ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Unmixing Hyperspectral Images with Fuzzy Supervised Self-Organizing Maps

T. Villmann[1], E. Merényi[2], and W.H. Farrand[3]

1- University of Appl. Sciences Mittweida - Dept. of Mathematics
Mittweida - Germany

2- Rice University Houston - Dept. of Statistics
Houston, TX - USA

3- Space Science Institute
Boulder, CO - USA

Abstract. We propose a powerful alternative to customary linear spectral unmixing, with a new neural model, which achieves locally linear but globally non-linear unmixing. This enables unmixing with respect to a large number of endmembers, while traditional linear unmixing is limited to a handful of endmembers.

1 Introduction

Linear mixture modeling ("unmixing") has been popular in the modeling of spectral imagery, for its low computational expense and easy interpretation [1]. With a few endmembers (distinct spectra representing extremes in mixing trends of the materials in the scene) the entire image can be characterized through the fractional contributions of the endmembers at a given pixel location, However, spectra interesting from the application point of view, or a larger number of endmembers can cause near-collinear situations and prevent matrix inversion that yield the fractions. In practice, linear mixture models with more than 5–6 endmembers are rare, for this reason. Hyperspectral images characterize complex scenes where unmixing with respect to a large number of endmembers is often desirable, but may not be possible with customary linear models.

2 Linear Unmixing

For a linear unmixing model of hyperspectral data we assume pixels (data vectors) $\mathbf{v} \in \mathbb{R}^{n_b}$ with n_b is the number of bands. Let N_C be the number of classes, M the number of endmembers and N the number of pixels (data). An endmember $\mathbf{em}_i = \left(em_i^1, \ldots, em_i^b, \ldots, em_i^{n_b}\right)$ is a spectral signature vector of a certain material. The linear mixing model by [1] is

$$\mathbf{v} = \sum_{i=1}^{M} f_i \mathbf{em}_i + \mathbf{r} \tag{1}$$

with f_i is the fraction of the endmember $\mathbf{em}_i$ (mixing coefficient) with the restriction $\sum_{i=1}^{M} f_i = 1$ contributing to the pixel $\mathbf{v}$ and $\mathbf{r}(\mathbf{v}) = \mathbf{r} = \left(r^1, \ldots, r^b, \ldots, r^{n_b}\right)$ is the residual vector for an given pixel $\mathbf{v}$. Depending on the mixing model, $f_i \geq 0$ may additionally be required leading to a convex problem. This limiting

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

case is not considered here in the used unmixing linear model (Spectral Mixture Analysis, SMA, [1]).

3 Nonlinear Unmixing by Fuzzy Supervised Self-Organizing Map

In the following we suggest a model for unmixing that estimates local mixing coefficients for local models, which can be non-linearly distributed in the data space. It is a prototype based semi-supervised vector quantization, whereby the prototypes are the local data models and the local mixing coefficients are learned according to a semi-supervised adaptation scheme. The underlying robust vector quantization scheme here is the self-organizing map (SOM) [3]. However, a transformation to similar robust vector quantizers like neural gas is straightforward.

3.1 The Fuzzy Supervised Self-Organizing Map (FSSOM)

The usual SOM model assumes data points $\mathbf{v} \in V \subset \mathbb{R}^{n_b}$ according to the data density $P(\mathbf{v})$. The prototypes $\mathbf{w_q}$ are assigned to an externally given (neuron) lattice A with neurons (nodes) $\mathbf{q} \in A$, whereby A is equipped with an underlying topological structure usually chosen as a regular grid and N_A is the number of nodes. Additionally, to each data point belongs a (fuzzy) class label vector $\mathbf{c}(\mathbf{v}) \in [0,1]^{N_C}$ describing the data point's degree of membership to the N_C classes. If $\sum_{i=1}^{N_C} c^i(\mathbf{v}) = 1$ holds, the labeling is called probabilistic and possibilistic otherwise. The data labeling is crisp if$\mathbf{c}(\mathbf{v}) \in \{0,1\}^{N_C}$ is valid additionally. Accordingly, $\mathbf{c}(\mathbf{w_q}) \in [0,1]^{N_C}$ denotes the class label vector of neuron $\mathbf{q}$ class assignments. We denote the grid dissimilarity between neurons $\mathbf{q}$ and $\mathbf{q}'$ by $d_A(\mathbf{q},\mathbf{q}')$, and $d(\mathbf{v},\mathbf{w_q})$ is the differentiable dissimilarity in the data space, frequently chosen as the squared Euclidean distance.

