DOI: 10.58895/ksp/1000128686-12 erschienen in:
OCM 2021 - 5th International Conference on Optical Characterization of Materials, March 17th – 18th, 2021, Karlsruhe, Germany : Conference Proceedings
DOI: 10.58895/ksp/1000128686 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000128686/

Generation of artificial training data for spectral unmixing by modelling spectral variability using Gaussian random variables

Johannes Anastasiadis and Michael Heizmann

Institute of Industrial Information Technology (IIIT),
Karlsruhe Institute of Technology (KIT),
Hertzstr. 16, 76187 Karlsruhe, Germany

Abstract A stochastic method how artificial training data for spectral unmixing can be generated from real pure spectra is presented. Since the pure spectra are modelled as Gaussian random vectors, spectral variability is also considered. These training data can in turn be used to train an artificial neural network for spectral unmixing. Non-negativity and sum-to-one constraints are enforced by the network architecture. The approach is evaluated using real mixed spectra and achieves promising results with the used datasets.

Keywords Spectral unmixing, spectral variability, Gaussian random variables, convolutional neural network

1 Introduction

Optical measurement techniques are often used for process monitoring in industry because they are non-contact and non-destructive. An important task is to determine the relative proportions of the components in substance mixtures. There are applications in many fields, such as food industry, medical technology, as well as in the processing of bulks. This task cannot be solved sufficiently with conventional colour images, because these only contain three colour channels (red, green, blue) per pixel, whereas hyperspectral images have a finely gained spectrum in each pixel that characterizes the materials much better [1]. If information of different materials is contained in one measured pixel,

spectral unmixing (SU) is needed to get the relative proportions, the abundances, of the pure materials in the pixel [2]. This is often done model-based [3]. However, these models usually assume a single spectrum for each pure substance involved. Actually, the spectra of the pure substances and also of the mixtures vary quite a lot [4,5].

Artificial neural networks (ANNs) achieve excellent results in many domains and have the benefit of not requiring model knowledge. Particularly for SU, the use of ANNs has further advantages [6]. First, the non-negativity and the sum-to-one constraints can be enforced by a normalising output layer. Second, spectral variability can be considered if it is contained in the training data. However, ANNs need a lot of significant training data to perform well, which are often not available in the domain of hyperspectral imaging.

In this approach we model the mixing characteristics including spectral variability of substances and use these models to generate training data for an ANN used for SU. Several spectra of each pure substance are needed for this approach. Even without the availability of real mixed spectra, most of the advantages of SU using an ANN can be exploited. Furthermore, performance can be improved if additional mixed spectra are available. This approach is therefore suitable for use in an industrial environment where the pure substances involved are known and hyperspectral images can be acquired in advance. We have already trained ANNs with artificially generated mixed spectra in a preceding work [7]. There, pure spectra randomly drawn from a set were mixed using models to obtain mixed spectra. In contrast, the approach in this paper models the spectra of the pure substances as Gaussian distributed random vectors. There is another approach, where SU is accomplished by directly applying Gaussian process regression [8], but not for training data generation. In this contribution, to generate the training data, the random vectors are combined for many different sets of abundances using the linear mixing model (LMM), the Fan Model (FM), the generalized bilinear model (GBM), and the linear quadratic model (LQM) [2,9–11].

The rest of the paper is structured as follows: In Section 2 the basics of SU are summarized. Section 3 then provides the probabilistic equivalents of the mixing models used. In Section 4 the experimental results are shown. Finally the paper is summarized and conclusions are drawn in Section 5.

