A SIMPLE SCHEME FOR UNMIXING HYPERSPECTRAL DATA BASED ON THE GEOMETRY OF THE N-DIMENSIONAL SIMPLEX

Paul Honeine[(1)], *Cédric Richard*[(2)]

[(1)] Institut Charles Delaunay (FRE CNRS 2848), LM2S, Université de technologie de Troyes, 10010 Troyes, France

[(2)] Observatoire de la Côte d'Azur (UMR CNRS 6525), Fizeau lab., Université de Nice Sophia-Antipolis, 06108 Nice, France

paul.honeine@utt.fr, cedric.richard@unice.fr

Based on observation satellites, remote-sensing with hyperspectral image analysis has gained wide popularity, with applications to mineral exploration, environmental monitoring, and military surveillance. With a large number of measured wavelength bands, each pixel has a complete spectrum. It is usually assumed that each spectrum can be linearly decomposed into spectra, each provided from some pure materials. For a given image, the hyperspectral unmixing problem involves the extraction of these *pure* spectra, the so-called endmembers, and the estimation of their abundances within each pixel. In this paper, we assume the number of endmembers is known; otherwise methods such as the one presented in [1] can be used to estimate it

The first step is the endmember extraction problem. Using convex geometry, this problem is equivalent to finding a data-enclosing simplex, as proposed in [2, 3]. One of the most popular (automated) algorithms for endmember extraction is Winter's N-Findr [3]. The main driving forces behind this algorithm is a simple iterative and scalable scheme. With a random initial set defining the simplex, it seeks to stretch the simplex in order to maximize its volume by visiting each pixel. The N-Findr technique gives a fast endmember extraction method with low computational cost, making it one of the most widely used. The second step involves estimating the abundance of each endmember for a given pixel. Many techniques have been proposed to estimate the proportions of pure spectra in a given spectrum. The main difficulties reside in constraints on proportions, with the sum-to-one and the nonnegativity (see [4, 5] and references therein).

In this paper, we give a simple algorithm for extracting the endmembers and their respective adundances, jointly in a single step. The proposed approach takes advantage of the well-known N-Finder scheme for endmember extraction. In the same spirit, we show that we can provide an efficient iterative algorithm for estimating the abundancies. Both the endmember extraction and the abundancies estimation are done in a single step, using the concept of barycentric coordinates, computed from volumes of simplices. This is done without any additional computational cost as opposed to classical least square estimation techniques [5].

In a hyperspectral image, the hypothesis of a linear mixed model is often used. The spectrum of a pixel is given as a linear combination of some *pure* spectra, called endmembers. Let $\boldsymbol{x}_\ell$ be the spectrum of the ℓ-th pixel, then the *mathematical model* is

$$\boldsymbol{x}_\ell = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_{\omega_i} + \epsilon, \tag{1}$$

where endmembers are defined by spectra $\boldsymbol{x}_{\omega_i}$, assumed represented by some pixels $\{\omega_1, \omega_2, \ldots, \omega_n\}$, and ϵ corresponds to the unfitness of the model, often treated as a Gaussian noise. In order for the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ to represent the physical abundance fraction associated to each endmember, two constraints must be imposed on theses coefficients: (1) the sum-to-one constraint $\sum_{i=1}^{n} \alpha_i = 1$, called hereafter the equality constraint, and (2) the nonnegativity constraint with $\alpha_i \geq 0$ for $i = 1, \ldots, n$. Solving the optimization problem (3) subject to both constraints requires advanced optimization techniques, as illustrated for instance in [5]. In this paper, we give a direct scheme to solve the equality-constrained optimization problem, incorporating naturally the endmember extraction procedure, and show a geometric interpretation to the violation of the nonnegativity constraint.

Unmixing hyperspectral data based on the linear model (3) involves two tasks: extracting the endmembers $\boldsymbol{x}_{\omega_i}$ and computing the coefficients α_i, for $i = 1, 2, \ldots, n$, and for the spectrum of each pixel. We begin next with the extraction method, as given in the N-finder scheme [3], and then drive in next paragraph the method to estimation the corresponding coefficients. But before, we give a definition of the volume of a simplex. Let $\boldsymbol{X} = \{\boldsymbol{x}_{\omega_1}, \boldsymbol{x}_{\omega_2}, \ldots, \boldsymbol{x}_{\omega_n}\}$ be the set of estimated endmembers,