The Heskes-variant of SOM allows an gradient descent scheme of prototype adaptation according to the cost function

$$E_{\mathrm{SOM}} = \int P(\mathbf{v})\, e_{\mathbf{s}(\mathbf{v})}(\mathbf{v})\, d\mathbf{v}. \tag{2}$$

The Heskes-mapping (winner determination) rule is given by

$$\mathbf{s}(\mathbf{v}) = \mathrm{argmin}_{\mathbf{q}\in A}\left(e_{\mathbf{q}}(\mathbf{v})\right) \tag{3}$$

with local costs

$$e_{\mathbf{q}}(\mathbf{v}) = \sum_{\mathbf{q}'\in A} h_\sigma^{SOM}(\mathbf{q},\mathbf{q}')\, d(\mathbf{v},\mathbf{w}_{\mathbf{q}'}) \tag{4}$$

where $h_\sigma^{SOM}(\mathbf{q},\mathbf{q}') = \exp\left(-\frac{d_A(\mathbf{q},\mathbf{q}')}{2\sigma^2}\right)$ is the neighborhood function [2]. The respective stochastic gradient learning rule is

$$\triangle \mathbf{w_q} = -h_\sigma^{SOM}(\mathbf{q},\mathbf{s}(\mathbf{v}))\, \frac{\partial d(\mathbf{v},\mathbf{w_q})}{\partial \mathbf{w_q}}. \tag{5}$$

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

The label information is not considered in SOM (unsupervised mode). The receptive field of the prototype $\mathbf{w_q}$ is defined as $\Omega_\mathbf{q} = \{\mathbf{v} \in V | \mathbf{q} = \mathbf{s}(\mathbf{v})\}$ with $n(\mathbf{q}) = |\Omega_\mathbf{q}|$ is number of data vectors belonging to the receptive field $\Omega_\mathbf{q}$in case of finite data sets.

The Fuzzy-Supervised SOM (FSSOM) is a semi-supervised learning variant of SOM taking into account the label information $\mathbf{c}(\mathbf{v})$ of data to adapt the prototype labels [4]. Further, in contrast to simple post-labeling, the prototype adaptation is also influenced by the class information. Particularly, the FSSOM model considers a *multiplicative* combination of the label dissimilarity and the data dissimilarity in a single deviation measure $D_\varepsilon(\mathbf{v}, \mathbf{w_r}, \gamma)$ replacing the usual data dissimilarity in (4) *during model training*. This combined dissimilarity is determined according to

$$D_\varepsilon(\mathbf{v}, \mathbf{w_q}, \gamma) = D_\varepsilon^\delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q}), \gamma) \cdot D_\varepsilon^d(\mathbf{v}, \mathbf{w_q}, \gamma) - \varepsilon_\delta \varepsilon_d \tag{6}$$

with the label dissimilarity measure $D_\varepsilon^\delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q}), \gamma) = (\gamma \cdot \delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q})) + \varepsilon_\delta)$ with differentiable dissimilarity $\delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q}))$ usually chosen as the Euclidean distance. The data space dissimilarity $D_\varepsilon^d(\mathbf{v}, \mathbf{w_r}, \gamma) = ((1-\gamma) \cdot d(\mathbf{v}, \mathbf{w_r}) + \varepsilon_d)$ is based on $d(\mathbf{v}, \mathbf{w_r})$, whereby the parameter vector $\varepsilon = (\varepsilon_\delta, \varepsilon_d)$ determines a small offset term, which is necessary to prevent unexpected behavior of the FSSOM under certain conditions [4]. It turns out that $D_\varepsilon(\mathbf{v}, \mathbf{w_q}, \gamma)$ is a quasi metric [7]. The parameter $\gamma \in [0, 1]$ determines the influence of the class information with $\gamma = 0$ yielding the standard Heskes-SOM.

