplex, yet such that* $\boldsymbol{x} \notin S$*. The projection* $\boldsymbol{x}' = P(\boldsymbol{x})$ *lies on some face of S and therefore,* $\exists A_s \subset \{\widehat{a}_1,\ldots\widehat{a}_p\} : \widehat{a}_i = 0, \forall\, \widehat{a}_i \in A_s$.

Proof. The convex hull of a subset of size q from the p defining points of S, with $q < p$, is a q–face of S and a simplex itself: $S_q \subset S$. If $\boldsymbol{x}' \in S_q$, then $\boldsymbol{x}' \in S_p$ and by definition $\sum_{i=1}^{q} \widehat{a}_i = 1$ and $\sum_{i=1}^{p} \widehat{a}_i = 1$. This implies $\exists I = \{i_1,\ldots i_{p-q}\} \subset \{1,2,\ldots p\}$: $\widehat{a}_i = 0, \forall\, i \in I$. Now suppose that $\widehat{a}_i \neq 0, \forall\, i \in \{1,2,\ldots p\}$. Then $\boldsymbol{x} \in \text{int}(S)$. The line connecting $\boldsymbol{x}'$ with $\boldsymbol{x} \notin S$ will intersect the surface of S in a point $\boldsymbol{y}$ between $\boldsymbol{x}'$ and $\boldsymbol{x}$. This means that $\|\boldsymbol{x}-\boldsymbol{y}\|_2 < \|\boldsymbol{x}-\boldsymbol{x}'\|_2$, proving that $\boldsymbol{x}'$ cannot be the minimum-distance (orthogonal) projection of $\boldsymbol{x}$ onto S. □

The aim is next to identify the abundance coefficient that has to be zero. For this, we first introduce the *incenter* of a simplex.

Definition 1. *The incenter* $\boldsymbol{c}$ *of a simplex S spanned by* $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_p\}$ *is the intersection of all* $(p-2)$*-dimensional planes that bisect the dihedral angles between the faces of S. This is also the center of the largest hyper-sphere one can inscribe in S. See Fig. 3.2.*

The coordinates of the incenter $\boldsymbol{c}$ are found via the $(p-2)$-dimensional volumes of the p faces of the simplex. Let V_i be the volume of the sub-simplex omitting $\boldsymbol{e}_i$, i.e. spanned by $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_{i-1},\boldsymbol{e}_{i+1},\ldots,\boldsymbol{e}_p\}$. We then have (see in Lin [2008]) that:

$$a_i^c = \frac{V_i}{\sum_{i=1}^{p} V_i}, \quad \boldsymbol{c} = \boldsymbol{E}\boldsymbol{a}^c \tag{3.8}$$

where the vector $\boldsymbol{a}^c$ contains the barycentric coordinates of the incenter and $\boldsymbol{c}$ its Euclidean coordinates. Once more, $\boldsymbol{E}$ is the matrix whose columns are $\{\boldsymbol{e}_i\}$.

Consider a point $\boldsymbol{x}$ lying in the simplex plane. The ray starting from the incenter $\boldsymbol{c}$ and going through $\boldsymbol{x}$ will intersect some face of the simplex. We define the set of all points giving intersections with a given face as the *bisective cone* of that face.

Definition 2. *The bisective cone Z_i of a p–simplex S spanned by $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_p\}$, ($\boldsymbol{e}_i \in \mathbb{R}^{p-1}, \forall\, i = 1,\ldots p$), is the set of points defined by:*

$$\boldsymbol{x} \in Z_i \Leftrightarrow \exists b_1,\ldots,b_p \geq 0 : \begin{cases} \boldsymbol{x} = \boldsymbol{c} + \sum_{j=1}^{p} b_j(\boldsymbol{e}_j - \boldsymbol{c}) \\ b_i = 0 \end{cases} \tag{3.9}$$

with $\boldsymbol{c}$ the incenter of S.

In Fig. 3.2 the concept of bisective cones is illustrated for $p = 3$. It is clear that these cones form a partitioning of the $(p-1)$-dimensional space spanned by the endmembers, i.e. the column space $C(\boldsymbol{E})$ of the endmember matrix $\boldsymbol{E}$:

$$\bigcup_i Z_i = \mathbb{R}^{p-1} \tag{3.10}$$

These bisective cones allow us to determine the abundance coefficient that has to be zero. To do this, we need the following theorem, which is a special case of the more general conjecture that follows.

Theorem 3. *Consider the canonical p–simplex, $S_c = \{(v_1,\ldots v_{p+1}) : \sum_i v_i = 1\,, \forall\, i : v_i \geq 0\}$, and a point $\boldsymbol{x}$ in the simplex plane, such that $\boldsymbol{x} \notin S$. The projection of $\boldsymbol{x}$ on S_c is $\boldsymbol{x}' = P(\boldsymbol{x})$ with barycentric coordinates (abundances): $\widehat{a}_1,\ldots,\widehat{a}_p$. We then have:*

$$\boldsymbol{x} \in Z_i \Rightarrow \widehat{a}_i = 0; \tag{3.11}$$

Proof. To start with, observe that barycentric- and Euclidean coordinates are identical in case of canonical simplex. To prove the theorem, we use the following lemma proven in Shwartz et al. [2006] and also used in Duchi et al. [2008]:

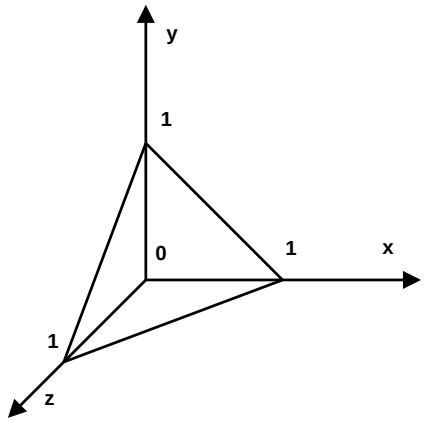

Figure 3.3: The canonical, standard or unit 3–simplex spanned by the three end-members $\{(1,0,0),(0,1,0),(0,0,1)\}$, effectively corresponding to the 2-dim. triangle.

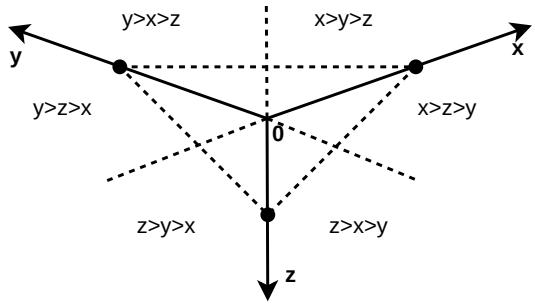

Figure 3.4: Two-dimensional plane of the canonical 3–simplex S_c, where the incenter coincides with the origin. The six options for the coordinates of a point $\boldsymbol{x}$ in this plane, and such that $\boldsymbol{x} \notin S_c$, are indicated.

Consider the canonical p–simplex, S_c. Let $\boldsymbol{y} \in S_c$ be the vector (point) that minimizes $\|\boldsymbol{y} - \boldsymbol{x}\|_2$ for a given $\boldsymbol{x}$, i.e. $\boldsymbol{y} = P(\boldsymbol{x})$. Let i and j be two coordinates of $\boldsymbol{x}$ such that $x_i > x_j$. We then have that $y_i = 0 \Rightarrow y_j = 0$.

When projecting onto a simplex, we know (see Lemma 1) that any point $\boldsymbol{x}$ can be replaced with its orthogonal projection on the simplex plane without affecting the projection onto the simplex. So we can restrict ourselves to points $\{\boldsymbol{x}\}$ in the simplex plane. We also know that, if a point $\boldsymbol{x}$ is in the simplex plane but outside the simplex S_c itself, at least one its barycentric coordinate or abundance coefficient has to be zero (see Lemma 2). A corollary of above lemma is then that the smallest x_i certainly corresponds to a zero barycentric coordinate, i.e.

$$\boldsymbol{x} \notin S_c,\ x_i = \min(x_1, \ldots, x_p) \Rightarrow y_i = 0 \tag{3.12}$$

See Fig. 3.3 and 3.4 for an illustration. Now the bisective cones Z_i correspond exactly to those subsets where a given coordinate x_i becomes minimal. A consequence of this corollary is that, at least for the canonical simplex, every point $\boldsymbol{x} \notin S_c$, $\boldsymbol{x} \in Z_i$, will have its corresponding barycentric coordinate $\widehat{a}_i = 0$. □

Next, we postulate that the above explained construct for determining abundance coefficients that have to be zero also works for general simplices. We base this supposition on extensive numerical testing.

Conjecture 4. *Consider a p–simplex S defined by endmembers $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p\}$, and a point $\boldsymbol{x} \notin S$. The projection of $\boldsymbol{x}$ is $\boldsymbol{x}' = P(\boldsymbol{x})$, with barycentric coordinates $\widehat{a}_1, \ldots, \widehat{a}_p$ (abundances). We then have:*

$$\boldsymbol{x} \in Z_i \Rightarrow \widehat{a}_i = 0; \tag{3.13}$$

Since this property allows us to determine which abundance coefficient has to be zero, we can construct a recursive algorithm: Suppose that $\boldsymbol{x} \notin S$, but $\boldsymbol{x} \in C(\widehat{\boldsymbol{E}})$ (see Eq. (3.5)). We know that orthogonally projecting $\boldsymbol{x}$ onto the S will yield a projection point $\boldsymbol{x}'$ residing on some face of S, i.e. the sub-simplex S_k defined by $\{\boldsymbol{e}_i\} \backslash \boldsymbol{e}_k$, where $\boldsymbol{e}_k$ is a vertex of S opposite to this face. Thus, projecting onto S_k instead of S does not change the abundance coefficients for those endmembers that are defining S_k. Orthogonal projection onto the plane of S_k reduces the implicit dimensionality of the problem by one, so we can continue recursively until we find a projected point inside the simplex, or we end up with a trivial one-dimensional problem. In the sequel we detail this algorithm, but first we give a word about some special cases.

3.3.2 Exception handling

We ought to recognize that Conjecture 4 is not always true, as counterexamples are findable for $p \geq 3$. The refutation examples all concern simplices that are highly skewed: their volume is very low compared to the smallest inter-vertex distance, and some dihedral angles are very obtuse. The configuration sketched in Fig. 3.5 hints at such a situation. The problem here is that the incenter will come close to $\boldsymbol{e}_1$ and $\boldsymbol{e}_4$, causing the point $\boldsymbol{x}$ to end up inside a wrong bisective cone and be projected onto $\boldsymbol{e}_2$ instead of $\boldsymbol{e}_3$ – the closest point to $\boldsymbol{x}$ on the simplex. We ought to remark that in our experience with real hyperspectral data such configurations have not manifested themselves often. One should also realize that an excessive mutual proximity of endmembers is likely to connote their mis-estimation, while the remoteness of data points from the simplex is proportional to noise.

Surely, to safeguard against any unfavorable situation, it is conceivable to run some simple checks. A first logical check can be to compare the mutual distances of the endmembers. In case of a suspected oblique simplex, one may choose to apply other methods (e.g. FCLSU) or use the SPU and reassert its result. The latter can be done efficiently via the Kolmogorov criterion that examines if some point is the closest-point projection [Bauschke and Borwein, 1993]:

$$\boldsymbol{y} = P_C(\boldsymbol{x}) \Leftrightarrow \forall \boldsymbol{z} \in C : (\boldsymbol{z} - \boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{y}) \leq 0 \tag{3.14}$$

where $P_C(\boldsymbol{x})$ is the projection of $\boldsymbol{x}$ onto a convex set C. Because a simplex S is a polyhedral set spanned by the endmembers, we only have to check whether the criterion is true for each endmember, i.e. for $\boldsymbol{z} \in \{\boldsymbol{e}_i\}, (i = \overline{1, p})$. Notice that inequality (3.14) follows directly from the cosine formula and a demand that the angle between $\overline{xy}$ and $\overline{yz}$ on the right-hand side of Fig. 3.5 is not smaller than 90°.

3.3.3 The algorithm

We start by describing the algorithm for projecting a single data point. Later, we extend this to arbitrary datasets. Suppose one wants to project a point $\boldsymbol{x} \in \mathbb{R}^d$ onto a p–simplex S_I defined by $I = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_p\}$ and find the corresponding abundances. One can do so by following these steps:

1. Orthogonally project $\boldsymbol{x}$ onto the $(p-1)$–dimensional plane containing I, yielding a projection point $\boldsymbol{y}$.

2. Check if $\boldsymbol{y} \in S_I$. If yes, go to Step 4, else go to Step 3.

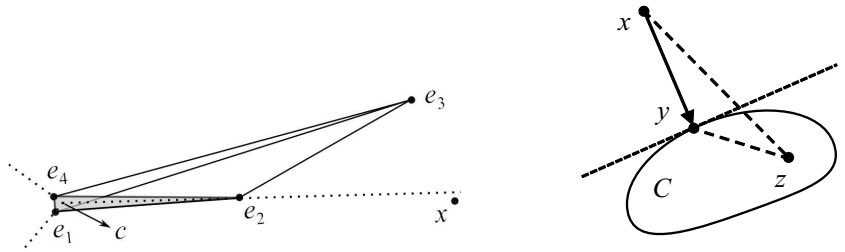

Figure 3.5: *Left*: A problematic configuration for the SPU ($\boldsymbol{e}_1$ and $\boldsymbol{e}_4$ must be very close). *Right*: Closest-point projection ($\boldsymbol{y}$) of a point ($\boldsymbol{x}$) onto the convex set C.

3. Determine the cone $Z_i : \boldsymbol{y} \in Z_i$. Remove $\boldsymbol{e}_i$ from I, yielding $I' = I \backslash \{\boldsymbol{e}_i\}$ and a new sub-simplex $S_{I'}$ of one dimension less. Set $\widehat{a}_i = 0$. Orthogonally project $\boldsymbol{y}$ onto the plane spanned by I'. Set $I = I'$ and go to Step 2.

4. Determine the abundance coefficients by writing $\boldsymbol{y}$ as a linear combination of the endmembers in I. This is now an exactly solvable system (since $\boldsymbol{y}$ is in the plane of S_I) respecting the constraints on the abundances.

The first step of the algorithm is to find the orthogonal projection $\boldsymbol{y}$ of a point $\boldsymbol{x}$ on the plane of S_I. This is done via Eq. (3.4). The same equation also allows to determine whether a point lies inside a simplex, which is needed for Step 2, as well as to find the abundances of an interior point, which is needed in Step 4. From this equation, we first define the partial abundance coefficients $\boldsymbol{v}$, as:

$$\boldsymbol{v} = (\widehat{\boldsymbol{E}}^{\mathrm{T}}\widehat{\boldsymbol{E}})^{-1}\widehat{\boldsymbol{E}}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{e}_1) \tag{3.15}$$

If all v_i, $i = 1, \ldots, p-1$ are positive and $\sum_i v_i \leq 1$, the orthogonal projection $\boldsymbol{y}$ of the point $\boldsymbol{x}$ lies in the simplex, and the corresponding abundance coefficients, $\widehat{\boldsymbol{a}}$ are given by concatenating $\widehat{a}_1 = 1 - \sum_i v_i$ and the vector $\boldsymbol{v}$:

$$\widehat{\boldsymbol{a}} = [1 - \sum_i v_i, \boldsymbol{v}] \tag{3.16}$$

The third step requires two main operations: Determining the cone Z_i that contains $\boldsymbol{x}$ ($Z_i : \boldsymbol{x} \in Z_i$), and projecting a point onto a simplex plane. The latter is once more accomplished via Eq. (3.4). To determine the cone, one must first calculate the incenter $\boldsymbol{c}$ via Eq. (3.8). Next, one must also compute the volumes of the p

sub-simplices, where the vertices are vectors in $\mathbb{R}^l$ (vectors with l components). One method to do this independently of l is to use the same Cayley-Menger determinant that we introduced in Sec. 2.2.1. In fact, the same equations (2.5) and (2.6) can be used, with d_{ij} now being the Euclidean distance between points $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$.

After finding the incenter $\boldsymbol{c}$, one can identify the cone $Z_i : \boldsymbol{x} \in Z_i$. If we define:

$$\boldsymbol{x}^c \equiv \boldsymbol{x} - \boldsymbol{c} \tag{3.17}$$

$$\boldsymbol{e}_i^c \equiv \boldsymbol{e}_i - \boldsymbol{c} \tag{3.18}$$

$$\boldsymbol{E}_i^c \equiv [\boldsymbol{e}_1^c, \ldots, \boldsymbol{e}_{i-1}^c, \boldsymbol{e}_{i+1}^c, \ldots, \boldsymbol{e}_p^c] \tag{3.19}$$

then the linear system of equations:

$$\boldsymbol{x}^c = \boldsymbol{E}_i^c \boldsymbol{b}^i \tag{3.20}$$

will have a solution for every index $i \in \{1, \ldots, p\}$. Because of the definition of bisective cone (3.9), we find that:

$$\boldsymbol{x} \in Z_i \Leftrightarrow \forall\, j : b_j^i \geq 0 \tag{3.21}$$

This observation enable the following algorithm for identifying the cone $Z_i : \boldsymbol{x} \in Z_i$,

1. Calculate the incenter $\boldsymbol{c}$ and the entities (3.17)-(3.19).
2. Solve Eq. (3.20) for every i, yielding a set of vectors $\{\boldsymbol{b}^i\} (i = \overline{1, p})$.
3. The set $\boldsymbol{b}^i$ with only positive entries identifies the cone that contains $\boldsymbol{x}$.

To solve Eq. (3.20) for $\boldsymbol{b}^i$, it suffices to consider a $(p-1) \times (p-1)$ linearly independent sub-matrix of $\boldsymbol{E}_i^c$ instead of the full $l \times (p-1)$ matrix, and use matrix inversion.

3.3.4 Implementation for large data sets

So far we have described the SPU algorithm dealing with a single point. In practice however, one normally needs to project a large number of data points corresponding to image pixels. Since several steps of the algorithm involve operations that take only the endmembers, a large performance boost can be gained by applying these operations to all points simultaneously. For example, the simplex-plane projection according to Eq. (3.4) can be done for an entire $d \times N$ data matrix, etc.

Consider a dataset $\{\boldsymbol{x}_n \in \mathbb{R}^d, n = \overline{1, N}\}$ and a set of endmembers $\{\boldsymbol{e}_i \in \mathbb{R}^d, i = \overline{1, p}\}$. The goal is to project all $\{\boldsymbol{x}_n\}$ onto the simplex spanned by $\{\boldsymbol{e}_i\}$, yielding the projected (or reconstructed) points $\{\boldsymbol{x}'_n\}$ and corresponding abundance vectors $\widehat{\boldsymbol{a}}_n = (\widehat{a}_{n1}, \ldots, \widehat{a}_{np})$. A pseudo-code that describes the recursive implementation is shown below. After finding the $\widehat{\boldsymbol{a}}_n$, the projected points are determined via $\boldsymbol{x}'_n = \boldsymbol{E}\widehat{\boldsymbol{a}}_n$.

Algorithm 1: simplex_project

input : $\boldsymbol{x}, \boldsymbol{E}$
$\boldsymbol{x}$: $d \times N$ (N data points)
$\boldsymbol{E}$: $d \times p$ (p endmembers)
output : $\boldsymbol{a}$
$\boldsymbol{a}$: $p \times N$ (N abundance vectors)

begin
  if $p = 1$ **then**
    $\boldsymbol{a} = \mathbf{1}$;
    return;
  $I = \{\}$;
  for $\forall n \in [1, \ldots, N]$ **do**
    Calculate partial abundance $\boldsymbol{v}^n$ with (3.15);
    if $\forall j : v_j^n \geq 0$ *and* $\sum_j v_j^n \leq 1$ **then**
      $\boldsymbol{a}_n = [1 - \sum_j v_j^n, \boldsymbol{v}^n]$;
    else
      Project $\boldsymbol{x}_n$ onto simplex plane with (3.4);
      $I := I \cup \{n\}$;
  Calculate endmember distance matrix D;
  for $\forall i \in [1, \ldots, p]$ **do**
    Calc. vol. V_i of sub-simplex $E_i = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{i-1}, \boldsymbol{e}_{i+1}, \ldots, \boldsymbol{e}_p]$ with (2.5);
  Calculate the incenter $\boldsymbol{c}$ with (3.8);
  for $\forall i \in [1, \ldots, p]$ **do**
    $I_i = \{\}$;
    for $\forall j \in I$ **do**
      Solve $\boldsymbol{x}_j = \boldsymbol{E}_i^c \boldsymbol{b}^i$ for $\boldsymbol{b}^i$;
      if $\forall k : \boldsymbol{b}_k^i \geq 0$ **then**
        $I_i := I_i \cup \{j\}$;
    if $I_i \neq \{\}$ **then**
      $\boldsymbol{a}^r = \text{simplex\_project}(\boldsymbol{x}_{I_i}, E_i)$;
      $\boldsymbol{a}([1, \ldots, i-1, i+1, \ldots, p], I_i) = \boldsymbol{a}^r$;
      $\boldsymbol{a}(i, I_i) = \mathbf{0}$;

3.3.5 Complexity

The algorithm is designed to finish in a finite number of recursions, where every recursion consists of a finite number of steps. The number of recursions that are necessary depends on the data. In the worst-case scenario, this number equals $\min(N(p-1), 2^p - 2)$, where N is the count of data points and $2^p - 2$ is the count of sub-simplices in a p–simplex. With real hyperspectral imagery, most pixels are likely to reside inside or close to the minimum-volume enclosing simplex and this number is expected to be much lower.

Within a single recursion, the highest computational cost is ascribed to:

- Projection onto the simplex plane: Involves the computation of a $p \times d$ projector matrix and its multiplication with a $d \times N$ data matrix.
- Determining the location of the incenter: Involves the computation of p determinants of symmetric and positive definite matrices with $(p-1)$ rows.
- Determining the vectors $\boldsymbol{b}^i$ for a given cone Z_i: Involves a single inversion of a $(p-1) \times (p-1)$ matrix, and a matrix multiplication of a $(p-1) \times (p-1)$ and a $(p-1) \times N$ matrix. This has to be executed p times.

All other steps concern index management or have negligible complexity. Note that the list does not include the computation of inter-vertex distances for the simplices at each stage, because one can compute the inter-point distances for the entire data once at the start, and then pass them between the calls. This does implicate that one must keep extra storage for $N(N-1)/2$ values. Even by this simply aspect, it is clear that the cost of the algorithm depends on the implementation details and used optimizations and is hard to exactly pinpoint. Therefore, we choose to examine this cost practically, by running the algorithm on simulated and real data using different settings and counting the runtime and recursive calls.

3.4 Experiments

3.4.1 Execution on synthetic data

As a first experiment, we test the SPU algorithm on synthetic data produced as linear mixtures with known abundances and endmembers. The data is constructed for a chosen count of data points N, the number of endmembers p, the dimensionality of the data d and a signal-to-noise ratio (SNR) in the following way. First, the $d \times p$ matrix E containing the endmembers as its columns is generated as a matrix of random numbers in the interval $[0,1]$. The $p \times N$ abundance matrix $\boldsymbol{A}$ is produced

by taking each column as a random sampling from the canonical simplex. For each column a of $\boldsymbol{A}$, this is done by sampling variables $\{a_i\}$ ($i = \overline{1,p}$) independently and uniformly in $[0,1]$, returning $-\log(a_i)$ (yielding an exponential distribution), and normalizing. An efficient algorithm for generating random samples from canonical simplex is also found in [Onn and Weissman, 2011]. Finally, a $d \times N$ noise matrix $\boldsymbol{\eta}$ is constructed by taking random numbers from a Gaussian distribution with a given SNR. According to [Harsanyi and Chang, 1994], the SNR is defined as half the reflectance divided by the standard deviation of the noise. The data set $\boldsymbol{x}$ is then given as: $\boldsymbol{x} = \boldsymbol{EA} + \boldsymbol{\eta}$.

For testing the accuracy and runtime, we compare the SPU algorithm with the FCLSU from Heinz and Chang [2001]. The techniques employed by the FCLSU were mentioned in Sec. 3.1 and it can be fairly said that this implementation has become a *de-facto* standard for computing the abundances of linear mixtures with known endmembers. Both algorithms were implemented in MATLAB on a standard Intel Core 2 Quad 3.0 GHz desktop computer with 4Gb of RAM. No specific effort was done to parallelize the execution of the FCLSU (which treats every data point sequentially and independently from previous runs), while the SPU algorithm could benefit from the simultaneous treatment of the entire dataset by design.