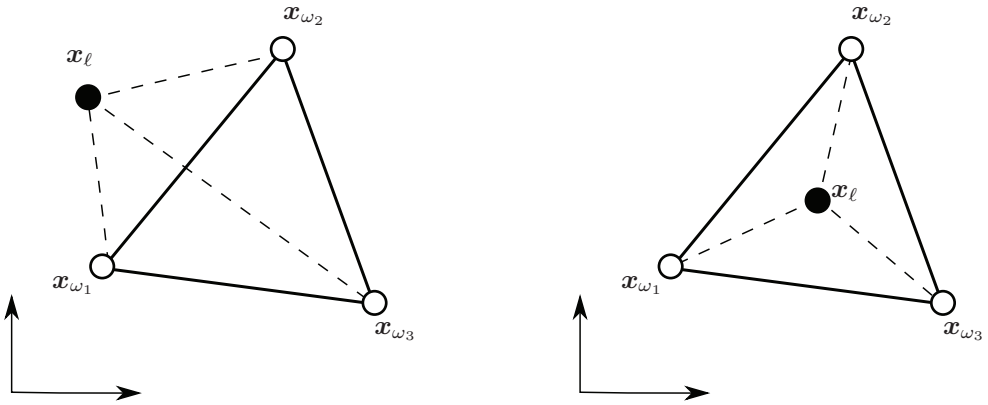

Fig. 1. Illustration of the simplex inclusion test in a 2-dimensional Euclidean space, when (left) the current data $\boldsymbol{x}_\ell$ is outside the simplex (here triangle) defined by $\{\boldsymbol{x}_{\omega_1}, \boldsymbol{x}_{\omega_1}, \boldsymbol{x}_{\omega_1}\}$, or (right) inside the simplex and thus can be removed from the candidates in that case.

with $\omega_i \in \{1, 2, \cdots\}$. The oriented volume of the simplex defined by the vertices $\boldsymbol{X}$ is given by

$$\mathcal{V}_{\boldsymbol{X}} = \frac{1}{(n-1)!} \det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{\omega_1} & \boldsymbol{x}_{\omega_2} & \cdots & \boldsymbol{x}_{\omega_n} \end{bmatrix}, \tag{2}$$

where $\det$ is the determinant operator. While taking the absolute value of this expression gives the (classical) volume of a simplex, the virtues of the oriented volume will be demonstrated in estimating the coefficients.

We take advantage of the fact that the set of endmembers defines the vertices of a simplex englobing all spectra of the image. Thus, this simplex has the largest volume among all simplices constructed from other spectra. Therefore, one seeks the simplex of largest volume, in an iterative manner by visiting each pixel. At initialization, a random set of pixels is selected as endmembers. The following process is iterated for each pixel, where $\boldsymbol{x}_\ell$ is its spectrum: One at a time, each endmember is replaced by the spectrum under investigation, and the oriented volume of the resulting simplex is computed, that is

$$\mathcal{V}_{\boldsymbol{X}\setminus\{\boldsymbol{x}_{\omega_i}\}\cup\{\boldsymbol{x}_\ell\}} \qquad i = 1, 2, \ldots, n,$$

where $\setminus$ denotes the set difference, $\mathcal{V}_{\boldsymbol{X}\setminus\{\boldsymbol{x}_{\omega_i}\}\cup\{\boldsymbol{x}_\ell\}}$ is the simplex (oriented) volume of vertices $\boldsymbol{X}\setminus\{\boldsymbol{x}_{\omega_i}\} \cup \{\boldsymbol{x}_\ell\}$, i.e. the set $\boldsymbol{X}$ with $\boldsymbol{x}_{\omega_i}$ removed and $\boldsymbol{x}_\ell$ added. Thus, we have n candidate sets, each defining a simplex, as well as the initial simplex. By comparing their volumes, two cases are can be distinguished:

- if $\max_i |\mathcal{V}_{\boldsymbol{X}\setminus\{\boldsymbol{x}_{\omega_i}\}\cup\{\boldsymbol{x}_\ell\}}| < |\mathcal{V}_{\boldsymbol{X}}|$, then the initial set of endmembers remains unchanged;
- otherwise, an entry of the initial set is substituted with $\boldsymbol{x}_\ell$ to give the new endmembers set. The outgoing spectrum, i.e. ω_i, is determined such as $i = \arg\max |\mathcal{V}_{\boldsymbol{X}\setminus\{\boldsymbol{x}_{\omega_i}\}\cup\{\boldsymbol{x}_\ell\}}|$.