2 Spectral unmixing

The estimation of the abundances in a mixture of substances based on its spectrum is called SU [2]. If the spectra of the pure substances involved are known, the term supervised SU is used. This is the case here. The estimation $\hat{\mathbf{a}} = [\hat{a}_1, ..., \hat{a}_P]^{\mathrm{T}} \in \mathbb{R}^P$ of the true abundances $\mathbf{a} = [a_1, ..., a_P]^{\mathrm{T}} \in \mathbb{R}^P$ of the up to P pure substances involved has to fulfil constraints in order to ensure physical validity. These are the non-negativity constraint and the sum-to-one constraint:

$$\hat{a}_p \geq 0 \quad \forall p\,, \qquad \sum_{p=1}^{P} \hat{a}_p = 1\,. \tag{2.1}$$

Many of the methods used for SU are model-based. The models used here approximate the mixed spectra using a parametric function. The most commonly used mixing model is the LMM, which works well for many applications [2,5,12,13]:

$$\mathbf{y} = \sum_{p=1}^{P} \mathbf{m}_p\, a_p + \boldsymbol{\omega} = \mathbf{M}\,\mathbf{a} + \boldsymbol{\omega}\,. \tag{2.2}$$

Here $\mathbf{y} \in \mathbb{R}^{\Lambda}$ is a discrete spectrum sampled at Λ wavelength channels and $\mathbf{M} = [\mathbf{m}_1, ..., \mathbf{m}_P] \in \mathbb{R}^{\Lambda \times P}$ are the spectra of the up to P involved pure substances. Differences between $\mathbf{y}$ and the weighted sum of the pure material spectra are considered by $\boldsymbol{\omega} \in \mathbb{R}^{\Lambda}$. The LMM is based on the assumptions that mixing takes place on a macroscopic scale and that photons interact with only one material before they hit the sensor [14]. There are also non-linear mixing models that are summarized in [3]. In this paper the GBM [10]

$$\mathbf{y} = \sum_{p=1}^{P} \mathbf{m}_p\, a_p + \sum_{p=1}^{P-1} \sum_{q=p+1}^{P} \gamma_{pq}\, a_p\, a_q\, \mathbf{m}_p \odot \mathbf{m}_q + \boldsymbol{\omega} \tag{2.3}$$

and the LQM [11]

$$\mathbf{y} = \sum_{p=1}^{P} \mathbf{m}_p\, a_p + \sum_{p=1}^{P} \sum_{q=1}^{P} b_{pq}\, \mathbf{m}_p \odot \mathbf{m}_q + \boldsymbol{\omega} \tag{2.4}$$

are used, where γ_{pq} and b_{pq} are the non-linearity coefficients and $\odot$ is the element-wise product. For $\gamma_{pq} = 1 \; \forall p, q$ the GBM is equivalent to the FM [9]. These mixing models also consider light that has interacted with up to two substances before hitting the sensor.

A commonly used approach considering (2.1) is the Fully Constrained Least Squares (FCLS) algorithm [15]. Here, the Lagrangian $L : \mathbb{R}^{P+1} \to \mathbb{R}$ with Lagrange multiplier $l \in \mathbb{R}$ is optimized:

$$L(\mathbf{a}, l) = \|\mathbf{y} - \mathbf{M}\,\mathbf{a}\|_2^2 - l \left(\sum_{p=1}^{P} a_p - 1 \right). \tag{2.5}$$

It is an iterative process that removes negative $\hat{a}_p$ and the corresponding pure spectra. For estimations based on non-linear mixing models, gradient based line search methods can be used [10].

The presented mixing models assume that each pure substance is represented by a single spectrum. In reality, however, pure substances may have varying spectra [4]. These variations are referred to as spectral variability, which is primarily caused by surface structure. There are also unmixing algorithms considering the spectral variability such as the extended linear mixing model (ELMM) [16]. The ELMM extends the LMM by a set of parameters that allow scaling of the pure material spectra.

The presented approach considers spectral variability by including it in the generated training data. The process of generating these data is described in the following section.

