

# A GRAPH-BASED METHOD FOR NON-LINEAR UNMIXING OF HYPERSPECTRAL IMAGERY.

*Rob Heylen, Dzevdet Burazerovic, Paul Scheunders*

IBBT-Visionlab, University of Antwerp  
Universiteitsplein 1, Building N  
B-2610 Wilrijk (Antwerp), Belgium

## 1. INTRODUCTION

An important aspect of hyperspectral image processing is unmixing, or the decomposition of individual pixel spectra into a combination of a small number of endmembers, representing the spectra of pure components. Many well-known algorithms for unmixing belong to the field of linear spectral mixture analysis, where one assumes that the constituents of a spectrum mix linearly, with mixing coefficients that are equivalent to their abundance. The physical interpretation of the mixing coefficients introduces two restrictions: positivity and normalization. Because of these two restrictions, every mixed pixel lies within a simplex in spectral space spanned by the endmembers. Most of the linear unmixing techniques (pixel purity index (PPI), N-findR, Simplex growing algorithm (SGA), D-max, ...) exploit this geometric notion in one way or another. For instance, the pixel purity index algorithm [1] projects the pixel spectra many times onto random vectors in spectral space, and tags the extremities. The spectra that got tagged the most are considered endmembers. The N-findR algorithm [2] searches for the largest volume simplex via an iterative procedure, and assigns the vertices of this simplex as endmembers.

These algorithms have in common that they assume that the spectra are composed of linear combinations of endmembers. In reality however, this is often not the case [3, 4]. Examples are the multiple scattering in shallow water or wet environments, secondary reflections through vegetation canopies or between light and dark surface materials [5], and the delicate spectral interactions due to multiple scattering between minerals for very short interaction lengths (e.g. in beach sands). In these cases, linear unmixing will often not yield correct results: One can find too much or wrong endmembers, and subsequently the abundance maps are erroneous as well.

Some methods to deal with non-linear unmixing of hyperspectral data involve neural networks or support vector machines [6, 7, 8, 9], or are based upon modeling the source of the non-linearity [4]. In this paper, we present a new approach for non-linear unmixing, by combining linear unmixing with a non-flat metric determined by the data manifold in spectral space. In order to do so, we combine parts of the linear unmixing technique N-findR with parts of the ISOMAP algorithm for non-linear dimensional reduction [10]. We do not require a specific model for the non-linear interactions between the spectra, and believe that this technique is applicable to most non-linear models.

## 2. DETERMINATION OF THE ENDMEMBERS

Suppose we have a hyperspectral image consisting of  $N$  pixels in a  $d$ -dimensional spectral space:  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ . If every pixel is a linear combination of  $p$  endmembers  $(\mathbf{e}_1, \dots, \mathbf{e}_p)$ , we can write

$$\mathbf{x}_i = \sum_{j=1}^p a_{ij} \mathbf{e}_j \quad (1)$$

with the endmember abundances given by the abundance coefficients  $a_{ij}$ . Furthermore, one assumes two restrictions on these abundance coefficients: positivity and normalization:  $\forall i, j : a_{ij} \geq 0, \sum_j a_{ij} = 1$ . With these constraints, the pixels will be contained within a  $(p-1)$ -dimensional simplex spanned by the  $p$  endmembers  $\{\mathbf{e}_i\}_i$ . To determine these unknown endmembers, we employ the same reasoning as in the original N-findR algorithm [2], where endmembers are determined as the vertices of a maximal volume simplex in spectral space. If a pure pixel is present for every endmember, this algorithm will return the correct endmembers.

The core of the algorithm is the calculation of simplex volumes. Suppose the distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are given by  $d_{ij}$ , and  $\mathbf{D}^2 = \{d_{ij}^2\}_{ij}$ . The volume  $V$  of a  $p - 1$ -dimensional simplex spanned by vertices  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$  can be written in terms of these distances [11]:

$$(-1)^p 2^{p-1} (p-1)!^2 V^2 = \det \begin{pmatrix} \mathbf{D}^2 & \mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \quad (2)$$

Using standard matrix identities, it is possible to rewrite this determinant as a product of a factor corresponding to an orthogonal distance of a single vertex to a  $(p - 2)$ -dimensional simplex, and a factor corresponding to the volume of this simplex. One advantage of working with distances between points instead of the actual spectral coordinates is that equation (2) is valid for any dimension  $p$ .

### 3. INCORPORATING NON-LINEARITY

If the endmember spectra mix non-linearly, the scatter plot of the pixels in spectral space will usually no longer form a simplex. The exact shape of this scatter plot depends on the physics underlying the spectral mixing of endmembers, or on the model used, and can vary significantly between situations. Since the endmember abundances still have to obey the positivity and normalization constraint, there exists a mapping between the space of abundance coefficients (which can be interpreted as coordinates in a standard  $(p - 1)$ -simplex), and the spectral space. We assume that this mapping preserves the topological structure of the simplex. In this case, we could use a non-linear dimensional reduction or transformation to preprocess the data before using it in a linear unmixing procedure. However, with any dimensional reduction, one loses information, and this might strongly affect the endmembers selected, certainly in geometric methods where extreme points and outliers will have a large influence on the end result.