First, we tested the numerical accuracy by calculating the abundance coefficients for several parameter sets with both algorithms and comparing them one by one. We found that, in a typical setting ($p = 5$, $N = 10^5$, $d = 4$, $SNR = 10$), about 99.7% of the abundances retrieved by the SPU differed by less than 10^{-7} from those obtained by the FCLSU (averaged over 100 runs). In the other 0.3% the errors can be attributed to the breakdown of the assumption used in the SPU algorithm, but even then most errors stayed very low. Similar results were gotten for higher values of p, although for any $p > 3$ data configurations yielding significant errors could be artificially induced (recall Sec. 3.3.2). Because of this practical compatibility between the two algorithms, the rest of the experiment was focused on charting the runtime.

Figure 3.6 shows several plots comparing the runtime of both algorithms when changing one of the data parameters while fixing the others. The details of these settings are found in the figure's caption. Starting from the top-left, one sees the dependence of the runtime on the SNR. The first thing to notice is that the runtime of the SPU scales linearly with the number of recursions (as could be expected). The second fact is that the SPU displays much shorter runtime than the FCLSU, which difference increases once the very low SNR values are passed. Within the most practical range, we see that the SPU outperforms the FCLSU by a factor 10-50.

The next plot to the right shows the runtime as a function of the number of endmembers p, on logarithmic scale. It can be observed that the runtime increases exponentially for both algorithms with an increasing p. The SPU once more per-

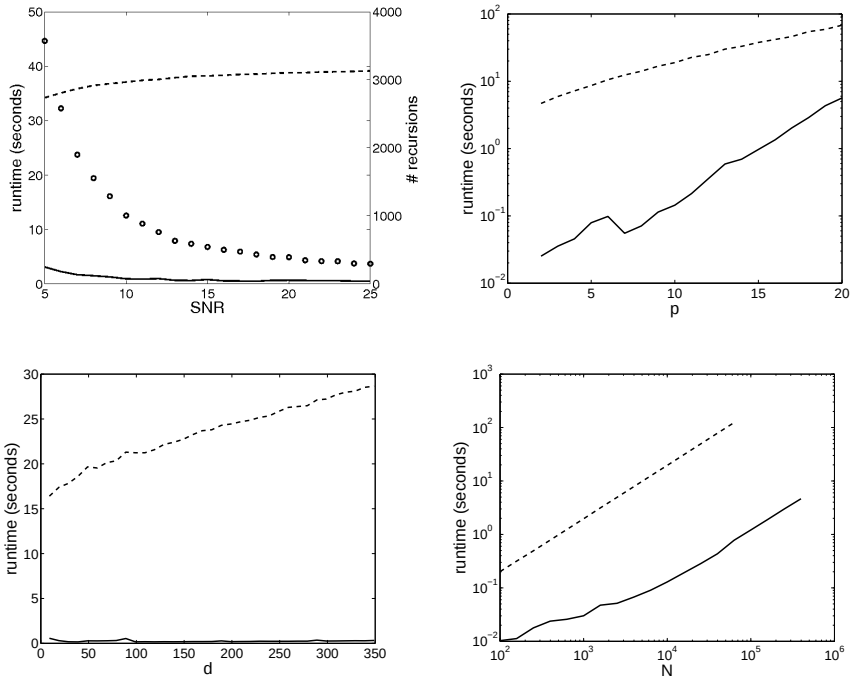

Figure 3.6: Runtime of the SPU (solid lines) and FCLSU (dashed lines) as a function of SNR ($p = 15$, $d = 50$, $N = 10^4$), the number of endmembers p ($N = 10^4$, SNR = 10, $d = 50$), dimensionality d ($N = 10^4$, SNR = 10, $p = 10$) and the number of data points N (SNR = 10, $p = 10$, $d = 50$). The circles in the top-left plot indicate the number of recursions in the SPU algorithm.

forms much better, but the slightly larger slope of the SPU's curve indicates that for very large p this can be reversed in favor of the FCLSU. Still, such high values of p are rarely encountered in practice.

In turn, the plot at the bottom left shows the runtime as a function of the dimensionality of the data d. One sees that the runtime of the FCLSU increases quasi-linearly with increasing d, whereas the curve of the SPU is mostly constant. The latter can be explained by the fact that the most intensive steps of the SPU algorithm involve the calculation of matrix inverses, where the size of these matrices depends only on the number of endmembers p. The overall conclusion is once more that it takes drastically shorter times to complete the SPU.

Finally, the bottom-right plot shows the dependence on the number of data points N. We see that the curve of the FCLSU is linear, as could be expected from the fact that the FCLSU processes data sequentially. The runtime of the SPU algorithm appears quasi-linear, although the exact relation is not so clear from the plot. The runtime of the SPU are again notably lower than those of the FCLSU. Moreover, the slope of the SPU curve is smaller, indicating that the gap would become even larger for very large datasets.

Execution of both algorithms with a profiler show that the FCLSU uses 75% of the runtime for calculating matrix pseudo-inverses. The other 25% is mainly used for memory management and other matrix manipulations. The SPU algorithm virtually spends all the time on matrix manipulations. This suggests that careful profiling and optimization could be instructive in realizing even larger performance gains.

3.4.2 Execution on a hyperspectral image

As we have seen, the SPU offers significantly lower runtime than the FCLSU when using random data. Such large gain clearly deserves to be asserted on real hyperspectral data as well. The dataset we consider here is an excerpt from the infamous AVIRIS Cuprite image that we also used earlier in Sec. 2.3.2. In this case, we select the entire "f970619t01p02_r02_sc04.a.rfl" file, yielding a $512 \times 614 \times 224$ data cube. After eliminating several low-SNR bands, we keep the same 188 bands as specified in [Chang and Wang, 2006]. An RGB view of this image is shown in Fig. 3.7.

Since both the FCLSU and SPU assume that the endmembers are known, the first thing to do was to select some endmember-extraction method to identify them. In this case we choose the N-findR algorithm that had been proven to perform well on this data set (see in Sec. 2.3.1). The next question was how to select an appropriate number of endmembers p. Taking into account various studies and results from the literature, and minding that very accurate endmember spotting was no longer our priority, we used all values of p in the range $[8, 15]$. For every value of p, the

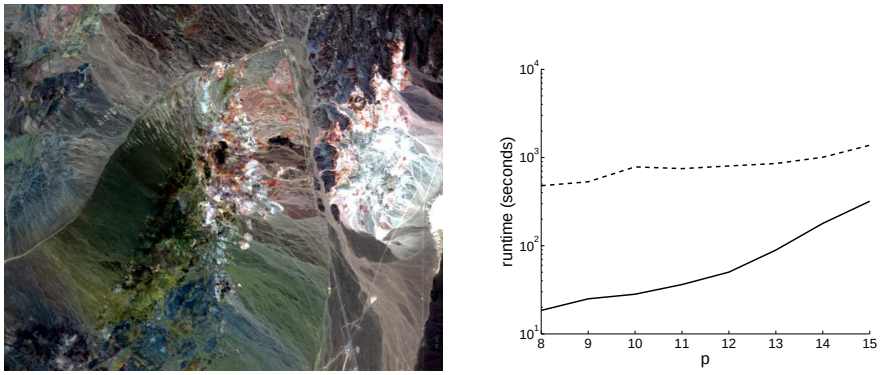

Figure 3.7: *Left*: An RGB view of the used excerpt from the Cuprite image (R=731, G=557, B=458nm). *Right*: The runtime for the SPU (solid line) and the FCLSU (dashed line) as a function of the number of endmembers p ($N = 314368$, $d = 188$).

extracted set included those endmembers that should be expected (see Sec. 2.3.1), so we settled with simply executing the SPU and FCLSU with different subsets of $\{p : 8 \leq p \leq 15\}$. The runtime as a function of p is shown in Fig. 3.7. From this result, we can fairly say that the SPU once more shows superior performance relative to the FCLSU (at least for this image and $p \in [8, 15]$).

3.5 Conclusion

We have presented a distinctive approach and solution for fully-constrained linear unmixing with known endmembers. The approach is inspired by the known equivalence between least-squares optimization and subspace projection, but takes this further, by involving the unmixing constraints to restrict the target subspace for the projection. The result is a simplex-projection unmixing (SPU) algorithm, which recursively projects the data onto sub-simplices of lower dimensions, thereby decreasing the dimensionality of the problem by one with each step. The algorithm is well suited for large data, because many of its operations depend solely on the endmembers and can be executed simultaneously for all data points. An implementation of the algorithm in pseudo-code is given, along with a motivation and proof of each step. While noncomplying cases are feasible, they concern highly skewed geometries that are less common in practice. An efficient procedure for detecting such cases has also been presented.

We have validated the SPU algorithm on random data and an excerpt from the AVIRIS Cuprite image, by comparing it with the benchmark FCLSU method. The SPU and the FCLSU yield almost identical results, except for negligible percentage of pixels. The runtimes of the SPU, however, are drastically lower for most practical settings. Only for very high numbers of endmembers does the SPU become computationally inefficient relative to the FCLSU.

One of the original motives for seeking an analytical solution to the FCLSU problem, which the SPU provides, was the prospect of reformulating such solution in terms of distance geometry. This would in turn allow incorporating the geodesic distances, thereby yielding a more proper nonlinear computation of abundances (relative to the one from Sec. 2.2.4) that would obey both unmixing constraints. A full implementation of this idea is described in a follow-up work by [Heylen and Scheunders, 2011].

PART II

APPLICATIONS

CHAPTER 4

UNMIXING FOR DETECTION OF ADJACENCY EFFECT

Abstract

The adjacency effect is an interesting phenomenon characterized by path interference between the reflectances coming from different ground covers. The effect is caused by atmospheric scatter, so a typical way to detect and correct it has been to model the radiation transfer and spectral correspondence at specific wavelengths. In our approach, we investigate the adjacency effect as being a spectral unmixing problem. This means that we opt to use unmixing to separate the true signature of a pixel from the background scatter reflected from its adjacent neighborhood. To account for different types of scatter, we consider both linear and nonlinear mixing models. We evaluate these models by comparison with a specialized method for detecting the adjacency effect in turbid waters surrounded by vegetated land. This is done both theoretically and on hyperspectral images acquired under varying atmospheric conditions.

N.B. Most of the content from this chapter coincides with the work published in [Burazerović et al., 2013]. In the chapter, the author of the thesis also provides a worthwhile update by giving a more formal (mathematical) backing for the case of unsupervised and fully-constrained linear unmixing.

4.1 Introduction

Spectral interference is an inherent property of reflectance data obtained by remotely sensing the Earth's surface. It refers to the situation where the spectra of multiple ground-cover materials contribute to the spectrum that is observed for a single pixel. One cause of this interference is related to the interaction of the incident and reflected solar radiation at the surface level. The other has to do with backscattering of this radiation in the atmosphere.

At the surface level, the limitations of the remote optical sensor entail that a single pixel will often be recorded as a mixture of disparate materials. The unraveling of such mixed pixels has been widely addressed as spectral unmixing, and some ways of approaching it have been discussed in our previous chapters.

Unlike the mixing that plays out at the surface of interest, another form of spectral interference is caused by atmospheric scattering (Mie and Raleigh scattering on water droplets, mixed gases and aerosol particles) changing the path of photons reflecting outside this surface. This is referred to as the *background-*, *environment-* or *adjacency effect*, and can be clarified by considering the sketch from Fig. 4.1. Most of the photons reaching the sensor's field of view (FOV) of the target surface will normally have reflected on this surface (path 1). However, due to the atmospheric (back)scatter, also photons that have dispersed in the air (path 2), or have reflected on the adjacent surfaces (paths 3, 4 and 5), may deflect into the same FOV. The adjacency effect is practically ascribed to the latter components, as these are responsible for the spectral pollution of the target by its surroundings. Detection and subsequent removal/correction of the adjacency effect can therefore be important for obtaining faithful information about a ground surface. (Think e.g. of retrieving concentrations of non-water constituents in water – see Chapter 5.)

For over three decades, the adjacency effect as described above has been studied (see e.g. [Richter et al., 2006] and its references) and different methods and tools for its detection and correction have since been developed. A longstanding and still popular methodology is to take the adjacency effect into account in atmospheric correction models like FLAASH [Anderson et al., 1999], ATCOR [Richter, 2008] and others [Kayadibi, 2011] by accounting for the surrounding average reflectance. Much effort has been made to investigate the adjacency effect through a model-driven approach, by determining the point spread function for a pixel and/or simulating the effect with Monte Carlo algorithms. Some representative algorithms from this category are found in e.g. [Semenov et al., 2011; Arai, 2002]. Recently, an image based approach for detection of adjacency effects has been proposed [Sterckx et al., 2011] that proved to be effective on MERIS (Medium Resolution Imaging Spectrometer) images over inland waters. The crux of this method is to exploit the knowledge

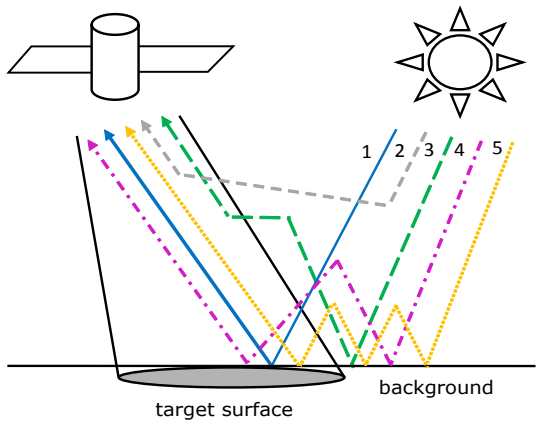

Figure 4.1: Reflectance and (back)scatter of solar radiation. The adjacency effect is ascribed to the reflectance coming from surfaces adjacent to the target surface (background), but deflecting into the FOV of the target (paths: 3, 4 and 5).

about the shape of the water spectrum in the NIR (near infrared) range, which is derived from a theoretical study of the water-leaving reflectance in turbid water in combination with seaborne measurements [Ruddick et al., 2006].

The purpose of our study presented here is to assess the spectral unmixing as a means of detecting and qualifying the adjacency effect. The main benefit of unmixing should reside in its data-oriented operation, and thus avoidance of extensive modeling or inquiry of physical parameters that are typical of specialized treatments (see above). Moreover, while such specialized treatments are mostly concerned with the adjacency effect in its totality, unmixing could effectively qualify its different parts, i.e. the types of reflectance and (back)scatter from Fig. 4.1. This could inadvertently provide extra knowledge about, e.g., the atmospheric conditions. Hence, we investigate unmixing on the basis of linear and nonlinear mixing models, and compare those with the aforementioned method for detecting the adjacency effect in turbid waters that was available to us as the reference.

The remainder of this chapter is organized as follows. Section 4.2 explains the essence of reference NIR spectral-similarity method. Section 4.3 brings in linear and nonlinear mixing models, and derives a mathematical relation between the linear unmixing and the mentioned reference. Section 4.4 describes comparison between the unmixing models, and with the reference, using real images. Section 4.4 summarizes with some general remarks and conclusions.

4.2 The NIR spectral similarity method

Recently, an image-based approach for detection of adjacency effects was described [Sterckx et al., 2011] based on the correspondence with the Near Infrared Similarity spectrum defined in [Ruddick et al., 2006]. The approach was originally developed for high resolution airborne imaging spectroscopy data, but later some successful applications were demonstrated with MERIS imagery. Since then, the approach has been referred to as SIMEC (SIMilarity Environment Correction) [Knaeps et al., 2010] and its essence can be explained as follows.

Based on a theoretical study of the water-leaving reflectance (visible and near-infrared light leaving the water column) in turbid water in combination with seaborne measurements, it was found that all water-leaving reflectance spectra are similarly shaped in the NIR, i.e. 700-900nm range. This spectral shape can be defined entirely by normalization at a single wavelength: 780nm, where water absorption is almost temperature independent, yielding a 'similarity' spectrum. This normalized water spectrum, $r_{w780}(\lambda) = r_w(\lambda)/r_w(780)$ is referred to as the 'Near infrared similarity spectrum' (NIR-ss) and is tabulated in [Ruddick et al., 2006]. SIMEC exploits it by investigating the deviation from its shape to quantify the adjacency effect (above water). The magnitude of the adjacency effect at wavelength λ is then effectively computed as:

$$E(\lambda) = x(780) - \frac{x(\lambda)}{r_{w780}(\lambda)} \quad \left(r_{w780}(\lambda) = \frac{r_w(\lambda)}{r_w(780)}\right) \tag{4.1}$$

where $x(\lambda)$ is the observed water-leaving reflectance. Hence, when $x \equiv r_w$, we have $E = 0$. In the case of water surrounded by *vegetated land*, henceforth 'vegetation', E will be largest at wavelengths where the vegetation reflectance is low compared to the reflectance at 780nm. In practice, λ should be chosen near the red edge, around 700nm. Such fixing of λ implies that E becomes wavelength independent.

Figure 4.2 further illustrates the central idea the SIMEC approach. The water spectrum has a fixed shape at 780nm, as specified by the NIR-ss, but this gets altered when water reflectance is mixed with that of vegetation due to the adjacency effect. The amount of alteration, ranging from 0 for pure water to some maximum value for pure vegetation, is taken as a measure of the adjacency effect.

It is worthwhile noticing that the SIMEC approach in principle enables a broader distinction between two features: water and 'other' that sufficiently differs from water (in the NIR range). This is seen from the use of the NIR-ss water spectrum R_{w780} in Eq. (4.1), where $x(\lambda)$ can be any water-leaving spectrum. When $x(\lambda)$ corresponds to a water spectrum affected by adjacency effects coming from nearby vegetated land, the maximum for $E(\lambda)$ will be observed for λ close to 700nm, but for

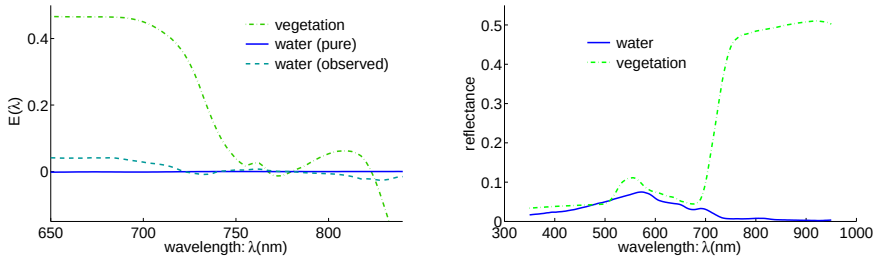

Figure 4.2: Error E computed according to Eq. (4.1) (*left*) using typical vegetation- and water spectra (*right*) and a real water spectrum taken from the test data.

water bordering with other ground-cover types this wavelength could be different. The situation of water being surrounded by vegetated land occurs frequently in nature, and as such it was the main focus in developing SIMEC.

4.3 Unmixing approach

In Sec. 4.1 we suggested that spectral unmixing can be utilized for detecting the adjacency effect, as well as qualification of its different components. This premise can be clarified by revisiting Fig. 4.1. The first thing we notice is that, in the absence of backscatter (paths 4, 5), both the target and the background contribute to the total reflectance independently. Such summation essentially connotes mixing according to the linear mixing model (LMM). The remaining secondary- (path 4) and higher-order interactions (path 5) should thus be expressible via extension to suitable nonlinear models.

4.3.1 Linear unmixing

Looking at Eq. (1.1) and Fig. 4.1, we see that the interaction between paths 1 and 3 can be described adequately by taking $p = 2$ and some $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ as the spectra of the 'target' and 'background'. There is also a clear physical motivation for insisting on the positivity: $a_i > 0$ $(i = 1,2)$, but less so for the additivity: $a_1 + a_2 = 1$. This is because a target pixel could in theory contain other contributions but the ones coming from the two depicted surfaces. The insistence on the latter can be practical, though, for obtaining more quantitative interpretations (when dealing with *reflectance* data and neglecting the higher-order reflections).

Interestingly, it turns that we can also derive a mathematical relation between the LMM and the earlier explained SIMEC. Let $\boldsymbol{w}$ and $\boldsymbol{v}$ designate the two endmembers, i.e. the spectra of water and vegetation, and let $\boldsymbol{w}$ correspond to a pure water spectrum. Thus, $w(\lambda)/w(780) = r_{w780}$ (see above). Assume that the reflectance of a pixel $\boldsymbol{x}$ is a linear mixture of $\boldsymbol{w}$ and $\boldsymbol{v}$. Writing out Eq. (1.1) and (4.1) then yields:

$$x(\lambda_p) = a_1 w(\lambda_p) + a_2 v(\lambda_p) \tag{4.2}$$

$$E = E(\lambda_p) = x(780) - \frac{w(780)}{w(\lambda_p)} x(\lambda_p) \tag{4.3}$$

where $a_i (i = 1,2)$ are the abundances obtained from linear unmixing and λ_p is the particular λ that was chosen to make E wavelength independent, i.e. $\lambda_p \approx 700$nm. Substituting $x(\lambda_p)$ into $E(\lambda_p)$ in Eq. (4.3) leads to:

$$E = a_2 \cdot \left[v(780) - \frac{w(780)}{w(\lambda_p)} v(\lambda_p) \right] \propto a_2 \tag{4.4}$$

since $\boldsymbol{v}$ and $\boldsymbol{w}$ are both known. In other words, the magnitude of the adjacency effect computed by SIMEC is proportional to the abundance of vegetation in water, as determined by linear unmixing. This unmixing in particular requires solving two equations in the form given by Eq. (4.2), since λ is taken at $700nm$ and $780nm$ and no further constraints are imposed.

Despite the mathematical resemblance, the two methods operate from two completely different angles. While SIMEC exploits prior model information in the form of carefully chosen wavelengths and the NIR-ss, the unmixing is data driven. Another difference is given by the constraints. Although SIMEC does not explicitly assume any constraints, the specific choice of the NIR-ss and the wavelength of $\lambda_p \simeq 700nm$ implies that E will be positive, since $x(\lambda_p) \simeq w(\lambda_p)$ (see also Fig. 4.2). When approaching the adjacency effect as a mixing problem, it is clear that the abundances should be positive. The sum-to-one constraint does not seem mandatory (see above), but with standard linear unmixing this constraint is normally also enforced.

Fully constrained unmixing

Equation (4.4) has shown that the unconstrained linear unmixing yields abundance a_2 that is up to a scaling factor equal to E obtained from SIMEC, where the choice of the vegetation endmember determines the scaling. This proportionality stays valid when imposing the sum-to-unity constraint. Namely, the derivation of Eq. (4.4) was made for any a_1 and a_2, so it also true when $a_1 + a_2 = 1$. However, when the knowledge about the two spectral bands used by SIMEC is not accounted for, the

fully constrained unmixing can also be performed unsupervised using all available spectral bands. For this, a water- and a vegetation endmember are obtained from the reflectance image (e.g. using some endmember-extraction algorithm) and the abundances are computed using the FCLSU. While such blind unmixing is beneficial in that it eliminates the need for any prior information, intuitively we expect that it will still yield results comparable to those produced by SIMEC.

The latter premise can be investigated more formally knowing that the fully constrained linear unmixing can be effectively performed after reducing the data to a subspace of dimension $p-1$, where p is the number of endmembers. This is typically realized via the PCA or some other linear DR technique (see in Sec. 1.3.3). The idea is now that the unmixing will implicitly emphasize the most telling spectral bands, which SIMEC uses by design. To examine this, assume once more that spectrum $\boldsymbol{x}$ is a linear combination of $\boldsymbol{w}$ and $\boldsymbol{v}$, but now also obeying the additivity:

$$\begin{gathered}\boldsymbol{x} = a_1\boldsymbol{w} + a_2\boldsymbol{v} = a_2(\boldsymbol{v}-\boldsymbol{w}) + \boldsymbol{w} \\ a_1, a_2 > 0\,,\ a_1 + a_2 = 1\end{gathered} \tag{4.5}$$

From earlier explained theory, we know that such $\boldsymbol{x}$ will reside on a 1-dimensional simplex, i.e. the line segment delimited by $\boldsymbol{w}$ ($\boldsymbol{e}_1$) and $\boldsymbol{v}$ ($\boldsymbol{e}_2$). See Fig. 1.5. Moreover, we know that the PCA computes a new orthonormal basis $\boldsymbol{Q} = \{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_m\}$ for data $\boldsymbol{X}^{\mathrm{T}} = [\boldsymbol{x}_1\, \boldsymbol{x}_2\, \ldots\, \boldsymbol{x}_n]$ in such a way that the first vector or principal component $\boldsymbol{q}_1$ maximizes the variance of the data projected onto it (see Sec. 1.3.3): $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{q}_1$, where $\boldsymbol{x}_i$ and $\boldsymbol{q}_i$ are $m \times 1$ vectors and $n > m$. For $\boldsymbol{X} \equiv \boldsymbol{x}$ from Eq. (4.5), $\boldsymbol{q}_1$ will point in the direction of $\boldsymbol{v}-\boldsymbol{w}$, because the only change happens along that direction (via a_2). Hence, we can express the projection $\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{v}-\boldsymbol{w})$ using Eq. (4.5), to yield:

$$a_2 = \frac{(\boldsymbol{x}-\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{v}-\boldsymbol{w})}{(\boldsymbol{v}-\boldsymbol{w})^{\mathrm{T}}(\boldsymbol{v}-\boldsymbol{w})} = \frac{1}{\|\boldsymbol{v}-\boldsymbol{w}\|^2}\sum_{j=1}^{m}[x(\lambda_j) - w(\lambda_j] \cdot [v(\lambda_j) - w(\lambda_j)] \tag{4.6}$$

We see that the terms with λ_j for which $\boldsymbol{v}$ and $\boldsymbol{w}$ differ the most will also contribute the most. In fact, with the adjacency effect being small ($\boldsymbol{x} \approx \boldsymbol{w}$), $\boldsymbol{v}$ and $\boldsymbol{w}$ having the shapes from Fig. 4.2 and a sparse sampling of λ, only few $\lambda_j > 700$nm will really count. So, there is a resemblance between a_2 and E from Eq. (4.3). In Sec. 4.4.3 we investigate this resemblance experimentally.