3 Stochastic mixing models

For the approach used in this paper the spectra $\mathbf{m}_p$ of the pure substances are modelled as Gaussian distributed random variables $\boldsymbol{M}_p \in \mathbb{R}^{\Lambda}$. These can be entirely described by their mean vector $\boldsymbol{\mu}_{\boldsymbol{M}_p} = \boldsymbol{\mu}_p \in \mathbb{R}^{\Lambda}$ and their covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{M}_p,\boldsymbol{M}_p} = \boldsymbol{\Sigma}_{p,p} \in \mathbb{R}^{\Lambda \times \Lambda}$. Mean vectors and covariance matrices of pure spectra are estimated using real data, which is why a set of measured spectra is required for each pure substance. Since the pure spectra of different pure materials do not depend on each other, they are assumed to be stochastically independent and therefore the cross-covariance matrix is $\boldsymbol{\Sigma}_{p,q} = \mathbf{0}$ if

$p \neq q$. A mixed spectrum $\boldsymbol{Y} \in \mathbb{R}^{\Lambda}$ is also modelled as a Gaussian distributed random vector with mean vector $\boldsymbol{\mu}_{\boldsymbol{Y}} \in \mathbb{R}^{\Lambda}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{Y},\boldsymbol{Y}} \in \mathbb{R}^{\Lambda \times \Lambda}$. The following moments result for the LMM:

$$\boldsymbol{\mu}_{\boldsymbol{Y}} = \sum_{p=1}^{P} a_p \, \boldsymbol{\mu}_p \,, \qquad \boldsymbol{\Sigma}_{\boldsymbol{Y},\boldsymbol{Y}} = \sum_{p=1}^{P} a_p^2 \, \boldsymbol{\Sigma}_{p,p} \,. \tag{3.1}$$

Appropriate non-linearity coefficients γ_{pq} and b_{pq} must be determined for the GBM and the LQM. In this paper a constant value is used for this purpose, which means for all p and q: $\gamma_{pq} = \gamma$ and $b_{pq} = b$. This is a limitation, but in practice it allows better results compared to the LMM and the FM. To determine suitable values, a small validation dataset with some real mixed spectra is required. The mean vector $\boldsymbol{\mu}_{\boldsymbol{Y}}$ for the GBM can be obtained by replacing $\mathbf{m}_p$ by $\boldsymbol{\mu}_p$ in (2.3). The following covariance matrix results for the GBM:

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{Y},\boldsymbol{Y}} = & \sum_{p=1}^{P} a_p^2 \, \boldsymbol{\Sigma}_{p,p} + \sum_{p=1}^{P-1} \sum_{q=p+1}^{P} \gamma^2 \, a_p^2 \, a_q^2 \, \boldsymbol{\Sigma}_{p\odot q,p\odot q} \\ & + \sum_{p=1}^{P} \sum_{\substack{q=1 \\ q\neq p}}^{P} \gamma \, a_p^2 \, a_q \, \left(\boldsymbol{\Sigma}_{p,p\odot q} + \boldsymbol{\Sigma}_{p\odot q,p} \right) \\ & + \sum_{p=1}^{P} \sum_{\substack{q=1 \\ q\neq p}}^{P-1} \sum_{\substack{l=q+1 \\ l\neq p}}^{P} \gamma^2 \, a_p^2 \, a_q \, a_l \, \left(\boldsymbol{\Sigma}_{p\odot q,p\odot l} + \boldsymbol{\Sigma}_{p\odot l,p\odot q} \right) . \end{aligned} \tag{3.2}$$

The (cross-)covariance matrices are calculated by:

$$\begin{aligned} \boldsymbol{\Sigma}_{p\odot q,p\odot q} &= \mathcal{D}(\boldsymbol{\mu}_p) \, \boldsymbol{\Sigma}_{q,q} \, \mathcal{D}(\boldsymbol{\mu}_p) + \mathcal{D}(\boldsymbol{\mu}_q) \, \boldsymbol{\Sigma}_{p,p} \, \mathcal{D}(\boldsymbol{\mu}_q) + \boldsymbol{\Sigma}_{p,p} \odot \boldsymbol{\Sigma}_{q,q} \,, \\ \boldsymbol{\Sigma}_{p,p\odot q} + \boldsymbol{\Sigma}_{p\odot q,p} &= \boldsymbol{\Sigma}_{p,p} \, \mathcal{D}(\boldsymbol{\mu}_q) + \mathcal{D}(\boldsymbol{\mu}_q) \, \boldsymbol{\Sigma}_{p,p} \,, \\ \boldsymbol{\Sigma}_{p\odot q,p\odot l} + \boldsymbol{\Sigma}_{p\odot l,p\odot q} &= \mathcal{D}(\boldsymbol{\mu}_q) \, \boldsymbol{\Sigma}_{p,p} \, \mathcal{D}(\boldsymbol{\mu}_l) + \mathcal{D}(\boldsymbol{\mu}_l) \, \boldsymbol{\Sigma}_{p,p} \, \mathcal{D}(\boldsymbol{\mu}_q) \,. \end{aligned} \tag{3.3}$$

Here, the output of operator $\mathcal{D}(\,\cdot\,)$ is a diagonal matrix with the elements of the input vector as its diagonal elements. If the input is a matrix, the output is a vector whose elements are the diagonal elements of the input matrix. The moments for the FM can be obtained by setting $\gamma = 1$.

The LQM also takes element-wise squared spectra into account. The following mean vector formula results for the LQM after transforming:

$$\boldsymbol{\mu}_Y = \sum_{p=1}^{P} a_p \boldsymbol{\mu}_p + b \boldsymbol{\mu}_{p\odot p} + \sum_{p=1}^{P-1} \sum_{q=p+1}^{P} 2\, b \boldsymbol{\mu}_p \odot \boldsymbol{\mu}_q \,, \tag{3.4}$$

where $\boldsymbol{\mu}_{p\odot p} = \boldsymbol{\mu}_p \odot \boldsymbol{\mu}_p + \mathcal{D}(\boldsymbol{\Sigma}_{p,p})$. The covariance matrix of a mixed spectrum is not shown for the LQM because it is similar to the one of the GBM. As for the other mixing models, all covariance and cross-covariance matrices are taken into account.

Using the Gaussian random variables calculated with the mixing models, training data for an ANN can now be generated. If training is performed over several epochs, new samples of mixed spectra can be drawn in every epoch. After data sampling, the spectra generated with non-linear mixing models are normalized. For this purpose they are multiplied by d_{GBM} or d_{LQM}, respectively:

$$d_{\mathrm{GBM}} = \left(1 + \sum_{p=1}^{P-1} \sum_{q=p+1}^{P} \gamma\, a_p\, a_q\right)^{-1}, \quad d_{\mathrm{LQM}} = \left(1 + P^2\, b\right)^{-1}. \tag{3.5}$$

This is necessary because, unlike the LMM, the prefactors do not add up to 1. However, since the light that contributes to the linear component cannot contribute additionally to the quadratic component, this normalization is useful [7]. Thus, values for γ and b greater than 1 are also reasonable. In the next section the described approach is evaluated with real hyperspectral datasets.

4 Experimental results

To evaluate our approach we use real hyperspectral data acquired in our image processing lab. All datasets have 91 wavelength channels with an average width of 4 nm from 450 nm to 810 nm and 400 spectra per mixture. Two of the datasets consist of mixtures of coloured quartz sand. The first of them (quartz-3) contains 45 mixtures of maximum 3 components including pure substances, which vary in abundance steps of 0.125 . The second one (quartz-4) contains 56 mixtures of maximum

4 components including pure substances, which vary in abundance steps of 0.2 . The third dataset consists of 56 mixtures of colour powders (colour-4), which also have a maximum of 4 components each. Here, the components are varied in steps of 0.2, too. Only pure spectra are used for training. For all datasets, 12 mixtures are used for validation and the rest (30 respectively 40) to test. We train our ANN using the artificial mixtures generated with the approach in Section 3. The validation datasets are used to determine good non-linear coefficients for the GBM and the LQM. After training we test the ANN with the real mixtures in our test datasets. The root-mean-square error

$$\Delta_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\frac{1}{P}\sum_{p=1}^{P}(\hat{a}_{pn} - a_{pn})^2}\,, \tag{4.1}$$

based on all N spectra including all mixtures of a test dataset, is used as a measure of performance. The results are compared to the FCLS algorithm and the ELMM based SU.