The FSSOM model leads to a prototype adaptation based again on the gradient of the cost function E_{SOM} (2), but which is now also influenced by the class agreement $\delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q}))$:

$$\triangle \mathbf{w_q} = -(1-\gamma) \cdot D_\varepsilon^\delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q}), \gamma) \cdot h_\sigma^{SOM}(\mathbf{q}, \mathbf{s}(\mathbf{v})) \cdot \frac{\partial d(\mathbf{v}, \mathbf{w_q})}{\partial \mathbf{w_q}} \tag{7}$$

and is accompanied by the label adaptation

$$\triangle \mathbf{c}(\mathbf{w_q}) = -\gamma \cdot D_\varepsilon^d(\mathbf{v}, \mathbf{w_q}, \gamma) \cdot h_\sigma^{SOM}(\mathbf{q}, \mathbf{s}(\mathbf{v})) \cdot \frac{\partial \delta(\mathbf{c}(\mathbf{v}), \mathbf{c}(\mathbf{w_q}))}{\partial(\mathbf{c}(\mathbf{w_q}))} \tag{8}$$

according to gradient $\frac{\partial E_{\mathrm{SOM}}}{\partial \mathbf{c}(\mathbf{w_q})}$ of the Heskes-SOM.[1] Thus, both, prototype vectors and their class assignment vectors, are parallely adjusted with subsequent renormalization of the class vectors to ensure $\mathbf{c}(\mathbf{w_q}) \in [0, 1]^{N_C}$. As the result, the FSSOM represents the data as well as their class distribution by the prototypes $\mathbf{w_q}$ and their fuzzy label $\mathbf{c}(\mathbf{w_q})$ determining the the gradual class assignments.

3.2 The Unmixing Model

We now give the description, how the FSSOM can be used as a non-linear unmixing model. The hyperspectral pixels are the data vectors. Further, endmembers

[1] At this point it is essential to use the Heskes-SOM, because it has a cost function whereas the usual SOM has not.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

are treated as class representatives with crisp labels $\mathbf{c}(\mathbf{v})$ in the FSSOM. In the following we make the trivial assumption that at least one sample per endmember is contained in the data to be analyzed, i.e. $n_i \geq 1$ holds for the overall number of pixels (samples) associated with the ith endmember $\mathbf{em}_i$. In this localized model we assume local endmembers$\mathbf{em}_i(\mathbf{q})$ associated with the ith endmember of the linear mixing model for the local area of the receptive field $\Omega_{\mathbf{q}}$. Let $n_i(\mathbf{q})$ be the number of pixels (samples) associated with the ith endmember $\mathbf{em}_i$ in the receptive field $\Omega_{\mathbf{q}}$. Averaging these pixels $\mathbf{em}_{i,j}(\mathbf{q}) = \mathbf{v}$ gives an estimate of the local endmembers, i.e. $\mathbf{em}_i(\mathbf{q}) \approx \frac{1}{n_i(\mathbf{q})} \sum_{j=1}^{n_i(\mathbf{q})} \mathbf{em}_{i,j}(\mathbf{q})$. The local linear mixing model for a prototype $\mathbf{w}_{\mathbf{q}}$ is obtained as

$$\mathbf{w}_{\mathbf{q}} = \sum_{i=1}^{M} \varphi_i(\mathbf{q}) \cdot \mathbf{em}_i(\mathbf{q}) \tag{9}$$

with local mixing coefficients (fractions) $\varphi_i(\mathbf{q})$, which are estimated by the already learned label vectors of the FSSOM, i.e. $\varphi_i(\mathbf{q}) \approx c^i(\mathbf{w}_{\mathbf{q}})$reflecting the relation $\varphi_i(\mathbf{q}) = \frac{n_i(\mathbf{q})}{n(\mathbf{q})}$. Finally, for a given pixel $\mathbf{v}$ the unmixing is obtained by the local estimates $\varphi_i(\mathbf{s}(\mathbf{v}))$ instead of the global unmixing coefficients f_i in the linear model. Hence, all pixels belonging to the same receptive field $\Omega_{\mathbf{q}}$ are treated according to the same local mixing model. Although the local unmixing is still linear, the overall new unmixing model is non-linear according to the usually non-linear mapping realized by the FSSOM mapping (3). A further improvement is achieved, if also the second winning unit $\mathbf{s}_2(\mathbf{v}) = \operatorname{argmin}_{\mathbf{q} \in A \setminus \mathbf{s}(\mathbf{v})}(e_{\mathbf{q}}(\mathbf{v}))$ is used. Interpolating between $\mathbf{s}$ and $\mathbf{s}_2$ we obtain the new estimate

$$\varphi_i(\mathbf{s}(\mathbf{v})) \approx \left(\frac{e_{\mathbf{s}_2}(\mathbf{v})}{e_{\mathbf{s}}(\mathbf{v}) + e_{\mathbf{s}_2}(\mathbf{v})}\right) c^i\left(\mathbf{w}_{\mathbf{s}(\mathbf{v})}\right) + \left(\frac{e_{\mathbf{s}}(\mathbf{v})}{e_{\mathbf{s}}(\mathbf{v}) + e_{\mathbf{s}_2}(\mathbf{v})}\right) c^i(\mathbf{w}_{\mathbf{s}_2}). \tag{10}$$