4.3.2 Generalized bilinear model

An extension to the LMM that also accounts for second-order interactions between the endmembers is the bilinear mixing model (see Sec. 1.3.3). A convenient form

called the generalized bilinear model (GBM) was recently studied in [Halimi et al., 2011a], which for a given pixel $\boldsymbol{x}$ is formulated as:

$$\boldsymbol{x} = \sum_{i=1}^{p} \widehat{a}_i \boldsymbol{e}_i + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \gamma_{ij} \widehat{a}_i \widehat{a}_j \boldsymbol{e}_i \odot \boldsymbol{e}_j + \boldsymbol{n} \tag{4.7}$$
$$0 \leq \gamma_{ij} \leq 1 \quad \forall i \in [1, p-1], \ \forall j \in [1, p]$$

We see that the main difference between the GBM and LMM is the summation of the interaction terms for endmembers taken by their Hadamard product: $\boldsymbol{e}_j \odot \boldsymbol{e}_k = [e_{j1}e_{k1}, \ldots, e_{jL}e_{kL}]^{\mathrm{T}}$. The convenience of the form given by Eq. (4.7) arises when examining the role and constraints on the coefficient γ_{ij}. (The constraints on a_i from Eq. (1.1) remain valid but now apply to $\widehat{a}_i$.) When $\gamma_{ij} = 0$, GBM reduces to LMM and $\widehat{a}_i$ becomes equal to a_i. This practically tells that γ_{ij} controls the amount of interaction between endmembers $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ in pixel $\boldsymbol{x}$, and thereby the degree of nonlinearity in the mixture. An implication of having such pixel-dependent γ_{ij} is in that it enables a highly flexible fitting of a bilinear model (one could argue that even more flexibility can be achieved by dropping any constraints on γ_{ij}).

The utility of GBM for detecting the adjacency effect can be understood directly by considering how it extends the LMM. In the case of two endmembers, $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, the GBM introduces an additional multiplicative term $\boldsymbol{e}_1 \odot \boldsymbol{e}_2$. Looking at Fig. 4.1, this term can be associated with the interaction between the 'target' ($\boldsymbol{e}_1$) and 'background' ($\boldsymbol{e}_2$) according to path 4. This reading also suggests that the remaining multiple backscattering (path 5) could be tackled by extending Eq. (4.7) with quadratic and higher-order powers; yet, such extensions are normally considered impractical and are therefore avoided (see [Somers et al., 2009] and Sec. 1.3.2).

The determination of the unknown parameters of the GBM: the abundances $\widehat{a}_i$, the nonlinearity-coefficients γ_{ij} and the noise variance σ_n^2 was implemented in [Halimi et al., 2011a] using a hierarchical Bayesian estimator. The benefit of this approach resides in its ability to converge and fit Eq. (4.7) to particular data, but this comes at the cost of notable computational complexity. Therefore, an alternative implementation from [Halimi et al., 2011b] opts to estimate the mentioned parameters by resorting to some common optimization strategies (e.g. gradient descent), and it is also the implementation we adopt for our experiments.

4.3.3 Data-driven nonlinear unmixing

One way to address multiple interactions between the endmembers but avoid extensive model fitting is to consider a data-driven approach. Here, we take as suitable

our algorithm described in Sec. 2, which we shall henceforth refer to as the 'nonlinear unmixing' (NLU). Because Eq. (2.2) does not presume an explicit F, it could cover both linear and bilinear data dependencies and beyond. This is what makes it attractive as an integral method for addressing all different interference paths depicted in Fig. 4.1. One may simply take two endmembers for the 'target' and 'background' (same as with the LMM and GBM) and use the fractional volumes from Eq. (2.5)–(2.8) with geodesic distances to determine their 'nonlinear abundances'. However, due the uninspiring geometry of having only 2 endmembers, using multiple (non-water) endmembers seems advisable for this method. Another aspect is that the implicitness of Eq. (2.2) disallows reconstructing the pixels spectra $\boldsymbol{x}$, which makes the NLU less suitable for quantitative comparisons.

4.4 Experiments

In this section we give an empirical assessment of our considered mixing models as a means detecting and qualifying the adjacency effect. This is done on real images via comparison with the adopted reference method: SIMEC, as well as by cross comparing the mixing models among themselves.

4.4.1 The dataset

The test data consists of two different sets of MERIS-FR images that were available to us from a VITO archive. The images had been acquired above coastal and inland waters under varying atmospheric conditions. The first set corresponds to a monitoring of the artificial Lake Manantali in Mali, Africa, while the second presents an extended view of the Belgian coastline. Both scenes had been judged suitable for studying adjacency effect, in that none was expected to include substantial degree of mixing on the surface of interest (water). In the case of the Belgian coastline, we additionally confined the view to a region-of-interest (ROI) starting at around 700m from the coast and ending at several kilometers into the open sea. This was done to exclude the likely stretches of shallow water (under normal tide), as well as the far-away stretches where the adjacency effect is anyhow not likely to occur. All this was to ensure that any impurity in the observed water-leaving reflectance could be fairly interpreted as a predominant manifestation of the adjacency effect.

The actual image selections are shown in Fig. 4.3 and 4.5, and are henceforth referred to as *set 1* and *set 2*. The mentioned ROIs used with *set 2* are also shown. Figure 4.4 furthermore shows some additional and freely available images that give more idea about the imaged site in *set 1*. All the hyperspectral images designate

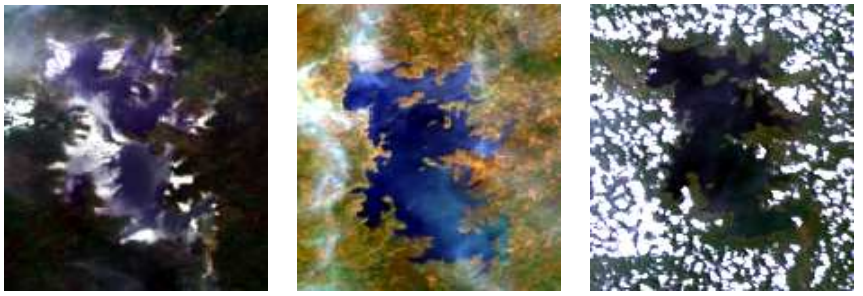

Figure 4.3: An RGB view (R=680, G=560, B=412nm) of the images constituting *set 1*, referred to as 1.1 (*left*), 1.2 (*middle*) and 1.3 (*right*).

Figure 4.4: An areal photo of the Manantali Dam and lake (source: *Wikipedia*) and an image of the lake acquired by NASA Landsat 7 satellite, in 2003 (source: *www.earthobservatory.nasa.gov*).

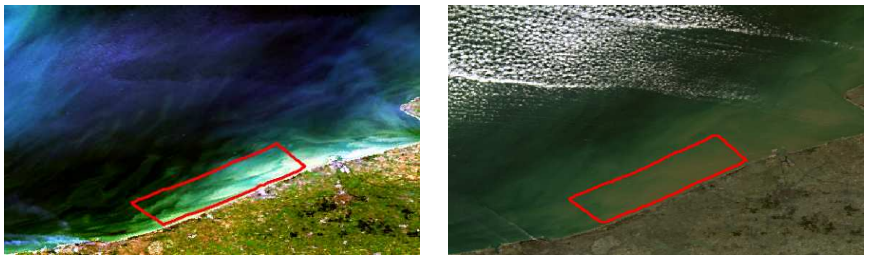

Figure 4.5: A RGB view of the images from *set 2*, denoted as 2.1 (*left*) and 2.2 (*right*). The quadrilaterals depicted in red designate the ROIs used for the actual assessment.

Table 4.1: Atmospheric parameters and meta-data (AOT = Aerosol Optical Thickness)

Image	Date	AOT at 675nm	AOT at 440nm	Water vapor (cm)
1.1	16/09/05	0.3398	0.3844	3.4344
1.2	09/11/05	0.1743	0.2036	1.6362
1.3	25/08/05	0.3839	0.4097	3.9198
2.1	27/09/11	0.2254	0.4386	1.3740
2.2	10/08/11	0.0925	0.1197	1.4807

a same (roughly linear but not uniform) 15-band sampling of the reflectance in the 412-885nm range, having a slightly varying spatial scale and sub-setting within each set. The image size amounts around 145x155 pixels for *set 1* and for *set 2* this is 250x420 pixels. In both sets full resolution MERIS data with a resolution at sub-satellite point 300m was used. Furthermore, all the images were atmospherically corrected using the same Modtran-based technique described in [Haan de and Kokke, 1996], but no specific adjacency correction was applied. The accompanying atmospheric parameters, as measured by the nearest AERONET station, are listed in Table 4.1.

4.4.2 Experimental setup

The main purpose was to compare unmixing and SIMEC using a same set of well-defined classes characterizing the target – water and its adjacent surroundings. In our imagery these classes were reasonably expected to include few endmembers abundantly represented as pure pixels. This is why we could afford to extract them in a supervised and largely manual fashion. Once known, these endmembers were usable according to Sec. 4.3 to unambiguously solve each of the unmixing equations.

In the case of *set 1*, two endmembers – one for water and one for vegetation were obtained by selecting or averaging representative pixels from the lake's center and surroundings. The spectral variability of *set 2* was larger, so the endmembers for this set were obtained via a two-step process: automatic extraction via algorithms such as N-findR (see Sec. 2.1) and SMACC [ENVI, 2009], followed by manual refinement. The latter step basically refers to discarding of endmembers based on our visual inspection of their spectra and unmixing results (notice e.g. the cloud-like

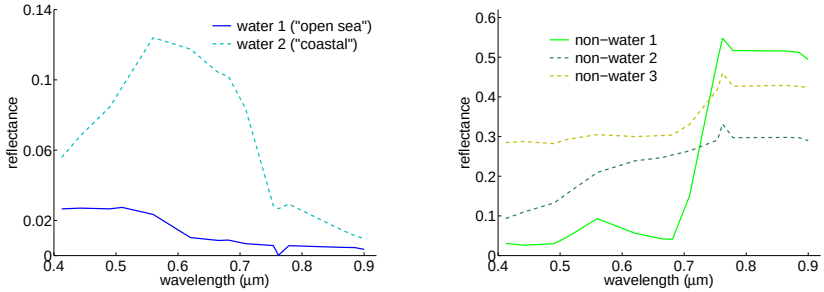

Figure 4.6: Endmembers extracted from *set 2*. The water endmembers have been named by interpreting their spatial location in the image.

content in image *1.3*). This process resulted in 1-2 endmembers for 'water' and 2-3 endmembers for 'non-water', which were then used to yield representative endmember(s) for each class. The water endmembers differed between the coastal and 'open-sea' regions, so the former was logically selected for unmixing the ROIs. The non-water endmember pixels were used in conjunction, selectively (minding their spatial proximity to the ROIs) or simply averaged. Figure 4.6 illustrates some of this. Note that we employed multiple non-water endmembers also as a way of inducing more complex data manifolds and thereby accommodating some of the mixing models, in particular the NLU from Sec. 4.3.3.

The endmembers were used in the mixing models according to Sec. 4.3 to generate three sets of abundance-maps: $a_i(x,y)$, $\widehat{a}_i(x,y)$ and $\tilde{a}_i(x,y)$, and in the case of GBM also the 'gamma-maps' $\gamma_{12}(x,y)$. In turn, the output of SIMEC was molded into an 'adjacency-map' by relating $1-\|E(x,y)\|$ to the intensity of pixel (x,y), with E given by Eq. (4.1) and $\|.\|$ a normalization confining it to interval $[0,1]$. (While Eq. (4.1) does not impose explicit limits on E, typical water- and vegetation shapes will ensure that E does not exceed $[0,1]$ by much. See also Fig. 4.2.) This transformation of E was done for the convenience of obtaining a same range and suitable visual interpretation for all the maps. This connotes that darker pixels are associated with larger deviations from 'pure' water, i.e. a stronger adjacency effect. The mapping of γ_{12} in its unaltered form means that brighter pixels in a 'gamma-map' will correspond to more nonlinearly mixed pixels.

Lacking a ground truth for our images, i.e. a reference information about the magnitude of adjacency effect playing in each pixel, a direct way to obtain *qualitative*

comparisons was to visually interpret the different maps over the regions-of-interest (ROIs). With *set 1*, the ROIs covered the entire water surface (the lake), while with *set 2* they were confined to the quadrilaterals depicted in Fig. 4.5 and motivated above. To isolate the ROIs, each map was multiplied with a matching and manually constructed binary mask. While such masking was already useful in accentuating the relevant detail in each map, further contrast enhancement was achieved via classical histogram manipulation [Gonzalez et al., 2003]. In particular, the abundance- and adjacency-maps were effectively transformed by translating their input range to a same new range, typically $[0,255]$, using linear scaling. The gamma-maps gave more skewed histograms and had to be treated using both linear and nonlinear contrast-stretching transformations.

4.4.3 The comparison

The LMM case

In Sec. 4.3.1 we established the equivalence between SIMEC and LMM, given that the latter is applied without unmixing constraints and using the specifications of SIMEC: the NIR-ss and the two spectral bands. This theory was confirmed by our computations, where the difference between a_1 and $1-\|E\|$ indeed consisted only of a largely constant scaling factor specific for each image (as it depends on the shape of the non-water endmembers). The question was, then, if notable resemblance with SIMEC would also follow for the fully-constrained LMM (FCLSU) using all spectral bands, as we anticipated in Sec. 4.3.2.

We show this comparison in Fig. 4.7, for brevity only for selected images from Fig. 4.3–4.5. Overall, we see an impacting symmetry between the results of FCLSU and SIMEC. The same picture was obtained with the remaining images and when using more than two endmembers. While these qualification concerns the contrast-stretched maps, the histograms of the un-stretched data (shown for the image with the most deviating results) give an idea about the scale. The same scale can be used to derive a sense about the magnitude of the adjacency effect, which is in essence proportional to the count of values deviating from 1 (pure water). Recall from Sec. 4.4.1 that no mixing was expected on the water itself, making all deviations from it mostly imputable to the adjacency effect.

Summarizing, we can say that, while there was some difference between FCLSU and SIMEC, it was showing mostly in the scale and not so much the spatial distribution of the affected pixels. It is worthwhile to point out that the largest differences were observed with the images from *set 2*, which can be partly attributed to a greater spatial scale and larger spectral variability of those data. This created more freedom

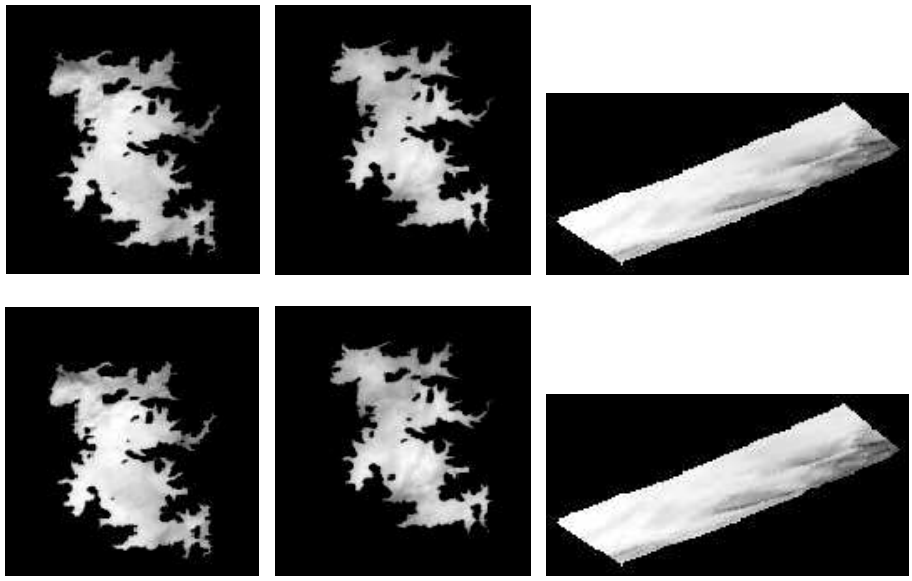

Figure 4.7: Contrast-stretched maps showing the adjacency-effect detection for images *1.2*, *1.3* and *2.1*. *Top*: Inverted spectral error $(1 - \|E\|)$ obtained from SIMEC. *Bottom*: Abundance-map for water (a_1) computed by FCLSU.

in selecting the endmembers and thus a larger chance for disagreement with the preset information used by SIMEC.

The GBM case

A meaningful exercise for the GBM was to examine the difference between its linear abundances $\widehat{a}_1$ and their counterparts a_1 produced by the LMM, and how it would relate to γ_{12}. Since the role of γ_{12} is to promote flexibility of fitting a bilinear mixing model (other models are mentioned in, e.g. [Halimi et al., 2011a]), the question was whether the GBM would yield a worthwhile distinction with the LMM. Figures 4.9 and 4.10 show the results of applying the GBM using two endmembers and the ROI images from *set 1* and *set 2*. One sees that there is a general resemblance between the abundance-maps produced by the GBM and their counterparts obtained by applying the LMM under the same conditions (see Fig. 4.7). For all images but *1.3*

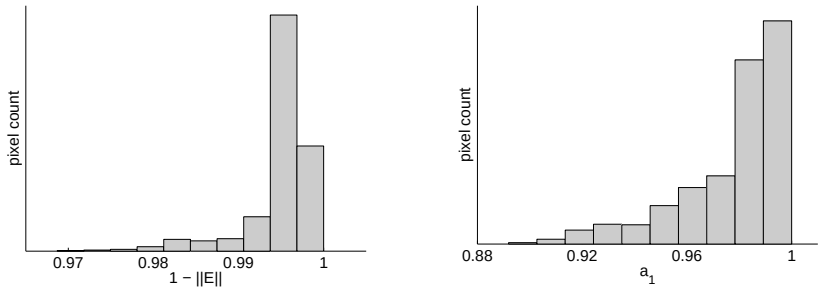

Figure 4.8: The un-stretched histograms of $1-\|E\|$ (*left*) and a_1 (*right*) corresponding to the two images from the far right column in Fig. 4.7.

and *2.1*, there is a non-trivial picture of $\gamma_{12} \neq 0$ largely complementing that of $\widehat{a}_1$. By non-trivial we mean other than the typical picture of $\gamma_{ij} \neq 0$ being concentrated at the spatial boundary between classes i and j [Halimi et al., 2011a]. Furthermore, we can say that all the gamma-maps appear 'logical', in the sense that $\gamma_{12} \neq 0$ are mostly confined to the coastal regions where one should also expect the largest chance for secondary reflections (between water and vegetated land).

The next step was to evaluate the unmixing results by computing some commonly used error measures [Keshava and Mustard, 2002]. One of the most prevalent such measures is the RMSE, which calculates the average difference between the original N pixels $\boldsymbol{x}$ and their reconstructions $\widehat{\boldsymbol{x}}$ obtained from particular model, as:

$$RMSE = \left(\frac{1}{NL} \sum_{i=1}^{N} \|\boldsymbol{x}(i) - \widehat{\boldsymbol{x}}(i)\|^2 \right)^{1/2} \tag{4.8}$$

where L is the number of spectral bands. Because the RMSE is often explicitly minimized when estimating parameters of unmixing models (as is the case with e.g. FCLSU), it is often prudent to consider additional measures, like the *spectral angle*:

$$SA = \frac{1}{N} \sum_{i=1}^{N} \arccos \left(\frac{\langle \boldsymbol{x}(i) \widehat{\boldsymbol{x}}(i) \rangle}{\|\boldsymbol{x}(i)\| \|\widehat{\boldsymbol{x}}(i)\|} \right) \tag{4.9}$$

A worthwhile distinction in assessing these measures was to be made between pixels with small and large γ_{12}. The motivation for this discernment is a premise that a larger γ_{12} should be justified by a better fit of the GBM compared to the LMM;

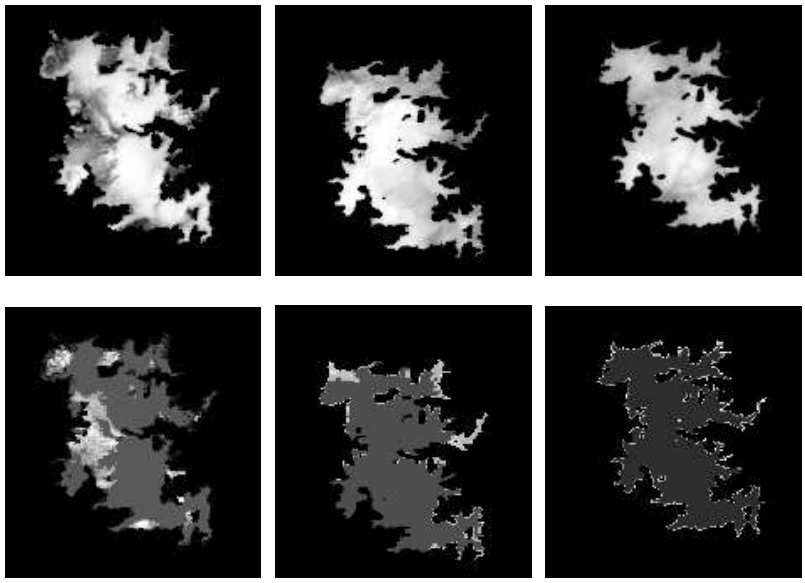

Figure 4.9: Contrast-enhanced results of GBM using two endmembers and images from *set 1*. *Top*: abundance-maps for water ($\widehat{a}_1$). *Bottom*: gamma-maps showing the nonlinearity coefficient between water and non-water (γ_{12}).

else, one can argue that it is always possible to enlarge γ_{12} at the expense of $\widehat{a}_1$ while maintaining the same RMSE. A fair complementarity of $\widehat{a}_1$ and γ_{12} is in fact already seen from Fig. 4.9 and 4.10. Thus, we subdivided γ_{12} into 2 subsets: a set of pixels with negligible values of γ_{12} (the dark-grey pixels in a gamma-map) and a set of pixels with values of γ_{12} that significantly differ from 0. Table 4.2 shows this assessment for the relevant images and the situation with two endmembers. We see that the subsets with large γ_{12} indeed consistently induced error decrements for the GBM relative to the LMM, an effect which is best visible in images from *set 1*. This suggests that, at least in some pixels, there was a genuine play of secondary reflections.

Notice that none of the above error measures was usable with SIMEC, due to the lack of an upper limit and thus absolute scale for E from Eq. (4.1). Another reason for not bringing in SIMEC is that we already established its resemblance to the LMM.

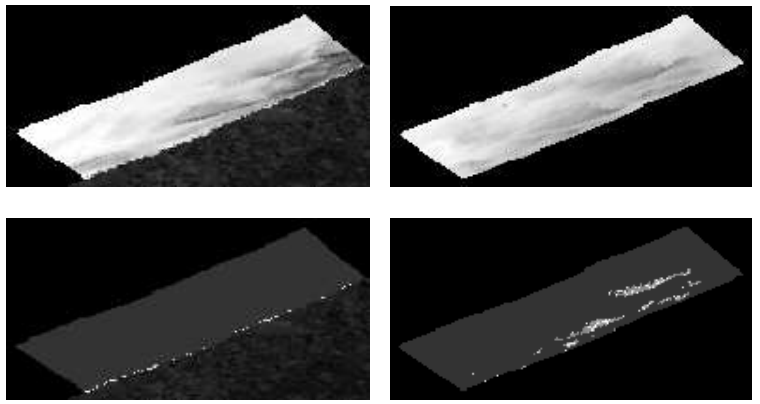

Figure 4.10: Contrast-enhanced results of GBM using two endmembers and images from *set 2*. The reading is the same as in Fig. 4.9. To place the γ_{12} for image *2.1* on the left, the non-water (coast) is also shown by its (attenuated) abundance, $\widehat{a}_2$.