The ANN we used for the evaluation is the convolutional neural network (CNN) from [17], which is the one dimensional version of [7]. It has three convolutional layers and two fully connected layers. The length of the convolution kernels is 3 and the numbers of feature maps from input layer to output layer are 1, 16, 32, 64, 64, and 1. The number of epochs depends on the dataset. Therefore, the CNN is trained with quartz-3 dataset for 251 epochs, quartz-4 dataset for 61, and colour-4 dataset for 31 epochs.

Different training datasets are generated using all presented mixing models and different step sizes $\mathrm{s} \in \mathbb{R}$ for the abundances $\mathbf{a}$ (see Fig. 4.1). In every epoch there are 400 spectra per mixture drawn from the previously determined Gaussian random vectors of the mixtures. Figure 4.1 shows the results for the different methods. The prefix CNN indicates that a CNN was trained for SU in order to get this result. The training data were generated with the corresponding stochastic mixing model.

It is evident that in almost all cases the presented approach leads to an improvement of the results compared to the FCLS or the ELMM based approach. This is due to spectral variability being modelled based on data and taken into account by the CNN. Which mixing

model yields the best results depends on the unmixing task. In the datasets used here the non-linear mixing models perform better. The FM delivers quite good results, although like LMM, it does not need any additional parameters. The colour-4 dataset shows, however, that the choice of the right mixing model can achieve a significant improvement. Smaller step sizes and thus more large training datasets lead in most cases to an improvement.

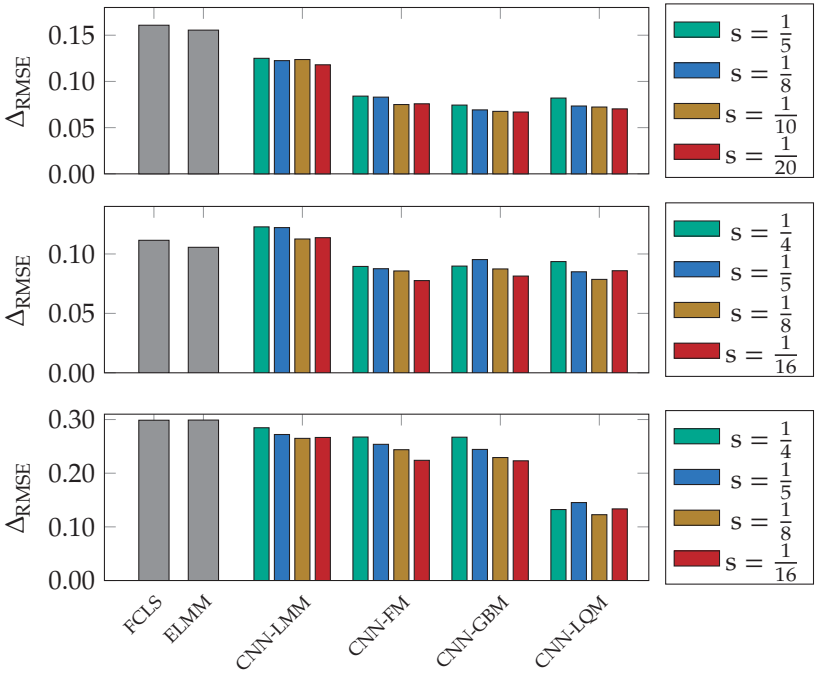

Figure 4.1: Root-mean-square error of the abundances for the compared methods applied on the quartz-3 (top), the quartz-4 (middle), and the colour-4 (bottom) dataset, respectively.

