HAL
open science

Nonlinear unmixing of hyperspectral images based on multi-kernel learning

Jie Chen, Cédric Richard, Paul Honeine

To cite this version:

Jie Chen, Cédric Richard, Paul Honeine. Nonlinear unmixing of hyperspectral images based on multi-kernel learning. Proc. IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), 2012, Shanghai, China. pp.1-4, 10.1109/WHISPERS.2012.6874231 . hal-01966012

HAL Id: hal-01966012

https://hal.science/hal-01966012

Submitted on 27 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NONLINEAR UNMIXING OF HYPERSPECTRAL IMAGES BASED ON MULTI-KERNEL LEARNING

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

[(1)] Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, France

[(2)] Université de Technologie de Troyes, CNRS, France

E-mail: jie.chen@utt.fr, cedric.richard@unice.fr, paul.honeine@utt.fr

ABSTRACT

Nonlinear unmixing of hyperspectral images has generated considerable interest among researchers, as it may overcome some inherent limitations of the linear mixing model. In this paper, we formulate the problem of estimating abundances of a nonlinear mixture of hyperspectral data based on a new multi-kernel learning paradigm. Experiments are conducted using both synthetic and real images in order to illustrate the effectiveness of the proposed method.

***Index Terms*—** Hyperspectral image, nonlinear unmixing, multi-kernel learning

1. INTRODUCTION

Spectral unmixing is an important issue to analyze remotely sensed hyperspectral data, due to the relatively low spatial resolution of images. This involves the decomposition of each mixed pixel into a combination of pure endmember spectra, and the estimation of the abundance value for each endmember [1]. Under the assumption that the endmembers have been identified, hyperspectral image unmixing thus reduces to an inverse problem for estimating the abundances. To be physically interpretable, the latter are required to satisfy two constraints: they must be nonnegative and their sum equal to one.

Although the linear mixture model has been widely used due to its simple physical interpretation and tractable estimation process, there are many situations in which it may not be appropriate and could be advantageously replaced by a nonlinear one. Nonlinear unmixing has generated considerable interest among researchers, and different methods have been proposed to account for nonlinear effects. Using supervised approaches, such as neural network training, is a way to bypass difficulties that come along with unknown complex mixing mechanism [2, 3]. Unsupervised approaches that do not require training samples have also been extensively studied in the literature. In [4], a nonlinear unmixing algorithm for generalized bilinear mixture model has been proposed based on Bayesian inference. In [5, 6], the authors have extended the collection of endmembers by adding artificial cross-terms of pure spectral signatures to model light scattering effects on different materials. In [7], the authors have addressed the nonlinear unmixing problem with an intimate mixture model.

Nonlinear algorithms operating in reproducing kernel Hilbert spaces (RKHS) have also been investigated. In [8], the authors have proposed a kernelization of the FCLS algorithm [9]. Unfortunately, even though this approach considers nonlinear distortions of spectral signatures, it does not explicitly include nonlinear interactions of the endmember spectra. To overcome this drawback, in our previous work [10], we have proposed a new framework based on reproducing kernel machinery in order to possibly take nonlinear photon interactions into account. The family of mixing models that was considered in that paper are partially linear models, which combines a linear mixing model parameterized by the abundances to be estimated, with a nonlinear nonparametric function in a RKHS. In this paper, we extend this approach by formulating the problem as a multi-kernel learning paradigm, in order to optimize the tradeoff between the linear and nonlinear terms in the mixing model. Experiments are conducted using both synthetic and real images in order to illustrate the effectiveness of the proposed method.

2. PRESENTATION OF THE ALGORITHM

2.1. Notations

Suppose that in a given hyperspectral image, there exist R significant endmembers with spectral signature $\boldsymbol{m}_i \in \mathbb{R}^L$, where L denotes the number of spectral bands. Note that, usually, we have $R \ll L$. Let $\boldsymbol{r} \in \mathbb{R}^L$ be an observed hyperspectral pixel, and $\boldsymbol{\alpha} \in \mathbb{R}^R$ the unknown abundance fractions associated to the latter. Let $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R] \in \mathbb{R}^{L \times R}$ be the matrix of the endmembers. For the sake of convenience, the ℓ-th row of $\boldsymbol{M}$ is denoted by $\boldsymbol{m}_{\lambda_\ell}^\top \in \mathbb{R}^L$, that is, the vector of the endmember signatures at the ℓ-th wavelength band.