Table 4.2: Errors of (bi-)linear unmixing with 2 endmembers

		RMSE ($\times 10^{-2}$)		SA ($\times 10^{-2}$)	
		LMM	GBM	LMM	GBM
Img. 1.1	a: $\gamma_{12} \simeq 0$	3.83	3.82	9.53	9.54
	b: $\gamma_{12} > 0$	2.00	1.82	6.10	5.57
Img. 1.2	a	1.07	0.98	3.86	3.87
	b	2.32	2.17	9.05	8.6
Img. 1.3	a	1.18	1.17	6.65	6.4
	b	2.62	2.56	12.22	12.03
Img. 2.2	a	4.88	4.88	7.20	7.14
	b	5.17	5.15	9.33	9.14

The NLU case

As explained previously, a suitable way to assess the NLU was to compute its abundance maps $\tilde{a}_i(x, y)$ and compare them visually with their counterparts produced by the GBM ($\widehat{a}_i$) and LMM (a_i). We can be brief about this by saying $\tilde{a}_i$ much resembled $\widehat{a}_i$ and a_i. A more rigorous comparison would examine the histograms. Ideally, the difference between a_i and $\tilde{a}_i$ would be equal to or larger than the difference between a_i and $\widehat{a}_i$, since the NLU is seen as a model that can theoretically encompass both the linear terms and nonlinearities beyond the bilinear multiplication.

Experimentally, we found that the difference between $\tilde{a}_i$ and $\widehat{a}$ was insignificant. This may be ascribed to both the absence of nonlinearity in the data and the NLU model itself. From the theoretical point of view, one should realize that the divergence of $\tilde{a}_i$ from their (bi-)linear counterparts depends on the curvature of the data manifold. Suppose that the nonlinear embedding in Fig. 2.1 was formed by slight bending the sheet of paper in-between the corners of the 2D triangle (which is well describable by a bilinear model). In this case, it would not make much difference if the volume ratios with geodesic distances – determining the $\tilde{a}_i$ (see Sec. 2.2.4), were measured along the bent surface or its 2D projection. At the same time, we have seen that even the GBM produced only a slight difference with the LMM.

Relation to atmospheric parameters

Besides comparing the unmixing results, one may wonder about their possible relation with the atmospheric data. Looking at Table 4.1, an intuitive interpretation of each parameter is that its increment should lead to a larger adjacency effect. Such correlation is indeed observed for some of our images; however, we can also find image pairs (e.g. images *2.1* and *1.2*) with comparable atmospheric data but different relative amounts of non-pure water pixels. Actually, this is not so strange if we realize that Table 4.1 shows only some of the possible atmospheric variables. In practice, such variables are likely to exhibit interdependence and more than just straightforward linear relation with the adjacency effect.

4.5 Conclusion

We have investigated the utility of spectral unmixing as a means of detecting and qualifying the adjacency effect. A distinctiveness of this approach resides in that a data-driven unmixing, which is normally used for unraveling of mixed pixels on the ground, is now engaged for treating an atmospheric effect, which is also typically treated in model-oriented fashion. We have implemented the said unmixing by three

different models, grading from the explicit linear and bilinear, to a more implicit nonlinear model. Each model has been motivated by its theoretical fit to a specific configuration regarding the solar reflection and atmospheric scatter composing the overall adjacency effect. As a reference, we have used a specialized adjacency-detection method that equally operates on a pixel basis but requires prior knowledge (model) about the reflectance of the target (water).

Our theoretical analysis and experiments with diverse imagery show that unmixing can indeed yield similar results as the model-oriented treatment. In this case, linear unmixing proved to be ample, but its bilinear extension was also useful in that it provided a meaningful enrichment of the linear results. Since our images included relatively few and uncorrelated (sparsely sampled) spectral bands, more general multivariate analysis could be instructive in forming additional judgment about the presence of nonlinearity in the data.

Remarks

We are indebted to the Flemish Institute for Technological Research (VITO) for acquisition and preprocessing of the dataset. We also wish to acknowledge B. Chatenet, LISA (Laboratoire Interuniversitaire des Systèmes Atmosphériques) and the PHOTONS and AERONET teams for their efforts in establishing, maintaining and calibrating the instruments and processing of the atmospheric data at Cinzana's aeronet station. Special thanks goes to the authors of Halimi et al. [2011b] for sharing a software implementation of their method. This work was partly supported by the project CHAMELEON of the Flemish Agency for Innovation by Science and Technology (IWT).

CHAPTER 5

UNMIXING FOR WATER-QUALITY RETRIEVAL

Abstract

Estimation of water quality in inland and coastal waters is a vital part of hydrological observation. In this chapter, we study this estimation as a spectral unmixing problem. Our starting point is the provision of an analytical model relating the reflectance of water to its impurities or constituents, as specified by their inherent optical properties (SIOP) and concentrations. To gauge these concentrations via unmixing, we explore a methodology that combines common unmixing with a suitable definition of endmembers exploiting the knowledge of the water-reflectance model. To mitigate the dependence on off-line definitions, we explore the extractability of the endmembers from the reflectance data by analyzing a dataset we construct using the water-reflectance model. Subsequently, we test the derived techniques with real water content, and we use the unmixing abundances to produce the actual concentration maps and compare those with the available reference. This is done for a hyperspectral image acquired over coastal waters of a shallow sea.

N.B. Most of the work described in this chapter has been published in [Burazerović et al., 2014]. In the chapter, the author of the thesis also gives additional clarification and illustrations for several of the used concepts, as well as a mention of their feasible elaboration.

5.1 Introduction

Monitoring of water quality in inland and coastal waters is a vital study area within the all-important field of hydrological observation and water-resource management. Traditionally, the composition and dynamics of water have been gauged via in *situ* measurements that essentially offer a point-based sampling of the water surface [Su et al., 2011]. However, due to the cost and limitations of this methodology in providing extensive spatial and temporal coverage, the use of remote sensing has been proliferating in recent years. The different modalities and techniques are here considered mostly in conjunction with strategic in *situ* sampling, which then serves to derive model parameters or establish a ground reference [Su et al., 2011].

The basic idea upon which rest all water-quality retrieval algorithms is that the water-leaving reflectance (visible and near-infrared light leaving the water column) is largely shaped by the identity and concentration of in-water substances (masses, bodies) or constituents. The exact definition of a water constituent can vary (see e.g. [Hommersom et al., 2011] and references therein), but in coastal and inland waters those that are optically active are: *colored dissolved organic matter* (CDOM), *totally suspended materials* (TSM), or equivalently *suspended particulate material* – SPM or *total suspended sediment/solids* – TSS, and *chlorophyll-a* (CHL). Furthermore, most inland and coastal waters can be classified as 'case 2', which means that all the constituents vary independently from each other [Su et al., 2011]. These observances have inspired many works in which the water-leaving reflectance is essentially inverted to obtain the unknown concentrations. This naturally assumes that all sources which, in addition to the water-leaving reflectance, contribute to the observed signal should be modeled and removed. Examples of these influences are sun- or sky glint Kay et al. [2009], atmospheric effects Salama et al. [2004] (see also our Chapter 4), etc. Figure 5.1 provides an illustration.

One of the prerequisites for inverting the water-leaving reflectance is to express it in terms of the total absorption and backscatter, or the *inherent optical properties* (IOP) of each water constituent. There exist several models that do so analytically using rational functions [Maul and Gordon, 1975] or more complex extensions thereof [Albert and Mobley, 2003]. The second element is to write IOP as products of concentration-*specific* IOP, or SIOP, and the respective concentrations. When the SIOP are known, e.g. from in *situ* or lab measurements, this enables to estimate the concentrations directly from the observed reflectance. One common way of doing so has been to apply the ordinary least squares (OLS), i.e. linear regression Hakvoort et al. [2002]. Another approach has used training of artificial neural networks [Schiller and Doerffer, 2005]. Yet another route has been to invert the spectrum indirectly, by comparing it with a spectral library constructed by substi-

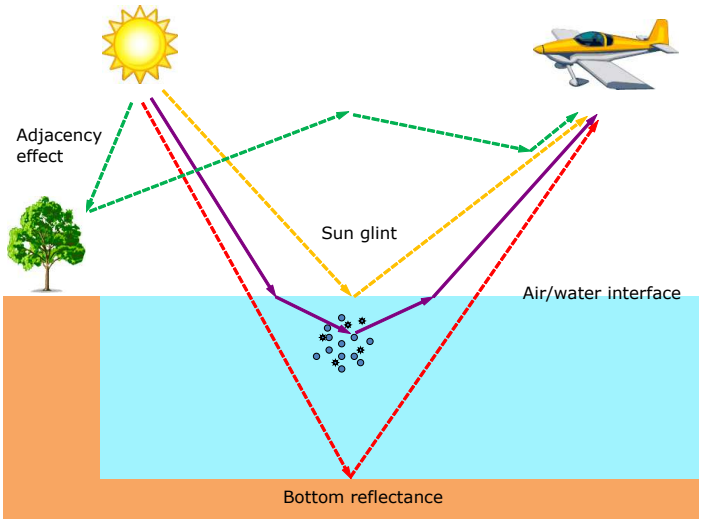

Figure 5.1: Water-quality retrieval via remote sensing. Aside from the interfering signals, the main object of analysis is the water-leaving reflectance shaped by the concentration-specific absorption and backscatter of water bodies, i.e. constituents.

tuting known concentrations into some water-reflectance model [Woerd van der and Pasterkamp, 2008]. A worthwhile study from the latter category, albeit with a different focus – mapping of microphytobenthos biomass, is also found in [Combe et al., 2005]. In general, the main limitation of all model-based methodology is its sensitivity to variations in the reflectance signal, which can be due to sensor noise, variation of SIOP, or inadequate estimation and correction of atmospheric effects.

A different take on the estimation of water-mass concentrations in coastal and inland waters was recently elaborated by [Hommersom et al., 2011] that makes use of spectral unmixing. The central idea was to employ the conventional *linear mixing model* (LMM), as we explained it in Sec. 1.3.1, wherein the endmembers are suitably defined as representations of different *water types* dominated by one or more constituents. Each endmember was then in effect to be calculated by substituting the known (in *situ* measured) SIOP and in turn utmost concentrations of each constituent into an adopted water-reflectance model. Despite the demonstrated ability of this approach to yield good reconstruction of the water spectra, its only qualitative translation of the unmixing result to concentrations and extensive reference to in *situ* measured data remain a limitation. We mitigate this by introducing

several add-ons. To start with, we explore extractability of the endmembers from the data. With the predefined endmembers, and those extracted from the real image, we then perform unmixing and use available ground-truth pixels to derive a model translating the unmixing abundances to actual concentration maps. We compare those with maps precomputed by a curve-matching method that were available to us as a reference. Through all this, we also give a more theoretical assessment of the unmixing approach in relation to the water-reflectance model.

The remainder of this chapter is organized as follows. Section 5.2 explains our framework and the core components in our approach. This includes a prevalent water-reflectance model and a derived endmember model that links the former to spectral unmixing, which we describe in detail in two separate subsection. Section 5.2 discusses our assessment of this endmember model and our method for endmember extraction. Section 5.5 validates the entire unmixing approach on a real hyperspectral data, by comparing its results with the available reference. Section 5.6 closes with conclusions and discussion.

5.2 Methodology

5.2.1 Water reflectance modeling

A common way to analytically relate reflectance of water to its constituents (see e.g. Hommersom et al. [2011], Knaeps et al. [2010], Hakvoort et al. [2002]) is to use the model formulated by Maul and Gordon [1975]:

$$\widehat{R}(0-,\lambda) = f \cdot \frac{b(\lambda)}{a(\lambda)+b(\lambda)} \tag{5.1}$$

where $R(0-)$ is the subsurface irradiance reflectance (depth = 0–) and we use the accent to distinguish the modeled reflectance from the observed (e.g. ship-based or airborne) one. Furthermore, a and b denote the total absorption and backscattering coefficients, and f is a parameter that is specifically related to illumination and viewing conditions but is usually assumed constant for particular water surface (e.g. Dekker [1993] found that for typical Dutch inland waters $0.2 \leq f \leq 0.59$). While other models certainly exist, it is fair to say that most connote some variation of Eq. (5.1) obtained by using more elaborate definitions of f, or using $b/(a+b)$ as a factor in more complex formulas including e.g. higher-order polynomials Albert and Mobley [2003]. We leave aside these elaborations and concentrate on Eq. (5.1), where we shall hence omit f and suppress the wavelength dependency to yield simplified notations. We shall bring this dependence up again wherever required by the context.

A typical assumption in the above analysis is that $a(m^{-1})$ and $b(m^{-1})$ can be decomposed as linear sums of contributions from pure water (seawater) and the other mentioned constituents: CDOM, TSM and CHL. The constituent terms can in turn be written as products of concentration-specific IOP or SIOP (a^* and b^*) and corresponding concentrations, so that $a_{\mathtt{CHL}} = a^*_{\mathtt{CHL}}\mathtt{CHL}$, $a_{\mathtt{TSM}} = a^*_{\mathtt{TSM}}\mathtt{TSM}$ and $a_{\mathtt{CDOM}} = a^*_{\mathtt{CDOM}}\mathtt{CDOM}$. Another convention decomposes the IOP into water, dissolved and particulate fractions Lee [2006]. This dictates splitting the absorption of TSM into its particulate (phytoplankton or algae) and the Non-Algae-Particles (NAP) part, i.e., $a_{\mathtt{TSM}} = a_{ph} + a_{\mathtt{NAP}}$ [Babin et al., 2003]. As still the NAP are dominant in complex waters under consideration, the NAP and TSM concentrations can be assumed to be equal and we can write $a_{\mathtt{NAP}} = a^*_{\mathtt{NAP}}\mathtt{TSM}$ Nechad et al. [2010]. Due to negligible (back)scatter for dissolved matter, one derives the following expressions for the IOP:

$$\begin{aligned} a &= a_w + a^*_{\mathtt{NAP}}\mathtt{TSM} + a^*_{\mathtt{CHL}}\mathtt{CHL} + a^*_{\mathtt{CDOM}}\mathtt{CDOM} \\ b &= b_w + b^*_{\mathtt{NAP}}\mathtt{TSM} \end{aligned} \tag{5.2}$$

where a_w and b_w are the absorption and backscatter of clean water, a^*_i, b^*_i are the SIOP and $\mathtt{TSM}$ (g/m^3) and $\mathtt{CHL}$ (mg/m^3) are concentrations. Concentration of CDOM is measured by its absorption [Kirk, 1994; Babin et al., 2003]:

$$a_{\mathtt{CDOM}}(\lambda) = a_{\mathtt{CDOM}}(\lambda_r)e^{-S(\lambda_r-\lambda)} \tag{5.3}$$

where $a_{\mathtt{CDOM}}(\lambda_r)$ is the absorption (m^{-1}) at a reference wavelength, typically 440nm. Minding the above notation we define $\mathtt{CDOM} \equiv a_{\mathtt{CDOM}}(\lambda_r)$ and $a^*_{\mathtt{CDOM}} \equiv exp(-S(\lambda_r-\lambda))$. A similar formula as given by Eq. (5.3) is also used for $a_{\mathtt{NAP}}(\lambda)$, and since absorptions of NAP and CDOM are difficult to distinguish, ocean color algorithms typically retrieve their sum [Aurin and Dierssen, 2012; Lee, 2006]. Inserting a and b from above into Eq. (5.1) that omits the wavelength dependency then yields:

$$\widehat{R} = \frac{b_w + b^*_{\mathtt{NAP}}\mathtt{TSM}}{a_w + b_w + (a^*_{\mathtt{NAP}} + b^*_{\mathtt{NAP}})\mathtt{TSM} + a^*_{\mathtt{CHL}}\mathtt{CHL} + a^*_{\mathtt{CDOM}}\mathtt{CDOM}} \tag{5.4}$$

or equivalently:

$$\begin{gathered} (a^*_{\mathtt{NAP}} + b^*_{\mathtt{NAP}}\tilde{R})\mathtt{TSM} + a^*_{\mathtt{CHL}}\mathtt{CHL} + a^*_{\mathtt{CDOM}}\mathtt{CDOM} = -a_w - b_w\tilde{R}, \\ \tilde{R} = (\widehat{R}^{-1} - 1) \end{gathered} \tag{5.5}$$

These formulas now permit to clarify some of the approaches for inverting $\widehat{R}$ that we mentioned above. Specifically, one way to estimate the concentrations $c_i \in \{\mathtt{TSM}, \mathtt{CHL}, \mathtt{CDOM}\}$ is to take $\widehat{R}$ and $\{a_i, b_i\}$ in Eq. (5.5) at $n \geq 3$ wavelengths, yielding

a full-column-rank system $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{y}$ that can be solved by OLS. The solution is then given as $\widehat{\boldsymbol{c}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$, where $\boldsymbol{c} = (\mathtt{TSM}, \mathtt{CHL}, \mathtt{CDOM})$ is the vector of concentrations,

$$\boldsymbol{A} = \begin{bmatrix} a^*_{\mathtt{NAP}}(\lambda_1) + b^*_{\mathtt{NAP}}(\lambda_1)\tilde{R}(\lambda_1) & a^*_{\mathtt{CHL}}(\lambda_1) & a^*_{\mathtt{CDOM}}(\lambda_1) \\ \vdots & \vdots & \vdots \\ a^*_{\mathtt{NAP}}(\lambda_n) + b^*_{\mathtt{NAP}}(\lambda_n)\tilde{R}(\lambda_n) & a^*_{\mathtt{CHL}}(\lambda_n) & a^*_{\mathtt{CDOM}}(\lambda_n) \end{bmatrix}$$

and $\boldsymbol{y} = (-a_w(\lambda_1) - b_w(\lambda_1)\tilde{R}(\lambda_1), \ldots, -a_w(\lambda_n) - b_w(\lambda_n)\tilde{R}(\lambda_n))$. Note that the inverse $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ must be computed for each data point, due to the dependence of both $\boldsymbol{A}$ and $\boldsymbol{y}$ on the input reflectance, via $\tilde{R}$. In the literature, authors have often rightfully remarked that the said OLS solution will not guarantee all positive concentrations Knaeps et al. [2010]. This has sometimes been circumvented by fixing one of those, such as CDOM in Hakvoort et al. [2002], which also seems justified considering the aforementioned indistinguishably of $a^*_{\mathtt{CDOM}}$ and $a^*_{\mathtt{NAP}}$. We ought to add that non-negativity of $\boldsymbol{c}$ as a solution to $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{y}$ should in principle be attainable by applying the NNLS algorithm by Lawson and Hanson [1974]. See also Sec. 3.1.

A popular alternative to this matrix inversion has been to avoid it altogether and construct a library of spectra $\{\tilde{R}_i\}$ to be matched the observed spectrum, $\widehat{R}$. These library spectra are then obtained by inserting the known SIOP (a^*_i, b^*_i) and concentrations $(c_i \in \{\mathtt{TSM}, \mathtt{CHL}, \mathtt{CDOM}\})$ into the reflectance model; particularly, the model from Eq. (5.4) allows reading out directly the $\{c_i\}$ that produced some $\widehat{R}$. There exist many approaches that use this basic curve matching idea and they differ mostly in how they assess the difference between $\tilde{R}_i$ and $\widehat{R}$. A recent study from Knaeps et al. [2010] does this by looking at both the classical *root-mean-squared error*, $\mathtt{RMSE} = E[(\tilde{R}_i - \widehat{R})^2]$ and a difference between specific wavelet features from the two spectra. Another approach from Kempeneers et al. [2005] uses *simulated annealing* to optimize the RMSE over all spectra. In the sequel, we shall use the term 'curve-matching' to refer to a combination of these two methods, as this had been used to produce results that we take as reference in our experiments.

We should remark that different treatments of the ocean color data certainly exist, where the above described methods belong to the prevalent group of *semi-analytical* algorithms. This term tells that the bio-optical stages of the *radiative transfer equation* are expressed by empirical relationships Ufermann [2003].

5.2.2 Unmixing approach

When aiming to reformulate the water-quality retrieval as an unmixing problem, one possibility is to take the mathematical view and examine the expressibility of Eq. (5.4) as an unmixing equation. For fixed SIOP, this may connote

Table 5.1: Combinations of $c_{i,\text{low}}$ and $c_{i,\text{high}}$ defining the endmembers via Eq. (5.4).

Endmember	TSM (g/m^3)	CHL (mg/m^3)	$a_{\texttt{CDOM}}(440)$ (1/m)
$\boldsymbol{e}_1$	0	0	0
$\boldsymbol{e}_2$	1	1	0.2
$\boldsymbol{e}_3$	1	1	3
$\boldsymbol{e}_4$	1	60	0.2
$\boldsymbol{e}_5$	1	60	3
$\boldsymbol{e}_6$	100	1	0.2
$\boldsymbol{e}_7$	100	1	3
$\boldsymbol{e}_8$	100	60	0.2
$\boldsymbol{e}_9$	100	60	3

seeking a *multi-univariate decomposition* [Gutierrez et al., 2002] of Eq. (5.4) as: $\widehat{R} = g(h_1(c_1), h_2(c_2), h_3(c_3))$, with $c_i \in \{\texttt{CHL}, \texttt{TSM}, \texttt{CDOM}\}$. The derived $g(\cdot)$ would ideally stage some known mixing function, giving the univariates $\{h_i(c_i)\}$ the role of independent endmembers. By inverting this $g(\{h_i\})$, and in turn each of the $\{h_i(\cdot)\}$, one could then directly translate the endmember abundances to wanted concentrations: $\{c_i\}$. Unfortunately, a clear obstacle for decomposing Eq. (5.4) is its low polynomial order. We could try using more complex water-reflectance models, but even then it is unsure what form of $g(\cdot)$, if any, we could obtain.

An alternative, thus, is to start with an mixing model we know how to invert and invoke Eq. (5.4) in contemplating a suitable definition of endmembers. This approach was recently worked out by Hommersom et al. [2011], whom proposed to use the standard fully-constrained linear unmixing (LMM), while defining the endmembers as representations of *water types* dominated by one or more water constituents. The latter was practically attained by inserting into Eq. (5.4) the SIOP, and in turn extreme concentrations of each constituent, as acquired from in *situ* measurements. In mathematical terms, this unmixing is expressed as given by Eq. (1.1) with the endmembers specified by Table 5.1. The table shows the possible combinations of $c_{i,\text{low}} = \min\{c_i(x,y)\}$ and $c_{i,\text{high}} = \max\{c_i(x,y)\}$, with (x,y) denoting the pixel coordinates. The values of $\{c_{i,\text{high}}\}$ were derived from a study area similar to ours (see Sec. 5.5.2), while $\{c_{i,\text{low}}\}$ were chosen different from 0, because pixels with 100% concentrations are unfeasible in water solutions. The $\boldsymbol{e}_1$ thus represents the spectrum of pure water, as characterized by e.g. [Pope and Fry, 1997].

At this point we ought to recognize some viable alternatives to this LMM-based approach. One is to apply nonlinear (e.g. bilinear) mixing models, using the same or

different definition of endmembers. The use of nonlinear unmixing seems justified by an intuitive notion that a sample of water comprises an 'intimate' mixture (see Sec. 1.3.2) of pure water and other substances. Yet, the endmembers from Table 5.1 already incorporate mixing, since several of them permit high concentrations for more than one constituent. This entails that writing out Eq. (1.1) with $\boldsymbol{x} = \widehat{\boldsymbol{R}}$, while inserting the $\{\boldsymbol{e}_i\}(i = \overline{1,9})$ from Table 5.1, will yield an elaborate expression for $\boldsymbol{R}$ featuring non-negligible cross products of powers of $\{c_i\}$. Hence a nonlinear relation between $\boldsymbol{R}$ and $\{c_i\}$ that determine its shape. Another distinctive take on unmixing may consist of storing the signatures of water endmembers in a dictionary and applying (sparse) regression [Iordache et al., 2011]. One counterargument for using this approach is that it principally resembles the aforementioned curve matching, whereby the water spectra are also matched with a library, only directly.

The ability of the described endmember model to reconstruct the water spectra was already demonstrated in [Hommersom et al., 2011]. Our goal next is to provide a more formal analysis and case for using this model, by considering additional actions, like extraction of the $\{\boldsymbol{e}_i\}(i = \overline{1,9})$ from the data and quantitative translation of their abundances $\{a_i\}$ to actual concentrations. If linearly combining the $\{\boldsymbol{e}_i\}$ can describe any water spectrum, then getting the concentrations by combining the $\{a_i\}$ should also be feasible. To explore the reach of the endmember model, we construct and analyze a representative test data using our knowledge of Eq. (5.4) and Table 5.1. From this, we design a specific endmember-extraction (EE) strategy. We then use the predetermined and extracted $\{\boldsymbol{e}_i\}$ to unmix a real image via the FCLSU (see Sec. 3.1), to obtain the abundances $\{a_i\}$. Finally, we use the available ground-truth pixels to derive a model that will translate these $\{a_i\}$ to full concentration maps.