4 Application to Hyperspectral Data

We unmix, with FSSOM, a hyperspectral image of the Lunar Crater Volcanic Field (LCVF), Nevada. The data and classification into $N_C = 23$ materials classes (spectral types) are described in [5, 6]. We compare the results to a regular (globally) linear mixture model, SMA (Spectral Mixture Analysis). Training spectra (endmembers) for both SMA and FSSOM ($N_A = 10 \times 5$) are the same, and multiple spectra represent each endmembers. Fig1 shows a comparison of 3-endmember mixtures ($M = 3$) from the two methods. We show 2 out of three endmember fraction images for reasons of space limitation. While the SMA exhibits more nuances, the high and low fractions are clearly at matching locations. In this experiment we used the top two winners for interpolation as in eq (10). Including more of the top winners in the interpolation will increase the nuances in the FSSOM output as well. The more substantive differences are due to the fact that the FSSOM keeps the fractions between 0 and 1, by virtue of better, globally non-linear, model fit whereas the globally linear, rigid SMA model is bound to have large negative or overpositive fractions. (These are clipped to 0 and 1, respectively, in the images shown, in order to get a direct visual comparison with the FSSOM fractions, and show larger areas of extreme values

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

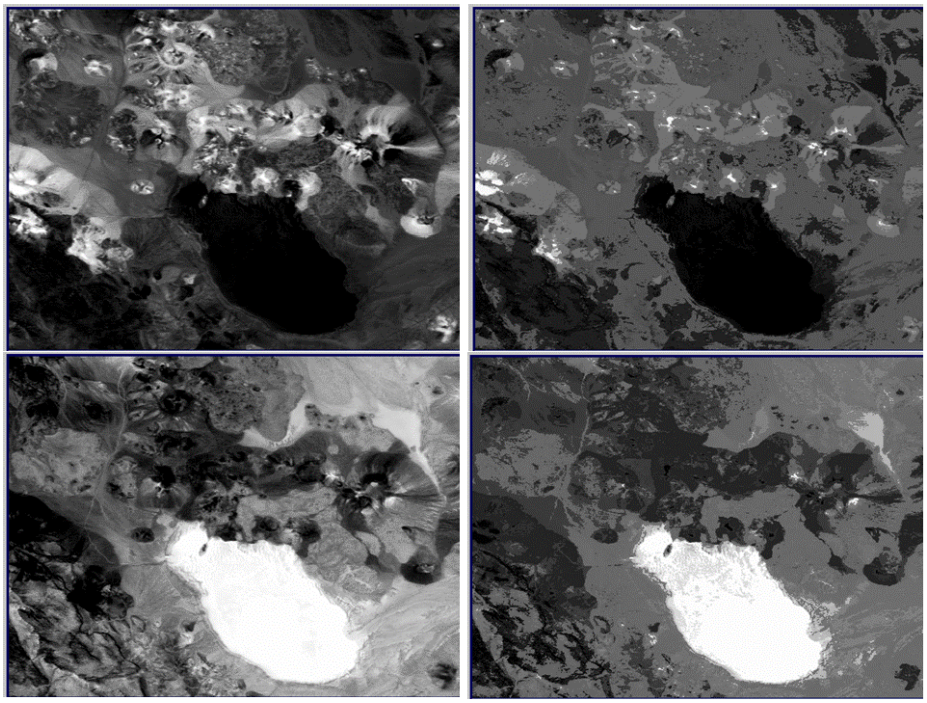

Figure 1: Fraction images of the Lunar Crater Volcanic Field site. Left column: fractions from linear unmixing; Right column: fractions produced by FSSOM. The top row shows the “cinder", bottom row shows the “playa" fractions.

than the SMA. Since the SMA puts no constraint on the fraction values (only that they add to 1), a few extreme fraction values (caused by poor model fit at those pixels) can cause an image scaling that makes visual comparison difficult. It is not trivial, however, how to clip and scale image values for straightforward, automatic comparison. This is subject of near-future investigation. The third fraction image (vegetation, not shown here)), is equally matching its SMA counterpart.