2.2. Formulation of the nonlinear unmixing problem

There are many situations where it is inappropriate to assume a linear model for hyperspectral data unmixing. See, e.g., the intimate mixture model [1]. This motivate us to consider

general mixing mechanisms of the form

$$\boldsymbol{r} = \boldsymbol{\Psi}(\boldsymbol{\alpha}, \boldsymbol{M}) + \boldsymbol{n} \tag{1}$$

with $\boldsymbol{\Psi}$ an unknown nonlinear function that defines the interactions between the endmembers in matrix $\boldsymbol{M}$, and $\boldsymbol{n}$ a measurement noise vector. To address the estimation of the abundance vector $\boldsymbol{\alpha}$, we suggest to consider the following inverse problem

$$\psi^* = \arg\min_{\psi \in \mathcal{H}} \frac{1}{2}\|\psi\|^2_{\mathcal{H}} + \frac{1}{2\mu}\sum_{\ell=1}^{L}(r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}))^2 \tag{2}$$

with $\mathcal{H}$ a given functional space, and μ a positive parameter that controls the trade-off between regularization and fitting. Clearly, this approach may fail if the functionals ψ of $\mathcal{H}$ cannot be adequately and finitely parameterized. Kernel-based methods can lead to efficient and accurate resolution of the inverse problem (2). In order to extract the mixing ratios of the endmembers, we define the function ψ in (2) as a partially linear model, which combines a linear mixing model parameterized by the abundances to be estimated, with a nonlinear nonparametric function, namely,

$$\begin{aligned} \psi(\boldsymbol{m}_{\lambda_\ell}) &= \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} + \psi_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}) \\ &\text{subject to} \quad \boldsymbol{\alpha} \succeq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{\alpha} = 1 \end{aligned} \tag{3}$$

with ψ_{nlin} any real-valued function on a compact $\mathcal{M}$, found from an RKHS denoted by $\mathcal{H}_{\text{nlin}}$. Let κ_{nlin} be its reproducing kernel. The constraints in (3) ensure that the abundance fractions are nonnegative, and that their sum is equal to one. We propose to conduct hyperspectral data unmixing by solving the following optimization problem

$$\begin{aligned} \psi^*, u^* = &\arg\min_{\psi, u} \frac{1}{2}\left(\frac{1}{u}\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} + \frac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) \\ &+ \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ &\text{where} \quad \psi = \psi_{\text{lin}} + \psi_{\text{nlin}} \\ &\text{with} \quad \psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{h}^\top \boldsymbol{m}_{\lambda_\ell} \\ &\text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \quad \text{and} \quad \boldsymbol{h} \succeq 0 \\ &\qquad\qquad 0 \le u \le 1 \end{aligned} \tag{4}$$

where u is used to optimize the tradeoff between the linear model ψ_{lin} and the nonlinear nonparametric function ψ_{nlin}. Learning simultaneously the parameter u, and functions ψ_{lin} and ψ_{nlin}, in a single optimization problem is known as the multi-kernel learning problem. See [11] and references therein. The rest of this section is devoted to the formulation and resolution of a convex optimization problem. Comparing equations (3) and (4), it is important to note that the sum-to-one constraint $\boldsymbol{1}^\top \boldsymbol{h} = 1$ has been given up due to the antagonistic action of the parameter u. To get $\boldsymbol{\alpha}$ after reaching convergence, an explicit normalization of $\boldsymbol{h}$ by the sum of its entries can be performed.

2.3. Convexity of the problem

Problem (4) is a convex optimization problem by virtue of the convexity of the function $f(u, \psi) = \|\psi\|^2_{\mathcal{H}}/u$ over $\mathcal{H} \times \mathbb{R}^*_+$. See [12]. This allows us to formulate the following two-stage optimization problem, with respect to ψ and u successively, in order to solve problem (4)

$$\min_u J(u) \qquad \text{subject to} \qquad 0 \le u \le 1 \tag{5}$$

where $J(u)$ is defined by (6), see next page.

The function $J(u)$ in subproblem (6) is defined as the optimum of a family of convex functions in (u, ψ), subject to convex constraints over ψ. As proven in [12], it turns out that $J(u)$ is convex in u and, as a consequence, that the constrained optimization problem (5) is convex.

2.4. Solution to the problem

By the strong duality property, we can derive a dual problem that has the same solution $J(u) = F(u, \psi^*)$ as the primal problem (6). Let us introduce the Lagrange multipliers β_ℓ and γ_r. The Lagrange function associated with problem (6) can be written as

$$\begin{aligned} G =& \frac{1}{2}\left(\frac{1}{u}\|\boldsymbol{h}\|^2 + \frac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ &- \sum_{\ell=1}^{L} \beta_\ell \left(e_\ell - r_\ell + \psi(\boldsymbol{m}_{\lambda_\ell})\right) - \sum_{r=1}^{R} \gamma_r\, h_r \end{aligned} \tag{8}$$

with $\gamma_r \ge 0$, where we have used $\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} = \|\boldsymbol{h}\|^2$. The conditions for optimality of G with respect to the primal variables are given by

$$\begin{cases} \boldsymbol{h}^* = u\left(\sum_{\ell=1}^{L} \beta_\ell^* \boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}^*\right) \\ \psi^*_{\text{nlin}} = (1-u)\sum_{\ell=1}^{L} \kappa_{\text{nlin}}(\cdot, \boldsymbol{m}_{\lambda_\ell})\, \beta_\ell^* \\ e_\ell^* = \mu\, \beta_\ell^* \end{cases} \tag{9}$$