5.3 Endmember extraction

When considering automatic extraction of our water endmembers, it is clear that reference to Eq. (5.4) and Table 5.1 cannot be avoided if we are to identify the extracted $\{\boldsymbol{e}_i\}$ and thereby enable interpretation of the unmixing result. So we could take this knowledge upfront into the extraction process. Nevertheless, posterior matching of the pre-specified $\{\boldsymbol{e}_i\}(i = \overline{1,9})$ to a small subset of candidates extracted in a data-driven fashion can be a convenient way of dealing with the expected homogeneity of water spectra. Therefore, we opt for the use of unsupervised (data-exploratory) extraction techniques.

We ought to point out that, in all our further discussion, we will assume that the SIOP are known. These SIOP will in effect correspond to data obtained from in *situ* measurements, as detailed in Sec. 5.6.

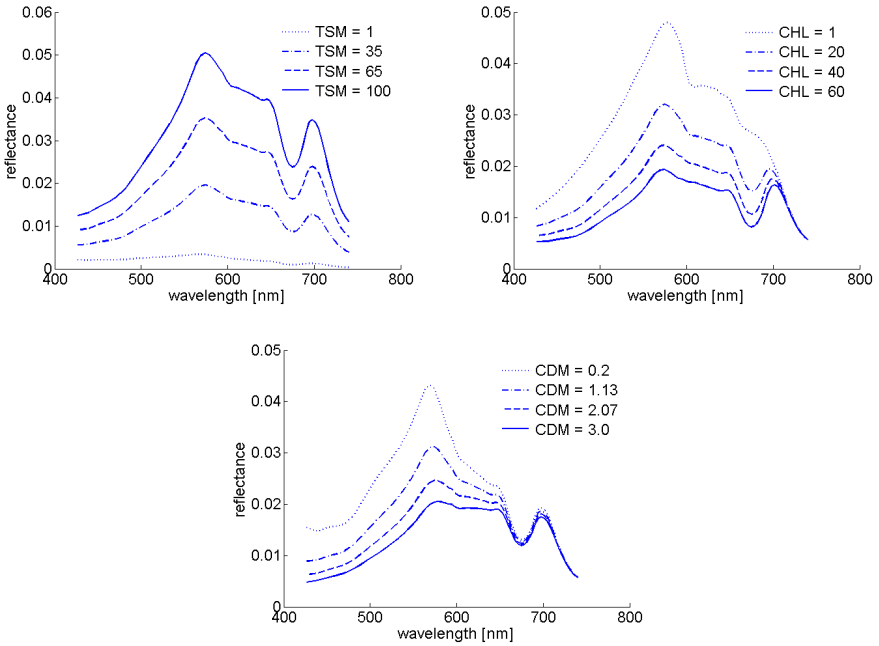

Figure 5.2: The effect of varying $c_i \in \{\texttt{TSM}, \texttt{CHL}, \texttt{CDOM}\}$ within its range defined by Table 5.1 while keeping $c_j, j \neq i$ constant (at their range average)

5.3.1 Assessment of the endmember model

The first question to ask is whether the $\{\boldsymbol{e}_i\}, (i = \overline{1,9})$ from Table 5.1 do in fact represent true (geometric) endmembers. A direct way to ascertain these $\{\boldsymbol{e}_i\}$ as the extremes of $\widehat{R}$ from Eq. (5.4) is to study this as a trivariate function of $\{\texttt{TSM}, \texttt{CHL}, \texttt{CDOM}\}$. However, the fact that $\widehat{R}$ is at the same time a function of the SIOP (which themselves depend on λ) makes this impractical. Still, by varying each c_i while fixing the others, we learn that different ranges of $\{c_i\}$ affect mostly distinct parts of $\widehat{R}$, which in turn implies that discerning not only the shape of a water spectrum but also its scale is relevant. Figure 5.2 gives an illustration. Notice that the visible wavelengths are also the most discerning, as all water spectra tend to converge in the NIR range Ruddick et al. [2006].

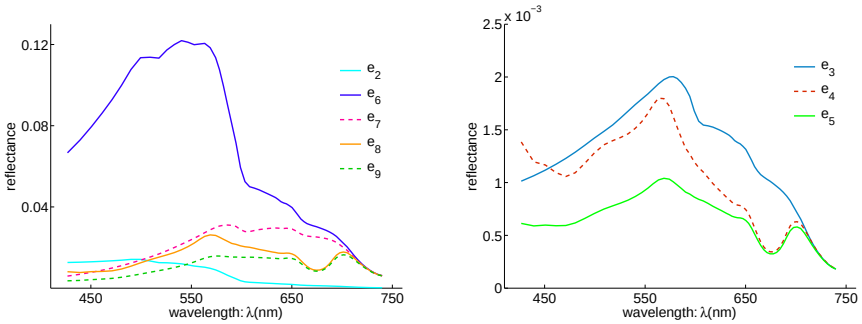

Figure 5.3: Endmembers yielded by the Gordon model when using values from Table 5.1 and in *situ* measured SIOP for our test image (e_1 is out of range and omitted).

Another and more empirical way to ascertain the $\{\boldsymbol{e}_i\}$ is to affirm them as being the vertices of a data-enclosing simplex. This is even more interesting since knowing Eq. (5.4) allows us to construct a perfectly representative test data. Namely, if Eq. (5.4) describes any water spectrum, and the SIOP and $\{c_{i,\text{low}}, c_{i,\text{high}}\}$ from Table 5.1 are known, then all possible spectra can be obtained by inserting into Eq. (5.4) random, uniformly distributed $c_i \sim U(c_{i,\text{low}}, c_{i,\text{high}})$. We use this notion to construct several sets counting from 10^3 to 20^3 samples, which practically corresponds to varying each c_i with steps between $5-10$ %. In addition, we account for perturbation in some of the SIOP ($a^*_{\texttt{CHL}}$, $a^*_{\texttt{TSM}}$, $b^*_{\texttt{TSM}}$) by shifting each randomly within a certain margin. We take these margins to be equal to the standard deviation (σ) from the mentioned in *situ* measurements, where the *signal-to-noise ratio* for the different SIOP (averaged over all wavelengths and computed as $SNR = 20\log_{10}(\mu/\sigma)$, with μ being the signal average) was found to reside between 10dB and 16dB.

To assess the $\{\boldsymbol{e}_i\}(i = \overline{1,9})$ as endmembers in this constructed data, several options can be pursued. One is to test the inwardness of the data relative to the p–simplex defined by $\{\boldsymbol{e}_i\}$. For this, we may compute the abundances $\{a_{ki}\}$ via the fractional volumes from Eq. (2.5)–(2.8) and check their additivity $\sum_i a_{ki} = 1$, after first projecting all k data points onto the simplex plane (see Sec. 3.2). Another possibility is to see which $\{\boldsymbol{e}_i\}$ can be identified by EE algorithms. One account we make from this testing is that most $\{\boldsymbol{e}_i\}$ were rightfully designated as endmembers, while $\boldsymbol{e}_3$, $\boldsymbol{e}_4$ and $\boldsymbol{e}_5$ gave more inconclusive results. Figure 5.3 shows these spectra, where we see that the three elusive endmembers are condensed at the lower end of the amplitude range and fairly resemble each other. It is now not difficult to imagine

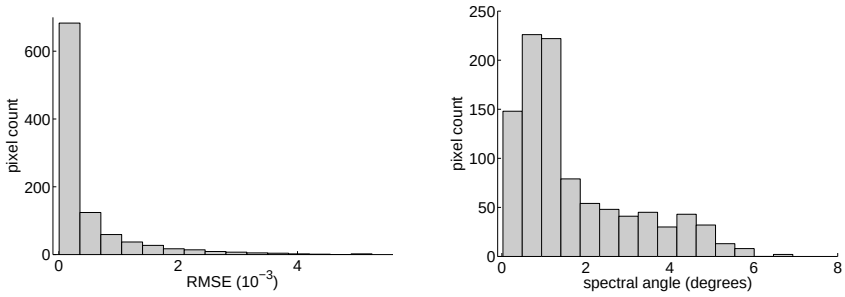

Figure 5.4: Reconstruction error for the endmember model on 1000 samples of the constructed (simulated) data.

how this could impede their discernment, depending on the order and depth to which they are sought as extremes in the data. One conclusion we draw from all this is that several $\boldsymbol{e}_i$ are likely to be indiscernible to typical unsupervised (geometric) EE algorithms when operating on real water content.

To examine the coverage of the endmember model, we unmix the constructed data using FCLSU and $\{\boldsymbol{e}_i\}$ from Table 5.1, and compute two common error measures – *root-mean-square error* (RMSE) and *spectral angle* Keshava and Mustard [2002], between the original and reconstructed spectra: $\{a_i \boldsymbol{e}_i\}$. Figure 5.4 shows this result. We infer that the majority of possible combinations of $c_i \in \{\texttt{TSM}, \texttt{CHL}, \texttt{CDOM}\}$ in Eq. (5.4)) is well reconstructed, with few apparently being 'out of reach' of the endmember model. This may in part be attributed to the use of unmixing constraints.

5.3.2 Feature extraction and clustering

Unlike with the simulated data, an indiscriminate use of unsupervised EE algorithms may yield poor detection of $\{\boldsymbol{e}_i\}$ in real (and large) water content, due to greater homogeneity of those spectra. We will alleviate this problem by reducing the search space for the EE algorithms. Inspired by our analysis from the previous section, we reckon that all water spectra exhibit limited shape variation that can be accommodated by grouping them into several clusters. One effect we expect from clustering is that it can set apart some $\{\boldsymbol{e}_i\}$ by making them appear at the boundaries of clusters.

Clustering of water spectra has been explored before for both classification and to define endmembers – as being the cluster centroids, where the input to this

clustering were typically the full spectra in their original, differentiated or normalized form (see Hommersom et al. [2011] and references therein). Similar use of clustering has been reported outside the water-retrieval context Dopido et al. [2011]; Glenn et al. [2013]. However, to us the main goal of clustering is different, namely to reduce the search space prior to EE.

We consider common clustering methods, such as *k-means* and others [Jain et al., 1999]. Instead of using the entire spectra, we will describe their shapes by extracting some fast-computable, and even more importantly, scale-invariant features. The shape invariance means that the features should reflect a spectral shape faithfully even if the spectra are offset, e.g. due to miscalculation of factor f from Eq. (5.1).

The first set of features we consider are the *Fourier Descriptors*:

$$r[k] = \sum_{n=1}^{N} R[n] e^{\frac{-i2\pi}{N}(k-1)(n-1)}, \ (k = \overline{1,N}); \quad f_i = \left| \frac{r[i+1]}{r[1]} \right|, \ (i = 1,2,3) \tag{5.6}$$

where $R[n] \equiv \widehat{R}(\lambda_n)$ is a water spectrum (water-leaving reflectance) and N is the number of wavelengths. The left-hand formula basically describes the discrete Fourier transform producing a signal r of the same length as the input signal R; yet, for meaningful discrimination of our spectra, and to be independent on the signal level, we take as features, $\{f_i\}$, the absolute values of few initial coefficients normalized by the first one. Another variable we consider is the *unsigned curvature*:

$$\kappa[n] = \frac{|\ddot{R}|}{(1+\dot{R}^2)^{3/2}}, \ f_4 = \frac{1}{N-1} \sum_{n=1}^{N} (\kappa[n] - \overline{\kappa})^2 \tag{5.7}$$

of which we take as feature its *standard deviation* as shown (the symbols $\dot{R}$ and $\ddot{R}$ stand for the first- and second derivative of R, and $\overline{\kappa}$ is the average of $\kappa[n]$). This way, we get a total of 4 features, which we take in their normalized form: $\|f_i\| = f_i / \max(f_i)$ to give each feature an equal weight.

We should say that this choice of features is certainly not exclusive, as other signal transforms, like *wavelets*, have been found useful in characterizing the water spectra Ampe et al. [2014]. However, where wavelets can be effective in capturing local shape variations Knaeps et al. [2010], our $\{f_i\}, (i = \overline{1,4})$ are adequate for discerning the entire spectra.

To test these features and clustering, we compute them from our constructed dataset described above. We validate the clustering using some common measures. One of those is the *silhouette index*, which measures heterogeneity and isolation of clusters by averaging a confidence score on the membership of each sample in particular cluster. Another is *Davies-Bouldin (DB) index* that identifies sets of

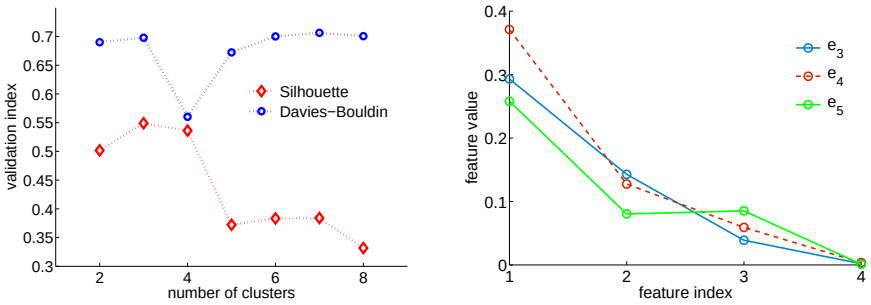

Figure 5.5: Feature extraction and clustering. *Left*: Cluster validation with simulated 15^3 spectra. *Right*: Features extracted for endmembers $\boldsymbol{e}_3$, $\boldsymbol{e}_4$ and $\boldsymbol{e}_5$.

compact and well-separated clusters by looking at intra- and inter-cluster distances. The exact mathematical formulation is found in e.g. [Bolshakova and Azuaje, 2002]. We compute these measures on the basis of multiple runs, with each run set up to use a same clustering algorithm (*k-means*) but a different number of clusters. The left-hand plot in Fig. 5.5 shows a typical result obtained when using all 4 features. We see that the iteration with 4 clusters produced a clear optimum: low *DB*- and high *silhouette* index [Bolshakova and Azuaje, 2002]. This tells us that the features are indeed capable of producing well-separated clusters.

An additional confirmation of this performance is given by Fig. 5.6, which shows random spectra from each cluster along with some of the target $\{\boldsymbol{e}_i\}$. Starting from the top, we see the 4 clusters obtained when using all 4 features. The bottom row shows another layer of clustering, i.e. two extra clusters obtained by subdividing two of the initial 4 clusters based on single features: f_1 and $f_5 = f_1 / f_2$. Specifically, the cluster from sub-figure indexed '31' (bottom left) is derived from the cluster '11' (top left) and the sub-cluster '32' from the cluster '21'. We see that several $\{\boldsymbol{e}_i\}$ are identifiable at the boundaries of these clusters, as we had also anticipated. Unfortunately, a comparison with the right-hand plot in Fig. 5.3 reveals that $\boldsymbol{e}_3$ and $\boldsymbol{e}_4$ remain fairly elusive. A comparison of their feature vectors at the right-hand side of Fig.5.5 give a more positive outlook, since at least $\boldsymbol{e}_3$ seems discernible from $\boldsymbol{e}_5$.

It is now clear that the result of this clustering – the cluster centroids (and the feature scaling factors $\{f_{i,\max}\}$) could be reused with other data, if that data will originate from water content with SIOP and $\{c_{i,\text{low}}, c_{i,\text{high}}\}$ residing in similar

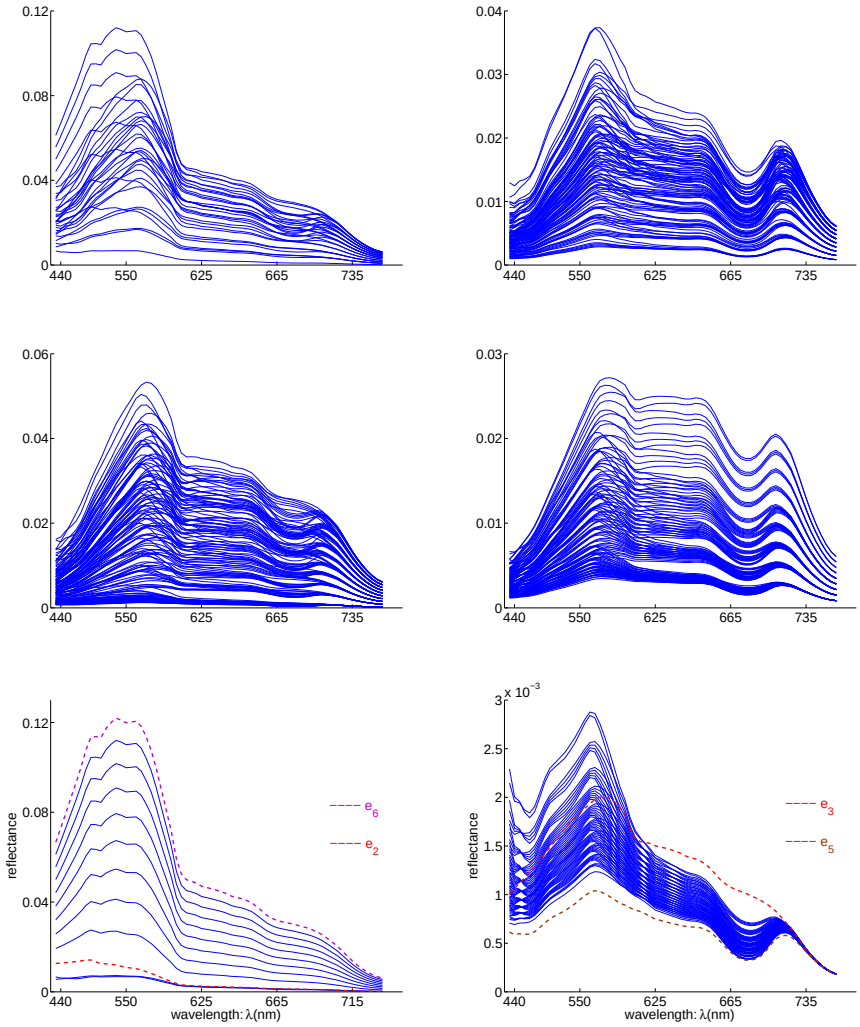

Figure 5.6: A result of multi-level clustering of 10^3 simulated spectra using the proposed feature-based representation. The bottom row shows two sub-clusters and the matching endmembers from Table 5.1. All spectra are slightly horizontally stretched due to alignment with indices of (non-uniform) spectral bands, i.e. wavelengths.

ranges. The benefit of such use is that each new pixel can be assigned to a cluster by minimum-distance comparison with few readily available points (the centroids), leading to fast, sequential processing.

5.3.3 Endmember extraction

For the actual endmember extraction, two strategies can be followed. Either we perform EE on each of the found clusters separately (cluster-level search), or alternatively we apply EE on the entire image, without using the clustering (image-level search). Doing both at the same time will yield an extended set of candidate endmembers that can be coalesced and sorted using several criteria. One of those criteria can be the ability of a subset of candidate endmembers to maximize the sum of their pairwise distances, as an indication of the largest simplex. Another can be the proximity of each candidate endmember to one of the predetermined cluster centroids or feature-based representations of $\{\boldsymbol{e}_i\}(i = \overline{1,9})$.

In our analysis of the simulated dataset in Sec. 5.3.1, we saw that the geometrical EE methods that search for extremes in the data were successful in detecting most of the $\{\boldsymbol{e}_i\}, (i = \overline{1,9})$. We also know that these $\{\boldsymbol{e}_i\}$ have been defined using the same Gordon model that is used to describe any other pixel. Therefore, for the EE step, we adopt some prevalent and accessible geometrical algorithms, such as N-findR Winter [1999], VCA Nascimento and Bioucas-Dias [2005b] and SMACC from ENVI [ENVI, 2009]. The imperfection of these algorithms in detecting some of the $\{\boldsymbol{e}_i\}$ should then be compensated by the clustering step.

It is worth noticing that, while we have intended the feature vectors as input for clustering, the same input could be used for the EE algorithms, now searching for fewer endmembers within the separate clusters. Going from spectra to features in a way realizes (nonlinear) dimensionality reduction, which could benefit the EE search. Still, one reason for keeping different input forms for the two operations (clustering and EE) is to promote the complementarity between them.

5.4 Translation of unmixing results

We consider unmixing of an image via FCLSU, with endmembers $\{\boldsymbol{e}_i\}$ specified by Table 5.1 or extracted from the image. This unmixing will yield abundances $\{a_{ki}\}$ for each water pixel k. Our goal is to translate the $\{a_{ki}\}$ to concentrations $\{c_{kj}\}$; however, to do this exactly (analytically) is far from trivial. Once more, this is seen when attempting to equate $\widehat{\boldsymbol{R}}$ expressed via Eq. (5.4) to $\boldsymbol{x}$ written out via Eq. (1.1), where the $\{\boldsymbol{e}_i\}$ are in turn also expressed via Eq. (5.4) using the same SIOP and some known values for $c_j \in \{\texttt{CHL}, \texttt{TSM}, \texttt{CDOM}\}$, like those from Table 5.1.

In Hommersom et al. [2011], a qualitative approach puts side-by-side the $\{a_{ki}\}$ and some $\{c_{kj}\}$ representing the in *situ* measured concentrations from a number of pixels. We propose a more quantitative approach and use the in *situ* measured values to derive a rule that will translate $\{a_i\}$ to $\{c_j\}$, then extrapolate this to the entire image. In other words, we seek a suitable combination of abundances that can yield a good fit to the ground truth, while also safeguarding generalization. We derive this combination by employing the classical *linear regression*.

Since not many ground-truth values are likely to be available, it is prudent to keep the model simple. For this, we make use of an intuitive reasoning that c_j should relate the most to $\{a_i\}$ from those $\{\boldsymbol{e}_i\}$ that include $c_{j,\text{high}}$ in their definition. We write: $\texttt{TSM} \sim f(a_j)$, $j \in \{6,7,8,9\}$, $\texttt{CHL} \sim g(a_k)$, $k \in \{4,5,8,9\}$ and $\texttt{CDOM} \sim h(a_l)$, $l \in \{3,5,7,9\}$, where f, g and h are some functions. The exact form of these functions is unknown, but since we are not using all $\{a_i\}$ $(i = \overline{1,9})$, we can reasonably demand it to connote more than a linear combination. Hence, we consider bilinear models, i.e. regression with the interaction terms:

$$\texttt{TSM} = \beta_{1,0} + \sum_{k\in\{6,7,8,9\}} \beta_{1,k}\,\boldsymbol{a}_k + \sum_{\substack{k,l\in\{6,7,8,9\}\\ k\neq l}} \beta_{1,kl}\,\boldsymbol{a}_k\,\boldsymbol{a}_l \tag{5.8}$$

$$\texttt{CHL} = \beta_{2,0} + \sum_{k\in\{4,5,8,9\}} \beta_{2,k}\,\boldsymbol{a}_k + \sum_{\substack{k,l\in\{4,5,8,9\}\\ k\neq l}} \beta_{2,kl}\,\boldsymbol{a}_k\,\boldsymbol{a}_l \tag{5.9}$$

where $\boldsymbol{c}_i = (c_{11}, \dots c_{1n}) \in \{\texttt{TSM}, \texttt{CHL}\}$ has the in *situ* measured c_i from n ground-truth pixels, $\{\boldsymbol{a}_i\}$ are the $n \times 1$ vectors of abundances a_i from each of those pixels, and $\beta_{1,\{.\}}$ are to be determined weights or *regression coefficients*. The same is applicable to CDOM after replacing a_4 with a_3 and a_8 with a_7 in Eq. (5.9).

It is worth pointing out that the whole exercise has more of *approximation* than of *learning* from data an entirely unknown *target function* [Abu-Mostafa et al., 2012]. The reason for this is that we've been able to reasonably stipulate the target and our hypothesis in Eq. (5.8)-(5.9) using our knowledge and comparison of the water-reflectance and endmember models. Hence, if the endmember model can yield good reconstruction of spectra for most pixels, and we get a low *in-sample* error by fitting Eq. (5.8)-(5.9) on a number of those, then by induction we should expect a low *out-of-sample* error from applying that fit to the rest of the image. This is also why we foresee a limited role for some practices from machine learning, like *regularization* – which we could use to constrain $\{\beta_{(.)}\}$, or *cross-validation* – which we could use to select one from multiple sets of $\{\beta_{(.)}\}$ trained of different subsets of the ground truth (as generated by e.g. the 'leave-one-out' method).

5.5 Experiments

This section describes our experiments with real hyperspectral data. To keep our objectives straight, we structure these experiments in two parts. After describing the image, we first discuss its unmixing with the prespecified endmembers from Table 5.1. This will include the step of translating the abundances to actual concentration maps and comparing those with a reference obtained from curve matching. In the second part, we test the ability of our EE approach to extract matching endmembers from the image and to produce similar result.

5.5.1 Description of the hyperspectral data

We consider an APEX image acquired over the Wadden Sea area in the Netherlands, in June, 2011. APEX is developed by a Swiss-Belgian consortium on behalf of ESA and is intended as a simulation, calibration and validation device for spaceborne imagers. The APEX sensors record data in about 300 (non-uniformly spaced) bands covering the range of 380-2500 nm Itten et al. [2008]. The radiometric, spectral and geometric calibration is performed by the Calibration Home Base (CHB) hosted at DLR Oberpfaffenhofen, Germany Gege et al. [2009]. The atmospheric and air-interface correction of the acquired data was in this case done with the MODTRAN-4 radiative transfer code following the algorithms and their implementation described in Haan de and Kokke [1996], Sterckx et al. [2011]. Residual sky glint was corrected by subtracting the reflectance remaining at 1203 nm.