After extracting the endmembers, we are now in a position to compute the coefficients. This is done at the same iteration, using the computed oriented volumes. Given the simplex defined by the vertices $\boldsymbol{x}_{\omega_1}, \boldsymbol{x}_{\omega_2}, \ldots, \boldsymbol{x}_{\omega_n}$, any $\boldsymbol{x}_\ell$ can be written as a linear combination of these vertices, i.e.

$$\boldsymbol{x}_\ell = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_{\omega_i} \qquad \text{with} \qquad \sum_{i=1}^{n} \alpha_i = 1, \tag{3}$$

To solving this constrained optimization problem, we combine the above expressions into the following incremented[1] linear system

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{\omega_1} & \boldsymbol{x}_{\omega_2} & \cdots & \boldsymbol{x}_{\omega_n} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 1 \\ \boldsymbol{x}_\ell \end{bmatrix},$$

[1]Using an incremented system has been previously introduced in the literature, with the fixed-point-free transform [6]. To our knowledge, this is the first time that barycentric coordinates are applied to provide a simple optimization scheme.

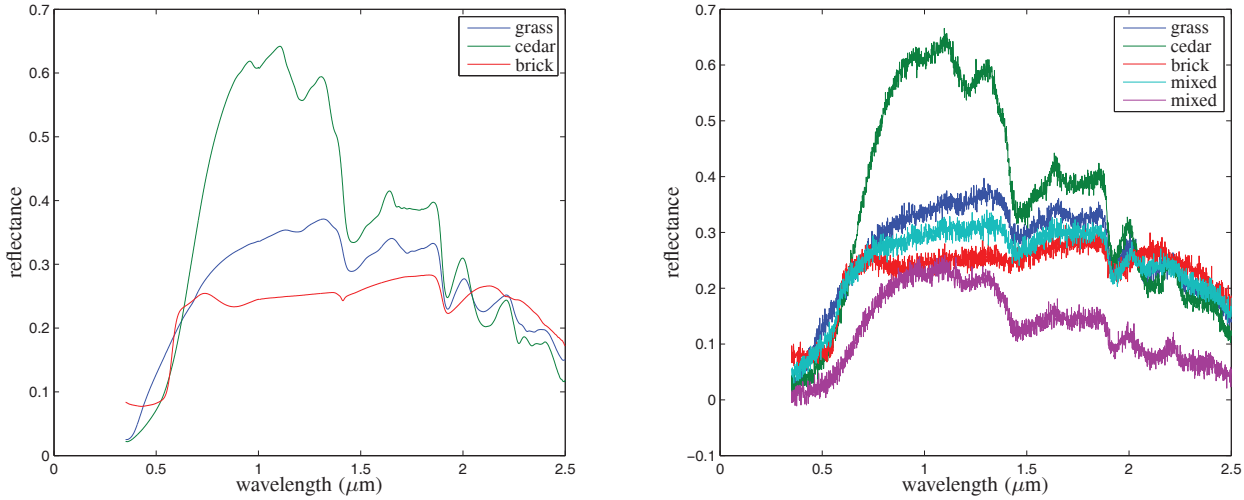

Fig. 2. Endmember spectra: (left) initial spectra and (right) their noisy versions used in generating the hyperspectral image, including two random mixed spectra.

where $\boldsymbol{\alpha}$ is the column-vector of coefficients α_i. Using Cramer's rule, the solution of this linear system can be expressed in terms of the determinants of the above matrix and of matrices obtained from it with one column substituted by the right-hand-side vector. Therefore, we can write

$$\alpha_1 = \frac{\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_\ell & \boldsymbol{x}_{\omega_2} & \cdots & \boldsymbol{x}_{\omega_n} \end{bmatrix}}{\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{\omega_1} & \boldsymbol{x}_{\omega_2} & \cdots & \boldsymbol{x}_{\omega_n} \end{bmatrix}} \quad \cdots \quad \alpha_n = \frac{\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{\omega_1} & \boldsymbol{x}_{\omega_2} & \cdots & \boldsymbol{x}_\ell \end{bmatrix}}{\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_{\omega_1} & \boldsymbol{x}_{\omega_2} & \cdots & \boldsymbol{x}_{\omega_n} \end{bmatrix}}. \tag{4}$$

Thus, each coefficient α_i equals a ratio of the oriented volumes of two simplices, the one with vertices $\boldsymbol{X}$ given by endmembers and the one resulting from the latter by substituting vertex $\boldsymbol{x}_{\omega_i}$ with $\boldsymbol{x}_\ell$. In other words, we have

$$\alpha_i = \mathcal{V}_{\boldsymbol{X} \setminus \{\boldsymbol{x}_{\omega_i}\} \cup \{\boldsymbol{x}_\ell\}} / \mathcal{V}_{\boldsymbol{X}},$$

for $i = 1, 2, \ldots, n$. Since these volumes are initially computed in order to stretch the initial simplex for enclosing most spectra, no additional computational cost is required to evaluate the coefficients at each iteration. The corresponding simplices are illustrated in Fig. 1 for two-dimensional data in two case, whether $\boldsymbol{x}_\ell$ lies inside or outside the endmembers simplex.