Since equation (2) allows us to calculate the volume of a simplex based upon the distances between the vertices, it is desirable to find a method to estimate the distances between pixels in spectral space, but taking into account the global non-linear structure of the data manifold. The approach we use is derived from the popular non-linear dimensional reduction technique ISOMAP [10]: First, we construct a weighted, connected, symmetric nearest-neighbor graph on the data. To construct such a graph, choose a constant  $K$ , connect every point in spectral space to its  $K$  neighbors with smallest Euclidean distance, and symmetrize the graph. The weight of every edge is its Euclidean length. The distance between two points is then defined as the shortest-path distance along this weighted graph, and can be calculated with the Dijkstra algorithm [12]. Since the graph structure will capture the overall geometry of the data manifold, the paths along which this distance is measured will follow the curvature of this manifold, and approximate the actual geodesic distance.

### 4. DERIVING THE ENDMEMBER ABUNDANCES

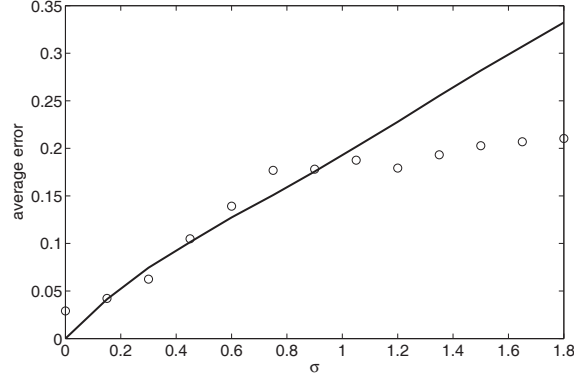
When we have found the endmembers, the final step is to decompose each pixel into its endmember abundances. In the case of linear mixing, this involves the solving of an overdetermined linear system of equations in the  $p$  endmember abundances  $a_{ij}$  for every pixel  $\mathbf{x}_i$ . Such systems can be solved via the least-squares method, or via constrained least-squares methods where (some of) the constraints on  $a_{ij}$  are taken into account. In the case of non-linear unmixing however, we have no such relation. Especially in our case here, where we did not assume any given model for the non-linearities, this abundance estimation seems to be a highly non-trivial task.

However, we can again exploit some geometric notions in order to obtain an estimate. Suppose we have a simplex spanned by  $p$  endmembers  $\mathbf{e}_i, i = 1, \dots, N$ , and a point  $\mathbf{x}_i = \sum_j a_{ij} \mathbf{e}_j$ , where the abundances obey positivity and normalization. It can then be shown that  $a_{ij}$  equals the volume of the simplex spanned by  $(\mathbf{e}_1, \dots, \mathbf{e}_{j-1}, \mathbf{x}_i, \mathbf{e}_{j+1}, \dots, \mathbf{e}_p)$  divided by the volume of the simplex spanned by all endmembers. The volumes of these simplexes can again be expressed in terms of distances between the vertices by (2), and by using the shortest-path distances in the graph for this calculation, we can determine these volume ratios in the curved data manifold. We use these volume ratios as estimations for the abundances.

### 5. THE ALGORITHM

The proposed algorithm for non-linear unmixing of hyperspectral images consists of the following steps:

1. Construct a weighted symmetric nearest-neighbor graph on the data with connection coefficient  $K$ .
2. Find the largest volume simplex in the data by starting from a randomly chosen initial simplex, and iteratively replacing vertices of this simplex while larger volumes are found. Use equation (2) for the volume calculation, and the Dijkstra



**Fig. 1.** The averaged error on the estimated abundance coefficients as a function of  $\sigma$ . Solid line: The error for the linear algorithm with Euclidean distance measure. Circles: The error for the non-linear algorithm with shortest-path distances.

algorithm to find the shortest-path distances from a point to all other points. The endmembers are the vertices of the largest simplex.

3. Use these endmembers to estimate the abundances by calculating simplex volume ratios.

Every one of these steps is explained in previous sections. There are however two parameters that still have to be determined: The connectivity constant  $K$ , and the number of endmembers  $p$ . For the connectivity constant  $K$ , one should choose a number that is neither too low (because the graph will become disconnected, and the distance calculation will fail, or give inaccurate results), nor too high (the graph will become highly connected, destroying the curvature information intrinsically present in the distance matrix). Empirically, we find that a value of  $K = 20$  seems to give accurate results.

Determining the number of endmembers can be done in several ways: One can estimate the intrinsic dimensionality of the data with several dimensional reduction techniques, or employ virtual dimensionality. In the technique outlined here, we used simplex volume ratio: If the data is well described by a simplex with  $p$  endmembers, the volume of any simplex with  $p + 1$  data points will be significantly smaller (ideally zero) when compared to the volume of the simplex with  $p$  endmembers. Plotting the volume ratios between simplexes of dimension  $p - 1$  and  $p$  as a function of  $p$  will yield a curve that has a sharp drop at a certain value of  $p$ . This determines the dimensionality of the dataset, and hence the number of endmembers.