The Wadden Sea area is regarded as suitable for studying water composition, because it includes coastal and shallow waters that are known to be subject to land, oceanic and human influences Hommersom et al. [2011]. The Marsdiep, the westermost tidal inlet of the Wadden Sea, is particularly suited for our purpose, because it contains different water types at short distance. Here, marine waters from the open sea in the west mix with freshwater originating from a channel in the south and a highly productive freshwater lake in the east. Figure 5.7 shows an extended view of the scene, where it is seen that the image is registered as a mosaic of multiple flight lines. Next to this image is a zoomed-in picture of a region of interest (ROI) that we select for unmixing.

The ROI was selected mostly to get a more workable image size, while including all the pixels bearing the results from in *situ* measurements. Another concern was to include both the deep main channel as well as the shallow waters near the coast, as a way to increase the chance of finding most of the $\{\boldsymbol{e}_i\}$, i.e. pixels with concentrations approaching those from Table 5.1. The spatial size of the ROI amounts around 1000×1000 pixels, with the pixel size 5.4m, while the spectral information has been

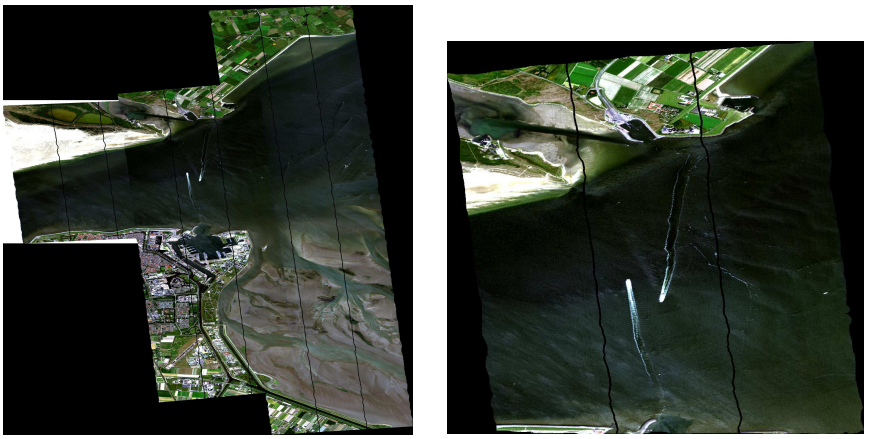

Figure 5.7: RGB view of the test data: an extended view of the scene (left) and the ROI used for unmixing (right).

confined to 65 bands in the 410-750nm range. Notice that the ROI image includes surfaces other than water, which we used as an extra reference for masking out non-water-like pixels.

The SIOP and concentrations were measured for TSM, CHL and CDOM from water samples collected at the pontoon of NIOZ. From the water samples, TSM concentration was determined by filtering the water on Whatman GF/F glass fiber filters according to the European reference method *EN*872(2005). Pigment analysis, including CHL-a concentrations were determined via HPLC. The measured values resided in the following ranges: $\texttt{TSM}\,(\mathrm{gm}^{-3}) \in [5.3, 10.2]$, $\texttt{CHL}\,(\mathrm{mgm}^{-3}) \in [5.1, 10.4]$.

The specific absorption spectra of non-algae particles and Chlorophyll were measured using the filter pad method using a LICOR integrating sphere attached to an ASD spectrometer following the methods described by [Tassan and Ferrari, 1995] and REVAMP protocols [Tilstone et al., 2003]. For the CDOM absorption, the water samples from the field campaign were temporarily stored in a cooled chamber and filtered through 0.2μm pore size filters. To retrieve the CDOM absorption coefficient of the water samples, the beam attenuation of the filtered water was measured with Ocean Optics equipment in a transparent cuvet. As the data from the ocean optics equipment are noisy for wavelengths $< 400nm$ and $> 950nm$ the exponential shape of the CDOM absorption was fitted based on the $420-750nm$ range. More details about the concentration and SIOP measurement is given in [Knaeps et al., 2012].

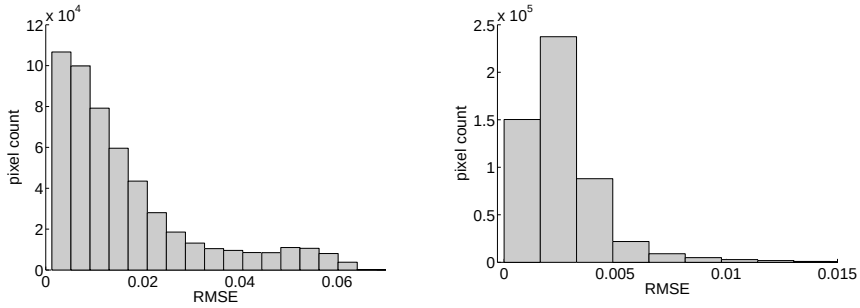

Figure 5.8: Reconstruction error of unmixing (FCLSU) the water spectra in the ROI image using the preset $\{\boldsymbol{e}_i\}(i=\overline{1,9})$ (*left*) and the extracted $\{\boldsymbol{e}_j\}$ from Fig. 5.13 (*right*)

Overall, there was a satisfactory resemblance between the in *situ* measured spectra and those read from the corresponding pixels in the APEX image. The mean and standard deviation ($\mu_x \pm \sigma_x^2$) of two prevalent error measures – RMSE and *spectral angle* computed between these spectra for 10 reference pixels were, respectively, 0.0059 ± 0.0028 and 10.7 ± 2.7 degrees. The same comparison between the in *situ* spectra and those obtained by inserting the in *situ* measured concentrations and SIOP into Eq. (5.4) were 0.0054 ± 0.0034 and 10.5 ± 3.8. We should say that few relative outliers were responsible for most of the deviation. The factor f required by Eq. (5.1) was also used and estimated at $f = 0.55$.

5.5.2 Unmixing with predefined endmembers

We start by unmixing the ROI image using FCLSU with all $\{\boldsymbol{e}_i\}, (i=\overline{1,9})$ from Table 5.1. This yields abundances $\{a_{ki}\}$ for each pixel k. First, we test the ability of the endmember model to represent the water spectra, which is verified by a low reconstruction error as shown in Fig. 5.8 on the left.

Our main goal is to derive concentration maps, and since we are most interested in CHL and TSM, we leave CDOM (for which no reference, curve-matching map was also available) out of the analysis. Figure 5.9 shows the six abundances of interest and the question is thus what models according to Eq. (5.8)-(5.9) should combine them. Since in our case $n = 10$, finding those models is prone to over-fitting. After some trials we retained these reduced sets of weights:

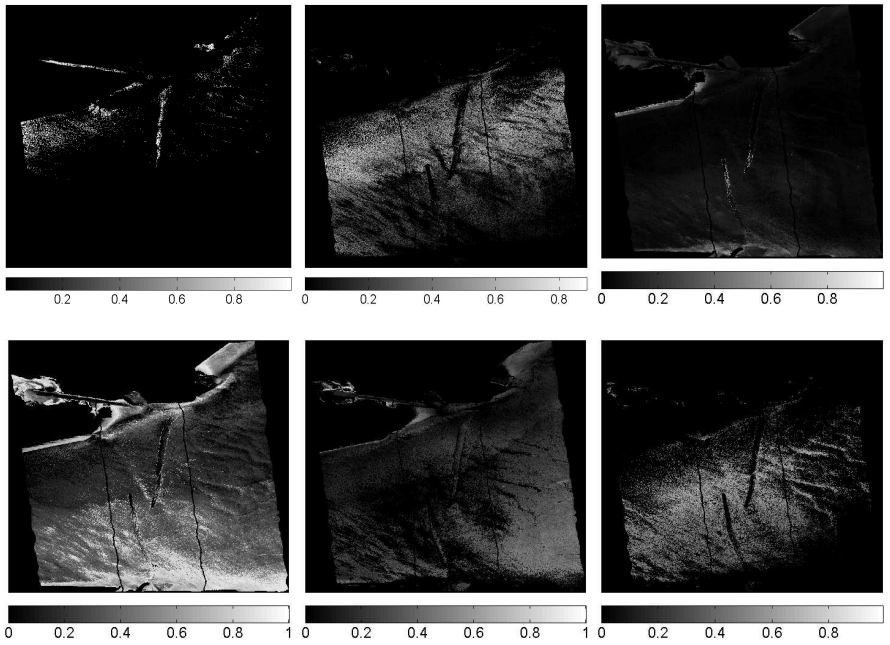

Figure 5.9: From left to right and top to bottom: abundance maps $a_4, a_5, \ldots a_8, a_9$ obtained from unmixing the ROI image with $\{\boldsymbol{e}_i\}$, $(i = 1, \ldots 9)$ from Table 5.1.

$$\begin{aligned} \beta_1 &= (\beta_{1,0}, \beta_{1,6}, \beta_{1,7}, \beta_{1,8}, \beta_{1,\{6,7\}}, \beta_{1,\{6,8\}}) \\ \beta_2 &= (\beta_{2,0}, \beta_{2,5}, \beta_{2,8}, \beta_{2,\{5,8\}}, \beta_{2,\{5,9\}}) \end{aligned} \tag{5.10}$$

where in the case of TSM we got a high *coefficient of determination*, $R^2 = 0.85$. This was notably lower with CHL ($R^2 = 0.6$), which can be ascribed to the fact that in all the ground-truth pixels there was a limited representation from a_5 and none from a_4. As for the generalization ability of these models, some idea is gotten from the unmixing statistics depicted on the right-hand side of Fig. 5.12, where we see that it took 5-6 endmembers to reconstruct most pixels, with $\{\boldsymbol{e}_3, \boldsymbol{e}_6\}$ being most underrepresented. Refer also to what was explained at the end of Sec. 5.4.

Figures 5.10 – 5.11 show the obtained concentration maps. As a comparison,

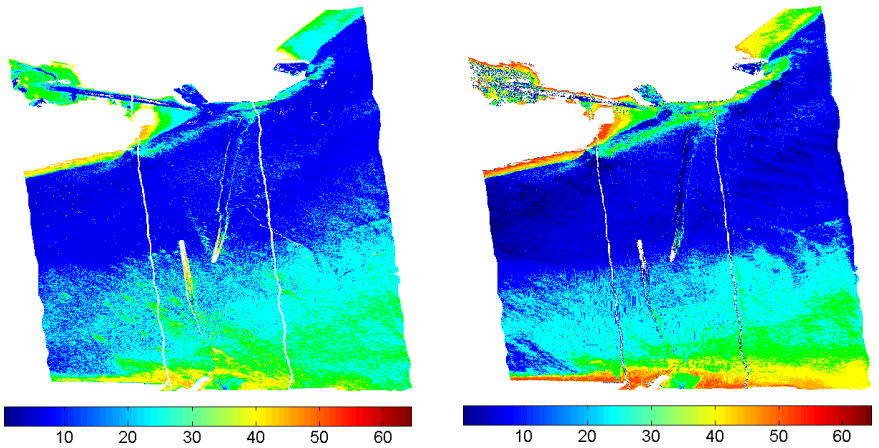

Figure 5.10: Concentrations TSM (g/m^3) obtained by combining the abundances $\{a_i\}$ from unmixing (*left*), and by the curve matching (*right*).

maps produced by the curve-matching method are shown as well. It is worthwhile pointing out that the number of ground-truth pixels can be artificially increased, by taking the values of the curve-matching method from their neighborhood and assuming those to be equal to the in-situ measured values. When doing so, similar maps were obtained. While it is clear that we had to be restrained with this ground-truth extension (or we would be fitting our results to the reference), the truth is that exaggerating it could provide an extra support for our argument about the coverage by the endmember model. We saw that combining the $\{\boldsymbol{e}_i\}$ $(i = \overline{1,9})$ *linearly* could well reconstruct most of the possible water spectra; now, we would see that a properly derived (still simple, bilinear) combination of the corresponding abundances could retrieve most of the possible concentrations.

5.5.3 Endmember extraction and unmixing

When extracting the $\{\boldsymbol{e}_i\}$ $(i = \overline{1,9})$ from real water content, one has to count in the likely event that not all of them will actually be present. As a sanity check, we perform an exhaustive search for these $\{\boldsymbol{e}_i\}$ in our image using full spectra and two error measures that we used previously. The left-hand side of Fig. 5.12 shows this result, from which we see that pure water was practically absent, while $\boldsymbol{e}_2$ and $\boldsymbol{e}_6$ give an

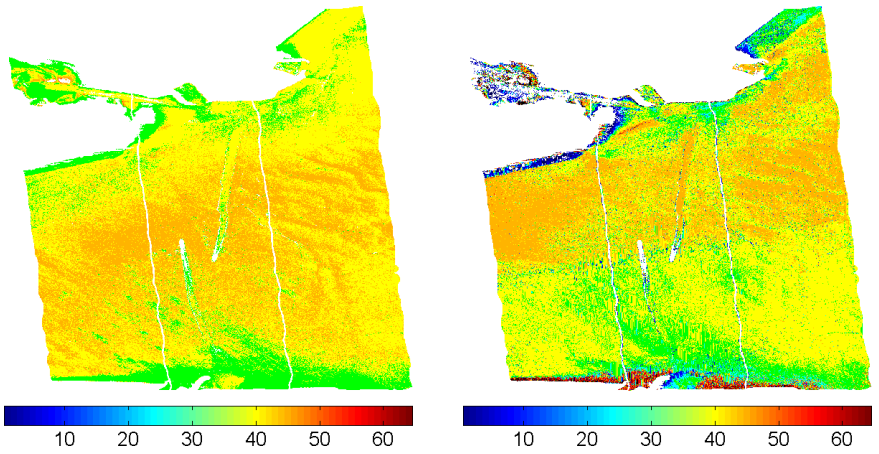

Figure 5.11: The same comparison as in Fig. 5.10 for CHL (mg/m^3).

inconclusive picture. It is interesting that a single pixel gave the lowest RMSE for both $\boldsymbol{e}_4$ and $\boldsymbol{e}_5$. To get additional hints, we also look at the previously mentioned usage statistics of these $\{\boldsymbol{e}_i\}$, as shown in Fig. 5.12 on the right.

The goal of the proposed clustering was to make a distinction between spectral shapes that may relate to the target endmembers. Instead of clustering the real dataset, we re-used the cluster centroids obtained from clustering the simulated dataset. After computing the features, the assignment of pixels to these centroids did not produce the anticipated 4 clusters; in fact, only 3 properly filled clusters were obtained. This suggests that the water spectra constituting the image did not exhibit all possible variety that we could theoretically expect based on our sampling of the Gordon model. Still, the second level of clustering did prove useful in spotting $\boldsymbol{e}_6$ and $\boldsymbol{e}_5$ at the sub-cluster boundaries, the same way this was illustrated in Fig. 5.6.

For the EE-step, two strategies explained in Sec. 5.3.3 were followed: cluster-based EE and an image-level search in which no clustering is applied prior the EE-step. At the image level, we settled with searching for 7-8 endmembers. Per cluster, we set the target number of endmembers between 2 and 4. The entire process was then generating up to 19 endmember candidates, which we sorted as explained in Sec. 5.3.3.

The final result is depicted in Fig. 5.13. Note that $\boldsymbol{e}_2$ is merely shown and not

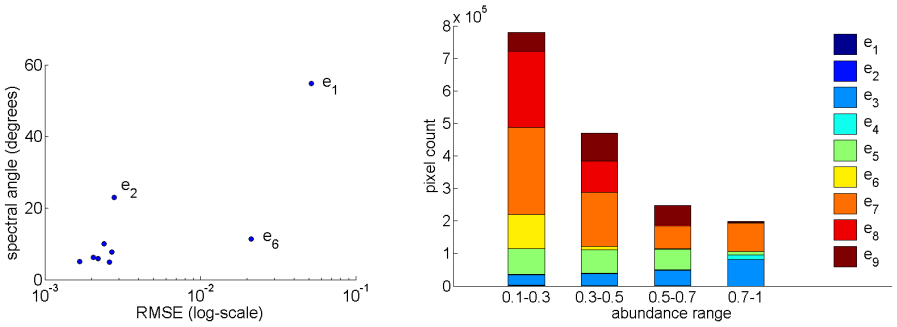

Figure 5.12: Endmember identifiability: deviation of $\{\boldsymbol{e}_i\}$ from Table 5.1 from their closest match found by a full search (*left*) and statistics of unmixing with these $\{\boldsymbol{e}_i\}$.

used in analysis afterwards, while $\boldsymbol{e}_4$ was not extracted, which would influence the derivation of a concentration map for CHL. While the plots show mostly the results of the combined cluster- and image-level search, we should say that good approximations for $\boldsymbol{e}_2$, $\boldsymbol{e}_8$ and $\boldsymbol{e}_9$ were also obtained from the image-level search alone. This was not true for $\boldsymbol{e}_7$ whose proper match was only found thanks to the use of clustering. By the look of Fig. 5.13, the best estimates for $\boldsymbol{e}_5$ and $\boldsymbol{e}_6$ – in terms of lowering the RMSE and spectral angle, are likely gotten by averaging the output from the cluster- and image-level search. While one may argue that the extracted $\boldsymbol{e}_5$ deviates intolerably from the tabulated one, assigning it was necessary to derive a decent-looking map for CHL following Eq. (5.9)-(5.10).

Finally, we derive the concentrations with the extracted $\{\boldsymbol{e}_i\}$ and compare them with those we had derived earlier with the preset $\{\boldsymbol{e}_i\}(i = \overline{1,9})$. First, we compare the reconstruction errors obtained with the two sets of endmembers, which are shown in Fig. 5.8 on the right. We see that the RMSE of the extracted $\{\boldsymbol{e}_i\}$ is lower, which can be explained by the relative homogeneity of the image spectra that was already evident after the clustering step. Figure 5.14 shows the resulting concentration maps. The maps are somewhat different from those in Fig. 5.10–5.11 but the global trends remain largely similar. Because of the lack of $\boldsymbol{e}_4$ in the analysis, the map obtained for CHL deviates more from the curve-matching reference.

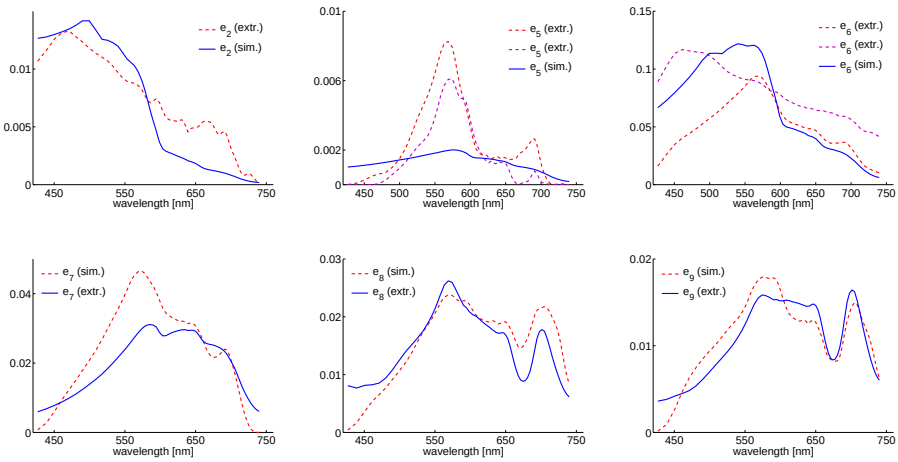

Figure 5.13: Results of EE. The full lines depict $\{e_i\}$ from Table 5.1 and the dashed lines the matching spectra extracted from the image. For e_5 and e_6, the dotted lines show the results from the image-level search.

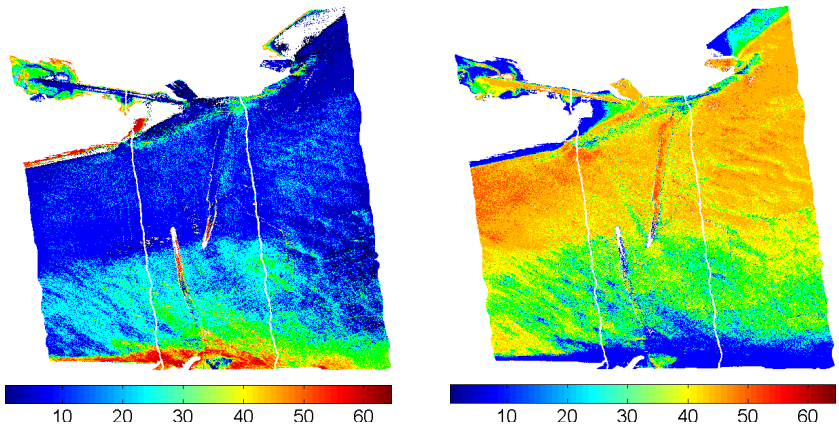

Figure 5.14: The counterparts to TSM from Fig. 5.10 (*left*) and CHL from Fig. 5.11 (*right*) obtained with extracted endmembers.

5.6 Conclusion

We have described a method for studying the water quality by means of spectral unmixing. One of the main goals of our investigation has been to reaffirm unmixing as a viable method of estimating the water composition when the endmembers, specifying different water types, are known. To this end, we have designed and performed an extensive analysis of simulated spectra, and conducted a comparison on real hyperspectral image using as reference a state-of-the-art, curve-matching method. One of our important conclusions is that linear unmixing with compounded endmembers defined from the water-reflectance model can amply cover the variability of water spectra, thus making it less pressing to use nonlinear unmixing models.

To further advance the unmixing take, we have proposed methods to extract the endmembers from the reflectance image and translate the endmember abundances to actual concentrations. The viability of this route has been confirmed by an acceptable resemblance between the unmixing-derived and reference concentration maps, as well as by impacting similarity between the extracted and prespecified endmembers. The proposed method for endmember extraction (EE) exploits a feature-based representation of the water spectra in combination with common data clustering and EE techniques. Several of these components by design exploit knowledge of the water-reflectance model and values of its parameters, such as the extreme concentrations of water constituents and their SIOP, while allowing some perturbation in those. This means that extendibility of the used methodology to other water content will depend on the properties and ranges of the said parameters. The regression model used to translate the abundances to concentrations has been derived from a limited ground truth, yet, the reasoning that led to its build up is also generally valid.

Remarks

We are indebted to the Flemish Institute for Technological Research (VITO) for the acquisition and preprocessing of the dataset. This work was partly funded by the project CHAMELEON of the Flemish Agency for Innovation by Science and Technology (IWT), and by the project IN PLACE of the Netherlands Organisation for Scientific Research (NWO).

PART III

SUMMARY

CHAPTER 6

SUMMARY

Remote sensing technology has advanced tremendously in recent decades. An important economic and strategic driver for this development has been the offering of wide spatial and temporal coverage by space- and airborne platforms, as well as the ever-increasing ability of sensors to record images with high spatial and spectral resolution. This has made remote sensing an attractive tool for collecting and interpreting data from the Earth's surface and atmosphere for various uses.

A modality that provides a bulk of data for remote sensing applications is hyperspectral imaging. This modality records the reflected solar radiation in contiguous and often numerous wavelengths or spectral bands. It thereby extends the standard photography, by enabling to treat each pixel individually as a spectrum discernible for each class of materials. One of the limitations of such imaging, where the spatial and spectral resolution are inescapably traded against one another, is the occurrence of mixed pixels and spectral mixing. Hence, the unraveling of spectral mixtures has been widely studied as spectral unmixing, where two main aspects are of interest: the estimation of the constituent spectra or endmembers, and the determination of their proportions or abundances, in the mixture. The work described in this thesis regards spectral unmixing from two objectives: advancement of unmixing methodology, and introduction of unmixing in new applications. This has motivated the structuring of the theses into two parts, each focusing on one of these two aspects.

Their common perspective has been to opt for data-driven approaches that can mitigate the dependency on physical parameters and models.

Part I describes two distinctive approaches and algorithms for unmixing with a given number of endmembers. The central theme here has been reformulation of the prevalent geometric and optimization-based treatments of the linear mixture model (LMM), by introducing distance-based and analytical formulations. Experiments with synthetic and real hyperspectral data indicate that both algorithms can notably improve the time- and memory efficiency compared to benchmark methods, as well as enable alternative (nonlinear) treatments of mixtures – by incorporating non-Euclidean distance measures. Another alternative treatment connoting a streaming implementation of endmember extraction has resulted from the same framework and is cited but not detailed in the thesis. This work should encourage investigation of a similar streaming-based computation of abundances, such that entire unmixing may be done in single pass and without pre-loading the image. This could involve the use of online optimization and statistical techniques, or more supervised unmixing methods, like e.g. (sparse) regression.