Encouraged by this result, we computed a 23-endmember mixture with FSSOM ($N_A = 15 \times 5$), where the $M = 23$ endmembers are defined by training pixels used in an earlier classification of the LCVF image [5]. There is no SMA equivalent since the SMA cannot handle such large number of endmembers, many of which are very similar to one another. However, we can compare the fraction images to the classification, arguing that the pixels with the highest fractions should delineate the locations of the material class that is characterized by the respective endmember spectra. Figure 2 indicates that this is indeed the case.

Future work needs to investigate the differences between SMA and FSSOM results, to understand where the FSSOM produces better or poorer model fit,

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

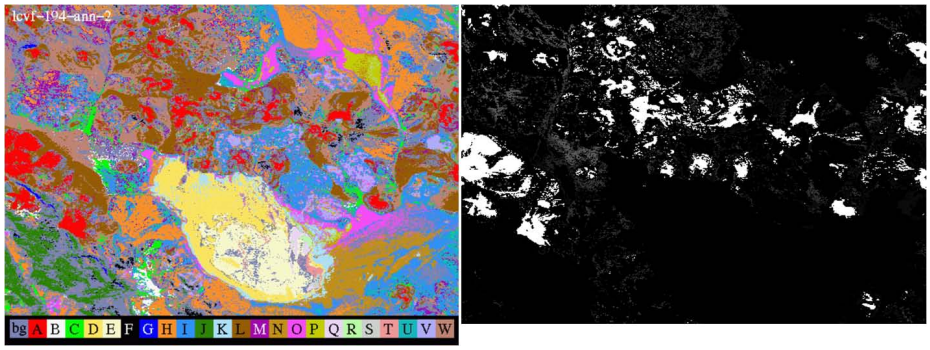

Figure 2: Unmixing of the LCVF image with 23 endmembers. Left: classification map from [6]. Right: fractions produced by FSSOM, for the "Cinder" endmember. In this case, only the top winner was used to compute the FSSOM fractions. The high-fraction locations match the Cinder class (A, red) very well. The rest of the 23 fraction images are similarly good matches to the corresponding classes.

and why. Overall, FSSOM provides an interesting non-linear alternative for unmixing based on local linear unmixing. In particular, problems with many endmembers can be handled by FSSOM, which is a significant improvment compared to customary linear spectral unmixing.

References

[1] J. B. Adams, M. O. Smith, and A. R. Gillespie. Imaging spectroscopy: Interpretation based on spectral mixture analysis. In C. Peters and P. Englert, editors, *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pages 145–166. Cambridge University Press, New York, 1993.

[2] T. Heskes. Energy functions for self-organizing maps. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 303–316. Elsevier, Amsterdam, 1999.

[3] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[4] M. Kästner and T. Villmann. Fuzzy supervised neural gas for semi-supervised vector quantization - theoretical aspects. *Machine Learning Reports*, 5(MLR-02-2011):1–12, 2011. ISSN:1865-3960, http://www.techfak.uni-bielefeld.de/~fschleif/mlr/mlr_02_2011.pdf.

[5] E. Merényi. Precision mining of high-dimensional patterns with self-organizing maps: Interpretation of hyperspectral images. In *Quo Vadis Computational Intelligence: New Trends and Approaches in Computational Intelligence (Studies in Fuzziness and Soft Computing, Vol 54, P. Sincak and J. Vascak Eds.)*. Physica Verlag, 2000.

[6] E. Merényi, W. Farrand, J. Taranik, and T. Minor. Classification of hyperspectral imagery with neural networks: Comparison to conventional tools. In T. Villmann and F.-M. Schleif, editors, *Machine Learning Reports*, volume 5, pages 1–15, 2011. ISSN:1865-3960. on-line http://www.techfak.uni-bielefeld.de/fschleif/mlr/mlr_04_2011.pdf. Also submitted to EURASIP Journal on Advances in Signal Processing.

[7] E. Pekalska and R. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*. World Scientific, 2006.