5 Summary

In this paper an approach is presented where artificial mixed spectra are generated using stochastic mixing models. These mixed spectra can then be used to train a CNN for SU. In this way, spectral variability is considered. Compared to methods from literature, which in part also include spectral variability, better results can be achieved with regard to the error of the estimated abundances. The choice of the mixing model in dependence of the problem significantly influences the quality of the estimation. Finer step sizes in the specified abundances for the training data can lead to an additional improvement.

In the future, the approach could be extended in such a way that the mean vectors and covariance matrices of the random vectors of mixed spectra are determined based on data instead of models. However, larger training datasets would be needed to train these networks.

References

1. A. Gowen, C. O'Donnell, P. Cullen, G. Downey, and J. Frias, "Hyperspectral imaging – an emerging process analytical tool for food quality and safety control," *Trends in Food Science & Technology*, vol. 18, no. 12, pp. 590–598, 2007.
2. N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE signal processing magazine*, vol. 19, no. 1, pp. 44–57, 2002.
3. N. Dobigeon, Y. Altmann, N. Brun, and S. Moussaoui, "Linear and nonlinear unmixing in hyperspectral imaging," in *Data Handling in Science and Technology*, C. Ruckebusch, Ed. Elsevier, 2016, vol. 30, pp. 185–224.
4. R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare, and C. Jutten, "Spectral variability in hyperspectral data unmixing: A comprehensive review," *arXiv preprint arXiv:2001.07307*, 2020.
5. W. Krippner, S. Bauer, and F. Puente León, "Considering spectral variability for optical material abundance estimation," *tm – Technisches Messen*, vol. 85, no. 3, pp. 149–158, 2018.
6. J. Anastasiadis and F. Puente León, "Spatially resolved spectral unmixing using convolutional neural networks (German paper)," *tm – Technisches Messen*, vol. 86, no. s1, pp. 122–126, 2019.

7. J. Anastasiadis, P. Benzing, and F. Puente León, "Generation of artificial data sets to train convolutional neural networks for spectral unmixing (German paper)," *tm – Technisches Messen*, vol. 87, no. 9, pp. 542–552, 2020.
8. Y. Altmann, N. Dobigeon, S. McLaughlin, and J. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2442–2453, 2013.
9. W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2951–2962, 2009.
10. A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, 2011.
11. I. Meganem, P. Déliot, X. Briottet, Y. Deville, and S. Hosseini, "Linear–quadratic mixing model for reflectances in urban environments," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 544–558, 2013.
12. S. Bauer, J. Stefan, and F. Puente León, "Hyperspectral image unmixing involving spatial information by extending the alternating least-squares algorithm," *tm – Technisches Messen*, vol. 82, no. 4, pp. 174–186, 2015.
13. W. Krippner and F. Puente León, "Band selection and estimation of material abundances using spectral filters (German paper)," *tm – Technisches Messen*, vol. 85, no. 6, pp. 454–467, 2018.
14. R. N. Clark and T. L. Roush, "Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research: Solid Earth*, vol. 89, no. B7, pp. 6329–6340, 1984.
15. D. Heinz, C.-I. Chang, and M. L. Althouse, "Fully constrained least-squares based linear unmixing," in *IEEE 1999 International Geoscience and Remote Sensing Symposium.*, vol. 2. IEEE, 1999, pp. 1401–1403.
16. M. A. Veganzones, L. Drumetz, G. Tochon, M. Dalla Mura, A. Plaza, J. Bioucas-Dias, and J. Chanussot, "A new extended linear mixing model to address spectral variability," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2014, pp. 1–4.
17. J. Anastasiadis and M. Heizmann, "CNN-based augmentation strategy for spectral unmixing datasets considering spectral variability," in *Image and Signal Processing for Remote Sensing XXVI*, L. Bruzzone, F. Bovolo, and

E. Santi, Eds., vol. 11533, International Society for Optics and Photonics. SPIE, 2020, pp. 188–199.