Chapter 2 describes a data-driven algorithm that estimates endmembers and their abundances under the nonlinear mixing assumption. The algorithm does so by searching for the largest data-enclosing simplex and subsequently computing the volumes of fractional sub-simplices for each data point, while expressing the volumes in terms of inter-point geodesic distances approximated by the shortest paths on a kNN graph constructed from the data. This approach virtually presents a computationally efficient equivalent to letting the linear unmixing be preceded by nonlinear (geodesic-based) dimensionality reduction. The main drawback is that the quasi-abundances derived from the 'geodesic volumes' lose their direct interpretatability as physical quantities, i.e. fractions of endmembers in a pixel. On the upside, they are not restricted to particular type of (nonlinear) surface. Future research should explore whether some relation can still be established between these quasi-abundances and those from a more explicit, e.g. bilinear, model. Another extension can be to also try other distance measures that may better reflect the data dependencies than the geodesic distance in particular situation (e.g. scene type).

Chapter 3 describes an algorithm that substitutes the conventional optimization-based computation of abundances with known endmembers (FCLSU) by an analytical solution. This is achieved by exploiting the demonstrated equivalence of solving the fully constrained least-squares problem and geometric projection of a point onto a simplex. The SPU algorithm introduces several geometrical properties for simplices to accomplish the latter. While these properties are not universally valid, the counterexamples typically concern skewed geometries induced by strong mutual proximity of vertices of a simplex, i.e. endmembers (which may imply their

mis-estimation). At the same time, the analytical nature of the SPU algorithm allows computing the abundances in a severely shortened amount of time. Investigating potential uses for the SPU algorithm outside the spectral-unmixing context may therefore be an interesting option.

Part II discusses unmixing in the context of two atypical applications: the detection of adjacency effect and the estimation of water quality in inland- and coastal waters. To introduce an unmixing approach in these applications, a same formula is followed whereby the use of particular mixing model(s) is theoretically motivated, possibly complemented with additional constructs and validated, through comparison with an available reference method. This includes quantitative comparisons on real hyperspectral images. An overall conclusion from all the comparisons is that the results of the said reference methods can be amply approached by those from the proposed unmixing-based treatments.

Chapter 4 describes an unmixing approach to detect the adjacency effect, which occurs when reflectances from a surface of interest (target) and its adjacent neighborhood superimpose on their path to the sensor due to atmospheric (back)scatter. The central idea of the approach is then to use unmixing to separate the true signature of a pixel on a target from the polluting scatter reflected from its larger neighborhood. As the test case, the often impacting adjacency effect in turbid waters surrounded by vegetated land has been studied. It has been demonstrated that a reference detection by a spectral-index ratio designed specifically to exploit a prior knowledge about the target (water) can be equally formulated as a case of supervised, unconstrained linear unmixing. It has also been shown that an unsupervised and fully constrained unmixing can produce a roughly scaled version of these results, if proper water and vegetation endmembers are applied. Unfortunately, the scaling also means the unsupervised use of unmixing is less suitable for the actual correction of the effect. Since the adjacency effect occurs mostly when the target pixel reflectance is lower that of its surroundings, it would be interesting to see whether unsupervised unmixing (using data dimensionality reduction) could still be useful for detection on other types of dark surfaces with relatively bright surroundings.

Chapter 5 elaborates and extends a recently introduced unmixing approach for estimating the quality of inland- and coastal waters. The basic of this estimation is the inversion of the imaged water spectrum via some model, as a way of retrieving the unknown concentrations of optically active impurities or constituents: CDOM, TSM and chlorophyll-a. A typical way of doing this and the reference to the study is curve matching, which compares each water spectrum to a library of spectra constructed by inserting the known (e.g. in *situ* measured) concentrations of each constituent into some water-reflectance model. The unmixing approach replaces this by performing fully-constrained linear unmixing of endmembers defined by

inserting into the water-reflectance model extreme concentrations of each constituent – in a way, this connotes linear unmixing with 'nonlinear endmembers'. To translate the endmember abundances to actual concentrations, a suitable model has been proposed and derived by regression on a limited number of ground-truth pixels. To further advance the unmixing take, a method has been proposed to extract the endmembers from the image by combining common data clustering and endmember-extraction techniques. An important conclusion that emerges from the experiments is that the adopted endmember model can amply cover the variability of water spectra, making it less pressing to use nonlinear unmixing models. Still, one fundamental questions that can be asked is whether mathematically more elaborate water-reflectance models can be decomposed as a mixing equation, such that abundances and concentrations can be related more directly. Another aspect is that images with ample numbers of ground-truth pixels are much desired prerequisite for proper use of machine-learning techniques and explicit optimization towards out-of-sample performance.

SAMENVATTING

Remote sensing technologie heeft in de afgelopen decennia enorm gevorderd. Een belangrijke economische en strategische drijfveer voor deze ontwikkeling is het aanbieden van uitgebreide ruimtelijke en temporele dekking door satellieten en vliegtuigen, evenals het steeds toenemende vermogen van sensoren om beelden met een hoge spectrale en spatiële resolutie op te nemen. Hierdoor is remote sensing een aantrekkelijk middel geworden voor het verzamelen en interpreteren van gegevens van het oppervlak en de atmosfeer van de Aarde voor verschillende doeleinden.

Een modaliteit die een grote hoeveelheid gegevens voor remote sensing aanlevert is hyperspectrale beeldvorming. Deze modaliteit registreert de gereflecteerde zonnestraling in aaneengesloten en vaak talrijke golflengten, oftewel spectrale banden. Door tevens een spatiële dekking te realiseren, breidt het standaard fotografie uit, door elke pixel individueel behandelbaar te maken als een gedetailleerd spectrum dat onderscheidend is voor ieder type materiaal in het beeld. Een van de beperkingen van dergelijk beeld, waarbij de spatiële en spectrale resolutie onontkoombaar tegen elkaar worden uitgeruild, is het optreden van gemengde pixels en spectrale menging. Vandaar dat het ontrafelen van spectrale mengsels uitgebreid is bestudeerd als spectrale ontmenging, waarbij vooral twee aspecten van belang zijn: de schatting van de samenstellende spectra oftewel endmembers, en de bepaling van hun fracties oftewel abundanties, in het mengsel. Het in dit proefschrift beschreven onderzoek benadert spectrale ontmenging vanuit twee doelstellingen: bevorderen van methodologie en de invoering van ontmenging in nieuwe toepassingen. Dit verklaart de indeling van de tekst in twee delen, waarbij elk gericht is op een van deeze twee aspecten. Hun gemeenschappelijke perspectief is om data gedreven verwerking te bevorderen die de afhankelijkheid van fysische modellen, of anders het typische gebruik van optimalisatietechnieken, kan beperken.

Deel I beschrijft twee verschillende algoritmen voor ontmenging met een gekend aantal endmembers. Het centrale thema is herformulering van de heersende geometrische en optimalisatie-gerichte behandeling van het lineaire mengingsmodel

(LMM), door de invoering van afstands-gerelateerde en analytische uitdrukkingen. Experimenten met synthetische en echte hyperspectrale beelden laten zien dat beide algoritmen computationeel veel efficiënter kunnen werken dan de referentie-methoden, terwijl ze ook alternatieve (niet-lineaire) behandelingen van spectrale mengsels mogelijk maken – door de opname van niet-euclidische afstandsmaten. Een andere alternatieve benadering die uit hetzelfde aanpak is ontstaan, die de schatting van endmembers via een sequentiële eerder dan een batch verwerking realiseert, wordt in het proefschrift aangehaald maar niet gedetailleerd. Dit werk zou verder onderzoek moeten inspireren naar een vergelijkbare sequentiële schatting van abundanties, zodanig dat volledige ontmenging in real-time en zonder het vooraf inladen van het hele beeld kan geschieden. Het is hierbij denkbaar om gebruik te maken van online-optimalisatie technieken, of voorkeur te geven aan de meer gesuperviseerde methodes voor ontmenging, zoals bv. de recent ontwikkelde 'schaarse regressie'.

Hoofdstuk 2 beschrijft een data gedreven algoritme dat endmembers en hun abundanties schat onder de veronderstelling dat spectrale menging niet-lineair is. Het algoritme zoekt effectievelijk naar de grootste data omsluitende simplex en berekent de volumes van deelsimplexen voor elk datapunt, waarbij de volumes worden uitgedrukt in geodetische afstanden benaderd door de kortste paden op een uit de data opgebouwde *kNN graaf*. Dit aanpak presenteert vrijwel een computationeel efficiënt equivalent van het laten voorafgaan van lineaire ontmenging door niet-lineaire dimensionaliteitsreductie (gebruikmakend van dezelfde afstandsmaat). Het grootste nadeel van dit aanpak is dat de quasi-abundanties die via de 'geodetische volumes' worden berekend hun uitdrukkelijke interpreteerbaarheid als fracties van endmembers in een pixel verliezen. Aan de andere kant hebben zij het voordeel om niet beperkt te zijn tot een bepaald soort (niet-lineair) oppervlak. Toekomstig onderzoek moet nagegaan of er een relatie kan worden gelegd tussen deze quasi-abundanties en parameters van een explicieter, bv. bilineair, ontmengingsmodel. Nog een uitbreiding kan zijn om ook andere afstandsmaten toe te passen die de data afhankelijkheden en mogelijke correlaties in bepaalde typen beelden beter dan de geodeten kunnen weerspiegelen.

Hoofdstuk 3 beschrijft een algoritme dat de heersende optimalisatie-gebaseerde berekening van abundanties in een mengsel met gekende endmembers vervangt door een analytische oplossing. Deze oplossing wordt behaald door de gelijkwaardigheid te benutten tussen het oplossen van het volledig beperkte kleinste-kwadraten probleem en geometrische projectie van een datapunt op een simplex. De SPU algoritme introduceert enkele geometrische eigenschappen voor simplexen om het laatstgenoemde te bewerkstelligen. Hoewel deze eigenschappen niet universeel geldig zijn, betreffen de tegenvoorbeelden in de regel afwijkende geometrieën,

waarbij de hoekpunten van de simplex, i.e. de endmembers, erg dicht bij elkaar liggen (wat op hun foutieve schatting kan duiden). Anderzijds maakt de analytische aard van het SPU algoritme dat de berekening van abundanties veel sneller wordt. Deze snelheid rechtvaardigt een zoektocht naar andere potentiële toepassingen voor het SPU algoritme buiten de spectrale ontmengingscontext, die eveneens beroep zouden kunnen doen op simplex projectie.

Deel II bespreekt spectrale ontmenging in het kader van twee atypische toepassingen: de detectie van nabijheids-effect en de schatting van kwaliteit oftewel samenstelling van binnenlandse en kustwateren. Om spectrale ontmenging in deze toepassingen te introduceren is dezelfde procedure gevolgd, waarbij een of meerdere ontmengingsmodellen theoretisch gemotiveerd zijn en vervolgens, eventueel aangevuld met extra constructies, vergeleken met gegeven referentiemethodes. Deze vergelijking omvat experimenten uitgevoerd op echte hyperspectrale beelden. Een algemene conclusie die uit dit werk naar voren is gekomen is dat de prestaties van de vastgestelde referentiemethoden ruimschoots benaderd kunnen worden door die van de voorgestelde, op ontmenging gebaseerde methoden.

Hoofdstuk 4 beschrijft een inzet van spectrale ontmenging voor het opsporen van nabijheids-effect. Dit effect treedt op wanneer de reflecties van een doeloppervlak en haar aangrenzende omgeving superponeren binnen het gezichtsveld van de sensor als gevolg van atmosferische verstrooiing. De kerngedachte van het voorgestelde aanpak is dan om ontmenging te gebruiken om de ware spectrale signatuur van een doelpixel te scheiden van de component die afkomstig is van haar bredere omgeving. Als een proefproces is de nabijheids-effect in troebele wateren omgeven door begroeide land onderzocht. Er is aangetoond dat een spectraal indexverhouding, dat specifiek is ontworpen om een voorkennis over het doeloppervlak (water) te benutten, eveneens kan worden uitgedrukt als geval van een gesuperviseerde, niet beperkte lineaire ontmenging. Ook is aangetoond dat een ongesuperviseerde en volledig beperkte lineaire ontmenging een ongeveer geschaalde versie van deze resultaten kan geven, als de juiste water en vegetatie endmembers worden toegepast. Aangezien de nabijheids-effect optreedt wanneer de reflectantie van een doeloppervlak lager is dan die van haar omgeving, zal het interessant zijn om te onderzoeken of het ontmengingsaanpak (gebruikmakend van dimensionaliteitsreductie) ook toegepast kan worden op andere donkere oppervlakken met relatief lichte omgeving.

Hoofdstuk 5 bestudeert en breidt uit een onlangs geïntroduceerde ontmengingsbenadering voor de bepaling van kwaliteit van binnenlandse en kustwateren. De hoofdgedachte is om water spectra te relateren aan concentraties van enkele optisch actieve onzuiverheden of bestanddelen: CDOM, TSM en chlorofyl-a, die de reflectantie van water (mede) bepalen. Een typische en tevens de referentiemethode om dit

te doen is curve benadering, dat elk water spectrum rechtstreeks vergelijkt met een bibliotheek van spectra die zijn verkregen door gekende (bv. in *situ* gemeten) concentraties van elk bestanddeel in een heersend water-reflectantie model in te steken. De ontmengingsaanpak vervangt dit door een lineaire ontmenging toe te passen, waarbij de endmembers gedefinieerd zijn door in hetzelfde water-reflectantie model de extreme concentraties van elk bestanddeel in te voegen – in zekere zin betekent dit een lineaire unmixing met 'niet-lineaire endmembers'. Om de endmember abundanties naar feitelijke concentraties te vertalen is een model voorgesteld waarvan de parameters zijn afgeleid door regressie op een beperkt aantal referentiewaarden die verkregen zijn uit in *situ* metingen. Om de ontmengingsbehandeling verder te bevorderen, is een methode voorgesteld om de voorgedefinieerde endmembers uit het beeld te halen door pure extractietechnieken met data clustering te combineren. De experimenten laten zien dat het vastgestelde endmember model de variabiliteit van alle water spectra ruimschoots kan beschrjven, waardoor het gebruik van niet-lineaire ontmengingsmodellen minder relevant wordt. Tegelijkertijd blijt er een fundamentele vraag over of er een complexer water-reflectantie model bestaat dat als een (niet-lineair) ontmengingsvergelijking ontleend kan worden, zodanig dat de abundanties en concentraties rechtstreekser aan elkaar kunnen worden gerelateerd. Een ander aspect is dat het beschikken over een dataset met voldoende referentiewaarden een voorwaarde blijft om goed gebruik te kunnen maken van machine-learning technieken die de out-of-sample prestatie (generalisatie vermogen) kunnen optimaliseren.

BIBLIOGRAPHY

Abid, M. (2005). *Spacecraft sensors*. Spacecraft Sensors. J. Wiley.

Abu-Mostafa, Y., Magdon-Ismail, M., and Lin, H. (2012). *Learning From Data*. AML-Book.

Adams, J., Sabol, D., Kapos, V., Filho, R., Roberts, D., Smith, M., and Gillespie, A. (1995). Classification of multispectral images based on fractions of endmembers: Application to land cover change in the Brazilian Amazon. *Remote Sens. Environ.*, 52:137–154.

Albert, A. and Mobley, C. (2003). An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters. *Opt. Express.*, pages 2873–2890.

Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. (2013). Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model. *IEEE Trans. Image Process.*, 22(4):1267–1276.

Ampe, E. M., Hestir, E. L., Bresciani, M., Salvadore, E., Brando, V. E., Dekker, A. G., Malthus, T. J., Jansen, M., Triest, L., and Batelaan, O. (2014). A wavelet approach for estimating chlorophyll-a from inland waters with reflectance spectroscopy. *IEEE Geosci. Remote Sens. Lett.*, 11(1):89–93.

Anderson, G. P., Pukall, B., Allred, C. L., Jeong, L. S., Hoke, M., and Chetwynd, J. H. (1999). FLAASH and MODTRAN4: State-of-the-art atmospheric correction for hyperspectral data. In *IEEE Aerospace Conference*, volume 4, pages 177–181.

Arai, K. (2002). Adjacency effect of layered clouds estimated with Monte Carlo simulation. *Adv. Space Res.*, 29(11):1807–1812.

Arngren, M., Schmidt, M. N., and Larsen, J. (2011). Unmixing of hyperspectral image using Bayesian nonnegative matrix factorization with volume prior. *J. Signal Process. Sys.*, 65(3):479–496.

Aurin, D. A. and Dierssen, H. M. (2012). Advantages and limitations of ocean color remote sensing in cdom-dominated, mineral-rich coastal and estuarine waters. *Remote Sens. Environ.*, 125(0):181–197.

Babin, M., Stramski, D., Ferrari, G. M., Claustre, H., Bricaud, A., Obolensky, G., and Hoepffner, N. (2003). Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around europe. *Journal of Geophysical Research: Oceans*, 108(C7):3211.

Bachmann, C., Ainsworth, T., and Fusina, R. (2005). Exploiting manifold geometry in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 43(3):441–454.

Bachmann, C., Ainsworth, T., and Fusina, R. (2006). Improved manifold coordinate representations of large-scale hyperspectral scenes. *IEEE Trans. Geosci. Remote Sens.*, 44(10):2786–2803.

Bachmann, C., Ainsworth, T., Fusina, R., Montes, M., Bowles, J., Korwan, D., and Gills, D. (2009). Bathymetric retrieval from hyperspectral imagery using manifold coordinate representations. *IEEE Trans. Geosci. Remote Sens.*, 47(3):884–897.

Bajorski, P. (2004). Simplex projection methods for selection of endmembers in hyperspectral imagery. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, volume 5, pages 3207–3210.

Ball, D. (2001). *The Basics of Spectroscopy*. SPIE P. Series. SPIE- The International Society for Optical Engineering.

Bauschke, H. H. and Borwein, J. M. (1993). On the convergence of von Neumann's alternating projection algorithm for two sets. *Set-Valued Analysis*, 1(2):185–212.

Belkin, M. and Niyogi, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv. NIPS*, 14:586–691.

Bennett, R. (1969). The intrinsic dimensionality of signal collections. *IEEE Trans. Inf. Theor.*, 15(5):517–525.

Bioucas-Dias, J. and Nascimento, J. (2008). Hyperspectral subspace identification. *IEEE Trans. Geosci. Remote Sens.*, 46(8):2435–2445.

Bioucas-Dias, J., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J. (2013). Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36.

Bioucas-Dias, J., Plaza, A., Dobigeon, N., Parente, M., Du, Q., Gader, P., and Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Topics Appl. Earth Observ.*, 5(2):354–379.

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, first edition.

Blumenthal, L. (1970). *Theory and applications of distance geometry*. Chelsea Publishing Co., New York, second edition.

Boardman, J. (1994). Geometric mixture analysis of imaging spectrometry data. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, volume 4, pages 2369–2371.

Bolshakova, N. and Azuaje, F. (2002). Cluster validation techniques for genome expression data. *Signal Processing*, 83:825–833.

Borengasser, M., Hungate, W., and Watkins, R. (2010). *Hyperspectral Remote Sensing: Principles and Applications*. Remote Sensing Applications Series. Taylor & Francis.

Broadwater, J. and Banerjee, A. (2009). A comparison of kernel functions for intimate mixture models. *IEEE Workshop on Hyperspectral Image and Sig. Proc.: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.

Broadwater, J. and Banerjee, A. (2010). A generalized kernel for areal and intimate mixtures. In *IEEE Workshop on Hyperspectral Image and Sig. Proc.: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.

Bruijn de, F., Bruls, W. H. A., Burazerović, D., and Haan de, G. (2002). Efficient video coding integrating MPEG-2 and picture-rate conversion. *IEEE Trans. Consumer Electron.*, 48(3):688–693.

Burazerović, D., Geens, B., Heylen, R., Sterckx, C., and Scheunders, P. (2013). Detecting the adjacency effect in hyperspectral imagery with spectral unmixing techniques. *IEEE J. Sel. Topics Appl. Earth Observ.*, 6:1070–1078.

Burazerović, D., Geens, B., Heylen, R., Sterckx, S., and Scheunders, P. (2012). Unmixing for detection and quantification of adjacency effects. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 3090–3093.

Burazerović, D., Gerrits, A., Taori, R., and Ritzerfeld, J. (2001). Time-scale modification for speech coding. In Heideman, G., editor, *Proc. 22nd Symposium on Information Theory in the Benelux*, pages 1–8.

Burazerović, D., Heylen, R., Raymaekers, D., Knaeps, E., Philippart, C. J. M., and Scheunders, P. (2014). A spectral-unmixing approach to estimate water-mass concentrations in case-2 waters. *IEEE J. Sel. Topics Appl. Earth Observ.*

Burazerović, D., Heylen, R., and Scheunders, P. (2011). Towards streaming hyperspectral endmember extraction. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 2519–2522.

Burazerović, D., Vandewalle, P., and Berretty, R.-P. (2009). Automatic depth profiling of 2D cinema - and photographic images. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 2365–2368.

Campbell, J. (2006). *Introduction to Remote Sensing*. Guilford Press.

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., and Malo, J. (2011). *Remote Sensing Image Processing*. Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan & Claypool Publishers.

Chang, C.-I. (2005). Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis. *IEEE Trans. Geosci. Remote Sens.*, 43(3):502–518.

Chang, C.-I. and Du, Q. (2004). Estimation of the number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 42(3):608–619.

Chang, C.-I., Ren, H., Chang, C.-C., D'Amico, F., and Jensen, J. (2004). Estimation of subpixel target size for remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.*, 42(6):1309–1320.

Chang, C.-I. and Wang, S. (2006). Constrained band selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 44:1575–1585.

Chang, C.-I., Wu, C.-C., Liu, W.-M., and Ouyang, Y.-C. (2006). A new growing method for simplex-based endmember extraction algorithm. *IEEE Trans. Geosci. Remote Sens.*, 44(10):2804–2819.

Chang, C.-I., Zhao, X.-L., Althouse, M., and Pan, J. (1998). Least squares subspace projection approach to mixed pixel classification for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 36(3):898–912.

Chatfield, C. and Collins, A. J. (2000). *Introduction to multivariate analysis*. Chapman and Hall, repr. edition.

Chaudhry, F. (2005). *Pixel Purity Index-based Endmember Extraction for Hyperspectral Data Exploitation*. University of Maryland, Baltimore County.

Chen, X. and Vierling, L. (2006). Spectral mixture analyses of hyperspectral data acquired using a tethered balloon. *Remote Sens. Environ.*, 103(3):338–350.

Chen, Y., Crawford, M., and Ghosh, J. (2006). Improved nonlinear manifold learning for land cover classification via intelligent landmark selection. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, pages 545–548.

Chi, J. and Crawford, M. (2013). Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity. *IEEE Geosci. Remote Sens. Lett.*, 10(4):711–715.

Christophe, E., Léger, D., and Mailhes, C. (2005). Quality criteria benchmark for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 43(9):2103–2114.

Clark, R. (1999). *Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy*. John Wiley and Sons, New York.

Close, R., Gader, P., Wilson, J., and Zare, A. (2012). Using physics-based macroscopic and microscopic mixture models for hyperspectral pixel unmixing. In *SPIE, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, pages 83901L–83901L–13.

Coifman, R. R. and Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30.

Combe, J.-P., Launeau, P., Carrère, V., D., D., V., M., Barillé, and C., S. (2005). Mapping microphytobenthos biomass by non-linear inversion of visible-infrared hyperspectral images. *Remote Sens. Environ.*, 98(4):371–387.

Coxeter, H. S. M. (1963). *Introduction to geometry*. Wiley, New York.

Cracknell, A. and Hayes, L. (2007). *Introduction to Remote Sensing*. Taylor & Francis.

Dekker, A. (1993). *Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing*. PhD thesis, Vrije Universiteit Amsterdam.

Dijkstra, E. (1959). A note on two problems in connexion with graphs. *Num. Math.*, 1:269–271.

Dobigeon, N., Moussaoui, S., Coulon, M., Tourneret, J.-Y., and Hero, A. O. (2009). Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery. *IEEE Trans. Signal Process.*, 57(11):4355–4368.

Dobigeon, N., Tourneret, J.-Y., and Chang, C.-I. (2008). Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. *IEEE Trans. Signal Process.*, 56(7):2684–2695.

Dobigeon, N., Tourneret, J.-Y., Richard, C., Bermudez, J., McLaughlin, S., and Hero, A. (2014). Nonlinear unmixing of hyperspectral images: Models and algorithms. *IEEE Signal Process. Mag.*, 31(1):82–94.

Dopido, I., Zortea, M., Villa, A., Plaza, A., and Gamba, P. (2011). Unmixing prior to supervised classification of remotely sensed hyperspectral images. *IEEE Geosci. Remote Sens. Lett.*, 8(4):760–764.