## 6. DEMONSTRATION ON ARTIFICIAL DATA

### 6.1. Construction of the dataset

We demonstrate the functioning of the algorithm on an artificial dataset, composed of non-linear mixtures of three endmembers. As endmembers, we choose three random mineral spectra from the USGS spectral library, interpolated at 50 wavelengths in the range 1.98-2.48  $\mu\text{m}$ . We generate 5000 spectral datapoints  $\mathbf{x}_i$  with these endmembers using the following equation, which is a multi-component extension of the mixing equation described in [5] for the non-linear spectral mixing of two components through secondary reflections:

$$\mathbf{x}_i = \frac{\sum_j^p a_{ij} \mathbf{e}_j + \sigma \sum_{jk} a_{ij} a_{ik} \mathbf{e}_j \mathbf{e}_k}{1 + \sigma \sum_{jk} a_{ij} a_{ik}} \quad (3)$$

The abundances  $a_{ij}$  are randomly generated, and respect the positivity and normalization constraints. The variable  $\sigma$  parametrizes the non-linearity of the mixing operation. Finally, we add the three endmembers to the collection of spectra as well, so that pure pixels are certainly present.

### 6.2. Results

We run the presented algorithm with two different distance measures: An Euclidean distance measure, yielding an algorithm closely related to the N-findR algorithm, and the graph-based nearest neighbor distance measure, incorporating the curvature of the data manifold. We choose  $K = 20$ , and the number of endmembers  $p = 3$  by design.

Next, we run these two algorithms for several values of the parameter  $\sigma$ , and compare the estimated abundance coefficients with the actual known abundance coefficients. The absolute difference between these two, averaged over all pixels, is plotted in figure (1) as a function of  $\sigma$ . Here we can clearly see that for  $\sigma = 0$ , linear unmixing yields the best results. This is to be expected, since the data can be correctly linearly unmixed, and the introduction of the shortest-path distances will certainly not improve this result. For  $\sigma$  between 0 and 1, one sees that the two algorithms yield comparable results. For larger values of  $\sigma$  however, the average error for linear unmixing keeps on rising almost linearly, while the error for the non-linear unmixing algorithm rises much less fast, indicating that the non-linear unmixing algorithm generates a much better estimation of the abundance coefficients than the linear algorithm.

## 7. CONCLUSIONS

We have presented an unsupervised unmixing algorithm that is capable to determine endmembers and their abundances in hyperspectral imagery under non-linear mixing assumptions. The algorithm is based upon the popular N-findR method, but uses distances between points in spectral space instead of the actual spectral values. The distances are defined as shortest-path distances in a nearest-neighbor graph, hereby respecting the non-trivial geometry of the data manifold in the case of non-linearly mixed pixels. Furthermore, also the abundance estimation can be expressed in terms of distances between points in spectral space via simplex volume ratios, allowing us to estimate these abundances even when the source of the non-linear mixing is unknown or cannot be modeled. The algorithm is demonstrated to perform well on artificial data, and yields much better results than an equivalent linear unmixing technique. Further investigations include a thorough comparison with other existing unmixing algorithms, testing the performance on real data, and expanding the technique to other, more efficient, unmixing algorithms (e.g. the simplex growing algorithm).

## 8. REFERENCES

- [1] J.W. Boardman, "Geometric mixture analysis of imaging spectrometry data," *Proc. Int. Geoscience and Remote Sensing Symp*, vol. 4, pp. 2369–2371, 1994.
- [2] E.M. Winter, "N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Imaging Spectrometry V*, pp. 266–277, 1999.
- [3] N. Keshava and J.F. Mustard, "Spectral unmixing," *IEEE Sig. Proc. Mag.*, vol. 19, pp. 44–57, 2002.
- [4] J.M.P. Nascimento and J.M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," *Image and Signal Processing for Remote Sensing XV*, vol. 7477, no. 1, pp. 74770I, 2009.
- [5] R.B. Singer and T.B. McCord, "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," *Proc. 10th Lunar Planetary Science Conf.*, pp. 1835–1848, 1979.
- [6] K.J. Guilfoyle, M.L. Althouse, and Chein-I Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 10, pp. 2314–2318, 2001.
- [7] J. Plaza, A. Plaza, P. Martinez, and R. Perez, "Nonlinear mixture models for analyzing laboratory simulated-forest hyperspectral data," *SPIE Image and Signal Processing for Remote Sensing*, 2003.
- [8] J. Plaza, P. Martinez, R. Perez, A. Plaza, and M.C. Cantero, "Nonlinear neural network-based mixture model for estimating the concentration of nitrogen salts in turbid inland waters using hyperspectral imagery," *SPIE Optics East Conference, Chemical and Biological Standoff Detection*, 2004.
- [9] Weiguo Liu and Elaine Y. Wu, "Comparison of non-linear mixture models: sub-pixel classification," *Remote Sensing of Environment*, vol. 94, no. 2, pp. 145–154, 2005.
- [10] J.B. Tenenbaum, V. De Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2004.
- [11] S. Rabinowitz, "The volume of an n-simplex with many equal edges," *Missouri Journal of Mathematical Sciences*, vol. 1, pp. 11–17, 1989.
- [12] E.W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.