From the literature of geometry of convex sets, the coefficients computed using (4) are called barycentric coordinates, defined by the vertices $\boldsymbol{X}$ of the simplex. It is well known that $\boldsymbol{x}_\ell$ is inside the simplex, if and only if all α_i's are nonnegative; otherwise, there exists at least one negative coefficient when it is outside. Therefore, when the solution of the optimization problem (3) gives at least one negative α_i, this means that the $\boldsymbol{x}_\ell$ is outside the simplex, and thus cannot be written in terms of linear combination of its vertices with both equality and nonnegativity constraints. It is worth noting that such limitation is valid for any simplex-based approach.

In order to illustrate the proposed method, we simulated a synthetic hyperspectral image from a linear combination of three pure materials. These materials are available from the USGS library [7], and correspond to golden grass, red brick, and cedar, with spectra illustrated in Fig. 2 (left). A 32-by-32 hyperspectral image is generated from pixels given by the model (3), where ϵ corresponds to a white Gaussian noise of variance 0.01. Using the same noisy model with the endmembers are incorporated in the image with the canonical coefficients $\{(1,0,0);(0,1,0);(0,0,1)\}$, as illustrated in Fig. 2 (right).

Most simplex-based methods require a preprocessing dimensionality reduction technique. For this purpose, we apply a classical PCA as preconised in [3]; however, more dedicated methods such as minimum noise fraction technique can be considered. We apply the proposed method, by setting the number of endmembers to 3. The resulting largest-volume simplex encloses 71% of the spectra, following from incorporating noise, not only in the coefficients by also in the endmembers. To measure the performance of the algorithm, we compute the spectral angle error [8] for each pixel, between its initial spectrum and the one

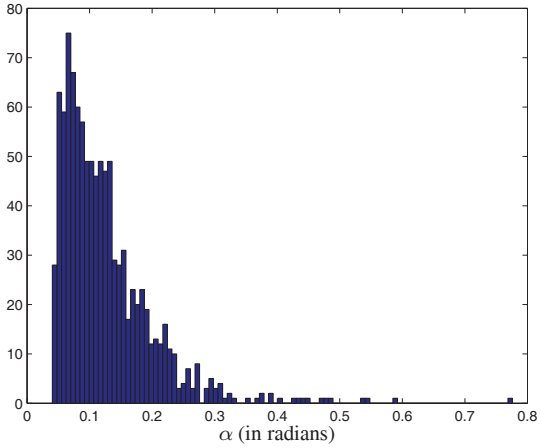

Fig. 3. Histogram of the error using the spectral angle between initial and obtained specta.

computed from the obtained coefficients, of the form

$$\alpha(\boldsymbol{x}_i, \boldsymbol{x}_j) = \cos^{-1} \frac{\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle}{\|\boldsymbol{x}_i\| \|\boldsymbol{x}_j\|}.$$

The histogram of these errors given in Fig. 3 shows a small angular error for the hyperspectral image.

1. REFERENCES

[1] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004.

[2] J. W. Boardman, "Automating spectral unmixing of aviris data using convex geometry concepts," in *Proc. Summ. 4th Annu. JPL Airborne Geosci. Workshop*, R. O. Green, Ed., 1994, pp. 11–14.

[3] M. Winter, "N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proceedings of SPIE: Imaging Spectrometry V*, vol. 3753, no. 10, 1999.

[4] N. Keshava, J. P. Kerekes, D. Manolakis, and G. A. Shaw, "An algorithm taxonomy for hyperspectral unmixing," in *Proceedings of the SPIE, SPIE AeroSense, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Detection and Identification I*, vol. 4049. Orlando, Florida, United States: SPIE, April 2000, pp. 42–63.

[5] D. Heinz and C. Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE trans on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, March 2001.

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

[7] R. N. Clark and Geological Survey (U.S.), *USGS digital spectral library splib06a [electronic resource]*, rev. sept. 20, 2007. ed. U.S. Geological Survey, Denver, CO, US, 2007.

[8] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, J. P. Barloon, and A. F. H. Goetz, "The spectral image processing system (sips): Interactive visualization and analysis of imaging spectrometer data," *Remote Sensing Environ*, vol. 2-3, no. 44, pp. 145–163, 1993.