Dowler, S. and Andrews, M. (2011). On the convergence of n-findr and related algorithms: To iterate or not to iterate? *IEEE Geosci. Remote Sens. Lett.*, 8(1):4–8.

Duchi, J., Shwartz, S. S., Singer, Y., and Chandra, T. (2008). Efficient projections onto the $l1$-ball for learning in high dimensions. *Int. Conf. on Machine Learning*, pages 272–279.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. Wiley, New York, second edition.

Duran, O. and Petrou, M. (2009). Spectral unmixing with negative and superunity abundances for subpixel anomaly detection. *IEEE Geosci. Remote Sens. Lett.*, 6(1):152–156.

Elte, E. (1912). *The Semiregular Polytopes of the Hyperspaces*. Hoitsema.

ENVI (2009). *ENVI (v.4.7) Tutorial: Using SMACC to Extract Endmembers*. Exelis Visual Information Solutions, Boulder, Colorado.

Fisher, P. (1997). The pixel: a snare and a delusion. *Int. J. Remote Sens.*, 18(3):679–685.

Floyd, R. W. (1962). Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345.

Gashler, M., Ventura, D., and Martinez, T. (2008). Iterative non-linear dimensionality reduction with manifold sculpting. *Adv. NIPS*, 20:513–520.

Gege, P., Fries, J., Haschberger, P., Schoetz, P., Schwarzer, H., Strobl, P., Suhr, B., Ulbrich, G., and Jan Vreeling, W. (2009). Calibration facility for airborne imaging spectrometers. *Journal of Photogrammetry and Remote Sensing*, 64:387–397.

Gillis, D., Bowles, D., and Winter, M. (2002). Using endmembers as a coordinate system in hyperspectral imagery. *Proc. SPIE, Imaging Spectrometry VIII*, 4816:346–354.

Giri, C. (2012). *Remote Sensing of Land Use and Land Cover: Principles and Applications*. Remote Sensing Applications Series. Taylor & Francis.

Glenn, T., Dranishnikov, D., Gader, P., and Zare, A. (2013). Subpixel target detection in hyperspectral imagery using piece-wise convex spatial-spectral unmixing, possibilistic and fuzzy clustering, and co-registered lidar. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, pages 1063–1066.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2003). *Digital Image Processing Using MATLAB*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Goodwin, N., Coops, N. C., and Stone, C. (2005). Assessing plantation canopy condition from airborne imagery using spectral mixture analysis and fractional abundances. *Int. J. Appl. Earth Obs.*, 7(1):11–28.

Green, A. A., Berman, M., Switzer, P., and Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sens.*, 26(1):65–74.

Guilfoyle, K., Althouse, M., and Chang, C.-I. (2001). A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks. *IEEE Trans. Geosci. Remote Sens.*, 39(10):2314–2318.

Gutierrez, J., Rosario, R., and Sevilla, D. (2002). On multivariate rational function decomposition. *J. Symb. Comput.*, 33(5):545–562.

Haan de, J. and Kokke, J. (1996). Remote sensing algorithm development toolkit: 1. Operationalization of atmospheric correction methods for tidal and inland waters. *BCRS Report, RWS-Survey Department, Netherlands*, page 91.

Hakvoort, H., de Haan, J., Jordans, R., Vos, R., Peters, S., and Rijkeboer, M. (2002). Towards airborne remote sensing of water quality in the netherlands – validation and error analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(3):171–183.

Halimi, A., Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. (2011a). Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Trans. Geosci. Remote Sens.*, pages 413–416.

Halimi, A., Altmann, Y., Dobigeon, N., and Tourneret, J.-Y. (2011b). Unmixing hyperspectral images using the generalized bilinear model. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, pages 1886–1889.

Han, T. and Goodenough, D. G. (2008). Investigation of nonlinearity in hyperspectral imagery using surrogate data methods. *IEEE Trans. Geosci. Remote Sens.*, 46(10):2840–2847.

Hapke, B. (1981). Bidirectional reflectance spectroscopy 1. theory. *Journal of Geophysical Research*, 86(B4):3039–3054.

Hapke, B. (2012). *Theory of Reflectance and Emittance Spectroscopy*. Cambridge University Press.

Harsanyi, J. C. and Chang, C.-I. (1994). Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection. *IEEE Trans. Geosci. Remote Sens.*, 32:779–785.

Harsanyi, J. C., Farrand, W., and Chang, C.-I. (1993). Determining the number and identity of spectral endmembers: An integrated approach using Neyman-Pearson eigenthresholding and iterative constrained rms error minimization. In *Thematic Conf. Geologic Remote Sensing*, volume 32, pages 1–10.

He, J., Zhang, L., Wang, Q., and Li, Z. (2009). Using diffusion geometric coordinate representation for hyperspectral imagery representation. *IEEE GRSL*, 6:767–771.

Heinz, D. and Chang, C.-I. (2001). Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, 39(3):529–545.

Heylen, R., Burazerović, D., and Scheunders, P. (2011a). Fully constrained least-squares spectral unmixing by simplex projection. *IEEE Trans. Geosci. Remote Sens.*, 49:4112–4122.

Heylen, R., Burazerović, D., and Scheunders, P. (2011b). Non-linear spectral unmixing by geodesic simplex volume maximization. *IEEE J. Sel. Topics Signal Process.*, 5(3):534–542.

Heylen, R., Parente, M., and Gader, P. (2014). A review of nonlinear hyperspectral unmixing methods. *IEEE J. Sel. Topics Appl. Earth Observ.*, PP(99):1–26.

Heylen, R. and Scheunders, P. (2011). Non-linear fully-constrained spectral unmixing. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, pages 1295–1298.

Hommersom, A., Wernand, M., Peters, S., Eleveld, M. A., van der Woerd, H. J., and de Boer, J. (2011). Spectra of a shallow sea - unmixing for class identification and monitoring of coastal waters. *Ocean Dynamics*, pages 463–480.

Honeine, P. and Richard, C. (2012). Geometric unmixing of large hyperspectral images: A barycentric coordinate approach. *IEEE Trans. Geosci. Remote Sens.*, 50(6):2185–2195.

Ifarraguerri, A. and Chang, C.-I. (1999). Multispectral and hyperspectral image analysis with convex cones. *IEEE Trans. Geosci. Remote Sens.*, 37(2):756–770.

Iordache, M., Bioucas-Dias, J., and Plaza, A. (2011). Sparse unmixng of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 6(6):2014–2039.

Iordache, M., Bioucas-Dias, J., and Plaza, A. (2014). Collaborative sparse regression for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.*, 52(1):341–354.

Itten, K., Dell'Endice, F., Hueni, A., Kneubühler, M., Schläpfer, D., Odermatt, D., Seidel, F., Huber, S., Schopfer, J., Kellenberger, T., Bühler, Y., D'Odorico, P., Nieke, J., Alberti, E., and Meuleman, K. (2008). APEX - the Hyperspectral ESA Airborne Prism Experiment. *Sensors*, 8:6235–6259.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323.

Johnston, S. and Cordes, J. (2003). Public good or commercial opportunity? case studies in remote sensing commercialization. *Space Policy*, 19(1):23–31.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, second edition.

Kay, S., Hedley, J. D., and Lavender, S. (2009). Sun glint correction of high and low spatial resolution images of aquatic scenes: a review of methods for visible and near-infrared wavelengths. *Remote Sensing*, 1:697–730.

Kayadibi, O. (2011). Evaluation of imaging spectroscopy and atmospheric correction of multispectral images (aster and landsat 7 etm+). In *Proc. Int. Conf. on Recent Advances in Space Technologies (RAST)*, pages 154–159.

Kempeneers, P., Sterckx, S., Debruyn, W., DeBacker, S., Scheunders, P., Youngje, P., and Ruddick, K. (2005). Retrieval of oceanic constituents from ocean color using simulated annealing. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, volume 8, pages 5651–5654.

Keshava, N., Kerekes, J., Manolakis, D., and Shaw, G. (2000). An algorithm taxonomy for hyperspectral unmixing. In Shen, S. S. and Descour, M. R., editors, *Proceedings of SPIE*, volume 4049, pages 42–63.

Keshava, N. and Mustard, J. (2002). Spectral unmixing. *IEEE Sig. Proc. Mag.*, 19:44–57.

Kirk, J. T. O. (1994). *Light and Photosynthesis in Aquatic Ecosystems, 2nd ed.* Cambridge University Press, New York.

Knaeps, E., Dogliotti, A., Raymaekers, D., Ruddick, K., and Sterckx, S. (2012). In situ evidence of non-zero reflectance in the olci 1020nm band for a turbid estuary. *Remote Sens. Environ*, 120(0):133–144.

Knaeps, E., Sterckx, S., Ruddick, K., Giardino, C., and Bresciani, M. (2010). Simec, an environment correction for meris based on the nir similarity spectrum. In *Proc. Ocean Optics XX Conference held in Anchorage, USA*, page 12.

Kruse, F., Boardman, J., and Huntington, J. (2003). Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.*, 41:1388–1400.

Kwok, J. T. Y. and Tsang, I. W. H. (2004). The pre-image problem in kernel methods. *IEEE Trans. Neural Netw.*, 15(6):1517–1525.

Landgrebe, D. A. (2005). *Signal Theory Methods in Multispectral Remote Sensing*. Wiley Series in Remote Sensing and Image Processing. Wiley.

Lawson, C. and Hanson, R. (1974). *Solving Least Squares Problems*. Prentice Hall, Englewood Cliffs, NJ.

Lee, Z.-P., editor (2006). *Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications*, volume No. 5 of *Reports of the International Ocean Colour Coordinating Group*. IOCCG, Dartmouth, Canada.

Liangrocapart, S. and Petrou, M. (1999). Feasibility study for the use of non-linear spectral unmixing. In *EUROPTO Conference on Image and Signal Processing for Remote Sensing*, pages 159–168.

Lillesand, T., Kiefer, R., and Chipman, J. (2008). *Remote sensing and image interpretation*. John Wiley & Sons.

Lin, I.-H. (2008). *Geometric linear algebra*, volume 2. World Scientific.

Liou, K. (2002). *An Introduction to Atmospheric Radiation*. International geophysics series. Academic Press.

Liu, W. and Wu, E. (2005). Comparison of non-linear mixture models: sub-pixel classification. *Remote Sens. Environ.*, 94(2):145–154.

Luo, B. and Chanussot, J. (2009). Unsupervised classification of hyperspectral images by using linear unmixing algorithm. In *IEEE Int. Conf. on Image Process.*, pages 2877–2880. IEEE.

Ma, L., Crawford, M. M., and Tian, J. (2010). Local manifold learning-based k-nearest-neighbor for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 48(11):4099–4109.

Ma, Y., Niyogi, P., Sapiro, G., and Vidal, R. (2011). Dimensionality reduction via subspace and submanifold learning [from the guest editors]. *IEEE Signal Process. Mag.*, 28(2):14–16.

Maaten van der, L. J. P., Postma, E. O., and Herik van den, H. J. (2008). Dimensionality Reduction: A Comparative Review. *Tilburg Universtity Technical Report TiCC TR 2009-005*. Available online.

Mahmood, Z., Akhter, M. A., Thoonen, G., and Scheunders, P. (2013). Contextual subpixel mapping of hyperspectral images making use of a high resolution color image. *IEEE J. Sel. Topics Appl. Earth Observ.*, 6(2):779–791.

Manolakis, D., Marden, D., and Shaw, G. A. (2003). Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116.

Maul, G. and Gordon, H. (1975). On the use of the Earth resources technology satellite (LANDSAT-1) in optical oceanography. *Remote Sens. Environ.*, 4(0):95 – 128.

Mesev, V. (2007). *Integration of GIS and Remote Sensing*. Mastering GIS: technology. Wiley.

Miao, L., Qi, H., and Szu, H. (2007). A maximum entropy approach to unsupervised mixed-pixel decomposition. *IEEE Trans. Image Process.*, 16(4):1008–1021.

Michelot, C. (1986). A finite algorithm for finding the projection of a point onto the canonical simplex of $\mathbb{R}^n$. *J. Optim. Theo. Applic.*, 50:195–200.

Mobley, C. (1994). *Light and Water: Radiative Transfer in Natural Waters*. Academic Press.

Morgan, D. and Falkner, E. (2001). *Aerial Mapping: Methods and Applications, Second Edition*. Mapping SScience. Taylor & Francis.

Müller, K.-R., Adali, T., Fukumizu, K., Principe, J. C., and Theodoridis, S. (2013). Special issue on advances in kernel-based learning for signal processing [from the guest editors]. *IEEE Signal Process. Mag.*, 30(4):14–15.

Nascimento, J. M. P. and Bioucas-Dias, J. M. (2005a). Does independent component analysis play a role in unmixing hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 43(1):175–187.

Nascimento, J. M. P. and Bioucas-Dias, J. M. (2005b). Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 43(4):898–910.

Nascimento, J. M. P. and Bioucas-Dias, J. M. (2009). Nonlinear mixture model for hyperspectral unmixing. In *Image Sig. Process. for Remote Sens. XV*, volume 7477, pages 74770I–74770I–8.

Natural Resources Canada (2014). Tutorial: Fundamentals of Remote Sensing, @http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9309.

Nechad, B., Ruddick, K., and Park, Y. (2010). Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sens. Environ.*, 114(4):854–866.

Onn, S. and Weissman, I. (2011). Generating uniform random vectors over a simplex with implications to the volume of a certain polytope and to multivariate extremes. *Ann. Oper. Res.*, 189(1):331–342.

Parente, M. and Plaza, A. (2010). Survey of geometric and statistical unmixing algorithms for hyperspectral images. In *IEEE Workshop on Hyperspectral Image and Sig. Proc.: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.

Parente, M. and Zymnis, A. (2005). Statistical clustering and mineral spectral unmixing in Aviris hyperspectral image of Cuprite, NV. *CS229 Report*. Available online.

Plaza, A., Martín, G., Plaza, J., Zortea, M., and Sánchez, S. (2011). Recent developments in spectral unmixing and endmember extraction. In Prasad, S., Bruce, L., and Chanussot, J., editors, *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*. Springer.

Plaza, A., Martínez, P., Pérez, R., and Plaza, J. (2004). A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 42(3):650–663.

Plaza, J., Plaza, A., Pérez, R., and Martínez, P. (2009). On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images. *Pattern Recognition*, 42(11):3032–3045.

Pope, R. and Fry, E. (1997). Absorption spectrum (380-700nm) of pure water. II.: Integrating cavity measurements. *Appl. Optics*, 36(33):8710–8723.

Prasad, S., Bruce, L., and Chanussot, J. (2011). *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*. Augmented Vision and Reality. Springer.

Rees, G. (1999). *The Remote Sensing Data Book*. Cambridge University Press.

Richter, R. (2008). *Atmospheric/Topographic Correction of Airborne Imagery. ATCOR-4 User Guide Version 4.3*. Wessling, Germany.

Richter, R., Bachmann, M., Dorigo, W., and Muller, A. (2006). Influence of the adjacency effect on ground reflectance measurements. *IEEE Geosci. Remote Sens. Lett.*, 3(4):565–569.

Robinson, I. S. (2010). *Discovering the ocean from space: the unique applications of satellite oceanography*. Springer/Praxis Publishing.

Roweis, S. and Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326.

Ruddick, K. G., De-Cauwer, V., and Park, Y. J. (2006). Seaborne measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters. *American Society of Limnolgy and Oceanography, Inc.*, 51(2):1167–1179.

S., J. and Qian, Y. (2009). Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.*, 47(1):161–173.

Salama, S., Monbaliu, J., and Coppin, P. (2004). Atmospheric correction of advanced very high resolution radiometer imagery. *Int. J. Remote Sens.*, 25(7-8):1349–1355.

Schiller, H. and Doerffer, R. (2005). Improved determination of coastal water constituent concentrations from meris data. *IEEE Trans. Geosci. Remote Sens.*, 43(7):1585–1591.

Schott, J. (2009). *Fundamentals of Polarimetric Remote Sensing*. SPIE tutorial texts. SPIE.

Semenov, A., Moshkov, A., Pozhidayev, V., Barducci, A., Marcoionni, P., and Pippi, I. (2011). Estimation of normalized atmospheric point spread function and restoration of remotely sensed images. *IEEE Trans. Geosci. Remote Sens.*, 49:2623–2634.

Shaw, G. A. and hua K. Burke, H. (2003). Spectral imaging for remote sensing.

Shimabukuro, Y. E. and Smith, J. A. (1991). The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Trans. Geosci. Remote Sens.*, 29(1):16–20.

Shwartz, S. S., Singer, Y., Bennett, P., and Parrado-Hernandez, E. (2006). Efficient learning of label ranking by soft projections onto polyhedra. *JMLR*, 7:1567–1599.

Silva de, V. and Tenenbaum, J. B. (2002). Global versus local methods in nonlinear dimensionality reduction. In Becker, S., Thrun, S., and Obermayer, K., editors, *NIPS*, pages 705–712. MIT Press.

Soergel, U. (2010). *Radar Remote Sensing of Urban Areas*. Remote Sensing and Digital Image Processing. Springer.

Somers, B., Asner, G., Tits, L., and Coppin, P. (2011). Endmember variability in Spectral Mixture Analysis: A review. *Remote Sens. Environ.*, 115(7):1603–1616.

Somers, B., Cools, K., Delalieux, S., Stuckens, J., der Zande, D. V., Verstraeten, W., and Coppin, P. (2009). Nonlinear hyperspectral image analysis for tree cover estimates in orchards. *Remote Sens. Environ.*, 113:1183–1193.

Sterckx, S., Knaeps, E., and Ruddick, K. (2011). Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: the use of the near infrared similarity spectrum. *Int. J. Remote Sens.*, 32(21):6479–6505.

Strang, G. (2009). *Introduction to linear algebra*. Wellesley-Cambridge Press.

Su, Z., Roebeling, R., Schulz, J., Hollemann, I., Levizzani, V., Timmermans, J., Rott, H., Mognard-Campbell, N., De Jeu, R., Wagner, W., Rodell, M., Salama, M. S., Parodi, G., and Wang, L. (2011). Observation of hydrological processes using remote sensing. In *Treatise on Water Science*, pages 351–399. Elsevier, Oxford: Academic Press. invited.

Tarabalka, Y., Benediktsson, J., Chanussot, J., and Tilton, J. (2010). Multiple spectral–spatial classification approach for hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, 48(11):4122–4132.

Tassan, S. and Ferrari, G. (1995). An alternative approach to absorption measurements of aquatic particles retained on filters. *Limnology and Oceanography*, 40(8):1358–1368.

Tenenbaum, J. B., Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

Thoonen, G., Spanhove, T., Haest, B., Vanden Borre, J., and Scheunders, P. (2010). Habitat mapping and quality assessment of heathlands using a modified kernel-based reclassification technique. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, pages 2707–2710, Honolulu, HI, USA.

Tilstone, G., Moore, G., Sorensen, K., Doerffer, R., Rottgers, R., Ruddick, K., Pasterkamp, R., and Jorgensen, P. (2003). Regional validation of MERIS chl products in North Sea coastal waters: Protocols for the validation of MERIS products in case-2 waters. In *Proc. ENVISAT validation workshop*. European Space Agency publications.

Toutin, T. (2004). Review article: Geometric processing of remote sensing images: models, algorithms and methods. *Int. J. Remote Sens.*, 25(10):1893–1924.

Tso, B. and Mather, P. (2009). *Classification Methods for Remotely Sensed Data*. Environmental engineering. CRC Press.

Tuenter, H. J. H. (2001). The minimum L2-distance projection onto the canonical simplex: A simple algorithm. *Algo Research Quarterly*, 4:53–55.

Ufermann, S. (2003). *Evaluation of a Semi-analytical Approach to the Retrieval of Water Quality Parameters from Optical Data in European Case-II Waters.* University of Southampton.

Veganzones, M. A. and Graña, M. (2008). Endmember extraction methods: A short review. In *Proc. Knowledge-Based Intelligent Information & Engineering Systems (KES)*, pages 400–407.

Velez-Reyes, M. and Rosario, S. (2004). Solving adundance estimation in hyperspectral unmixing as a least distance problem. *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, 5:3276–3278.

Villa, A., Chanussot, J., Benediktsson, J., and Jutten, C. (2011). Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution. *IEEE J. Sel. Topics Appl. Earth Observ.*, 5(3):521–533.

Wang, X., Kumar, S., Ramos, F., Kaupp, T., Upcroft, B., and Durrant-Whyte, H. (2006). Probabilistic classification of hyperspectral images by learning nonlinear dimensionality reduction mapping. In *Proc. 9th Int. Conf. Information Fusion*, pages 1–8.

Winter, E. (1999). N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Proc. SPIE*, 3753:266–275.

Woerd van der, H. and Pasterkamp, R. (2008). Hydropt: A fast and flexible method to retrieve chlorophyll-a from multispectral satellite observations of optically complex coastal waters. *Remote Sens. Environ.*, 112:1795–1807.

Wu, B., Zhou, Y., Yan, L., Yuan, J., Li, R., and Wang, D. (2009). Object detection from HS/MS and multi-platform remote sensing imagery by the integration of biologically and geometrically inspired approaches. In *ASPRS Ann. Conf.*

Yangchi, C., Crawford, M., and Ghosh, J. (2005). Applying nonlinear manifold learning to hyperspectral data for land cover classification. In *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, pages 4311–4314.

Zare, A. and Gader, P. (2007). Sparsity promoting iterated constrained endmember detection in hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.*, 4(3):446–450.

Zhang, Z. and Zha, H. (2004). Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment. *SIAM J. Sci. Comput.*, 26(1):313–338.

CURRICULUM VITAE

Dževdet Burazerović (∗ 2 August 1971, Sarajevo, Bosnia and Herzegovina, then Yugoslavia) graduated in May 1986 from the Second Gymnasium in Sarajevo, receiving the qualification of Mathematical Technician. In August 2000, he obtained the Master of Science degree from the Department of Electrical Engineering at Eindhoven University of Technology, The Netherlands, with specialization in digital signal processing. In the same year he joined Philips Research in Eindhoven, where he worked until late 2009 in several departments as Research Scientist, performing applied and topical research related to digital speech, audio and video. Part of this work was done in support of IP development and several standardization activities. In November 2009, he joined the Vision Lab, Department of Physics of the University of Antwerp, Belgium, as a Ph.D. student in the domain of hyperspectral image processing for remote sensing. There, he carried out research in spectral unmixing, the largest part of which relates to the work described and cited in this Ph.D thesis.

PUBLICATIONS

Journal articles

✓ Bruijn de, F., Bruls, W. H. A., Burazerović, D., and Haan de, G. (2002). Efficient video coding integrating MPEG-2 and picture-rate conversion. *IEEE Trans. Consumer Electron.*, 48(3):688–693

✓ Heylen, R., Burazerović, D., and Scheunders, P. (2011b). Non-linear spectral unmixing by geodesic simplex volume maximization. *IEEE J. Sel. Topics Signal Process.*, 5(3):534–542

✓ Heylen, R., Burazerović, D., and Scheunders, P. (2011a). Fully constrained least-squares spectral unmixing by simplex projection. *IEEE Trans. Geosci. Remote Sens.*, 49:4112–4122

✓ Burazerović, D., Geens, B., Heylen, R., Sterckx, C., and Scheunders, P. (2013). Detecting the adjacency effect in hyperspectral imagery with spectral unmixing techniques. *IEEE J. Sel. Topics Appl. Earth Observ.*, 6:1070–1078

✓ Burazerović, D., Heylen, R., Raymaekers, D., Knaeps, E., Philippart, C. J. M., and Scheunders, P. (2014). A spectral-unmixing approach to estimate water-mass concentrations in case-2 waters. *IEEE J. Sel. Topics Appl. Earth Observ.*

Conference proceedings

✓ Burazerović, D., Gerrits, A., Taori, R., and Ritzerfeld, J. (2001). Time-scale modification for speech coding. In Heideman, G., editor, *Proc. 22nd Symposium on Information Theory in the Benelux*, pages 1–8

- ✓ Burazerović, D., Vandewalle, P., and Berretty, R.-P. (2009). Automatic depth profiling of 2D cinema - and photographic images. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pages 2365–2368

- ✓ Burazerović, D., Heylen, R., and Scheunders, P. (2011). Towards streaming hyperspectral endmember extraction. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 2519–2522

- ✓ Burazerović, D., Geens, B., Heylen, R., Sterckx, S., and Scheunders, P. (2012). Unmixing for detection and quantification of adjacency effects. In *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, pages 3090–3093

Patents and patent applications

- *EP*: 1618744 *B*1, 1380029 *B*1, 1889203 *A*2, 1842374 *A*2

- *US*: 20030033140 *A*1, 20050226330 *A*1, 2005013496 *A*1, 20060104357, 2006262846 *A*1, 2007041447

- *WO*: 2004080050, 2004080081, 2004056112, 2005074296, 2005074297, 2005036886, 2006043192, 2006048807, 2007036888, 2007039871, 2008018042