As expected, note from equation (9) that u and $1-u$ control proportions of the linear and nonlinear components in the model. By substituting equation (9) into (8), we get the dual problem (11), see next page, where $\boldsymbol{K}_{\text{nlin}}$ is the Gram matrix associated with ψ_{nlin}, defined by

$$[\boldsymbol{K}_{\text{nlin}}]_{\ell,p} = \kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p}). \tag{10}$$

Problem (11) is a quadratic program (QP). Numerous candidate methods exist to solve it, such as interior point, active set and projected gradient. Function $J(u)$ can thus be evaluated at any point u_0 by solving this QP problem, with $u = u_0$.

Once the problem (11) has been solved, the variable u in the problem (5) is updated using a gradient descent strategy. The existence and computation of the derivatives of supremum functions such as $J(u)$ have been largely discussed in

$$J(u) = \begin{cases} \min_{\psi} F(u,\psi) = \frac{1}{2}\left(\frac{1}{u}\|\psi_{\text{lin}}\|^2_{\mathcal{H}_{\text{lin}}} + \frac{1}{1-u}\|\psi_{\text{nlin}}\|^2_{\mathcal{H}_{\text{nlin}}}\right) + \frac{1}{2\mu}\sum_{\ell=1}^{L} e_\ell^2 \\ \qquad \text{subject to} \quad e_\ell = r_\ell - \psi(\boldsymbol{m}_{\lambda_\ell}) \quad \text{with} \quad \psi = \psi_{\text{lin}} + \psi_{\text{nlin}} \\ \qquad \text{and} \quad \psi_{\text{lin}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{h}^\top \boldsymbol{m}_{\lambda_\ell} \quad \text{with} \quad \boldsymbol{h} \succeq \boldsymbol{0} \end{cases} \tag{6}$$

$$J(u) = \begin{cases} \max_{\boldsymbol{\beta},\boldsymbol{\gamma}} G'(u,\boldsymbol{\beta},\boldsymbol{\gamma}) = -\frac{1}{2}\left(\frac{\boldsymbol{\beta}}{\boldsymbol{\gamma}}\right)^\top \left(\begin{array}{c|c} \boldsymbol{K}_u + \mu\boldsymbol{I} & u\boldsymbol{M} \\ u\boldsymbol{M}^\top & u\boldsymbol{I} \end{array}\right)\left(\frac{\boldsymbol{\beta}}{\boldsymbol{\gamma}}\right) + \left(\frac{\boldsymbol{r}}{\boldsymbol{0}}\right)^\top \left(\frac{\boldsymbol{\beta}}{\boldsymbol{\gamma}}\right) \\ \qquad \text{subject to} \quad \boldsymbol{\gamma} \succeq \boldsymbol{0} \\ \qquad \text{with} \quad \boldsymbol{K}_u = u\boldsymbol{M}\boldsymbol{M}^\top + (1-u)\boldsymbol{K}_{\text{nlin}} \end{cases} \tag{11}$$

the literature [13]. The derivatives of J at u_0 can be calculated as if the corresponding minimizer $(\boldsymbol{\beta}_0^*, \boldsymbol{\gamma}_0^*)$ was independent of u_0, i.e.,

$$\begin{aligned} \left.\frac{dJ(u)}{du}\right|_{u=u_0} &= \frac{\partial G'(u,\boldsymbol{\beta}_0^*,\boldsymbol{\gamma}_0^*)}{\partial u}\Big|_{u=u_0} \\ &= -\frac{1}{2}\left(\|\boldsymbol{M}^\top\boldsymbol{\beta}_0^* + \boldsymbol{\gamma}_0^*\|^2 - \boldsymbol{\beta}_0^{*\top}\boldsymbol{K}_{\text{nlin}}\boldsymbol{\beta}_0^*\right) \end{aligned} \tag{12}$$

An appropriate selection of the step size allows the gradient descent algorithm to ensure that the constraint $0 \leq u \leq 1$ is satisfied. After a given stopping criterion is met, such as small modulus of the derivative (12), pixel reconstruction can be performed using

$$\boldsymbol{r}^* = [\psi^*(\boldsymbol{m}_{\lambda_1}), \ldots, \psi^*(\boldsymbol{m}_{\lambda_L})]^\top \tag{13}$$

with $\psi^*(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{m}_{\lambda_\ell}^\top \boldsymbol{h}^* + \psi^*_{\text{nlin}}(\boldsymbol{m}_{\lambda_\ell})$ defined in (9). Finally, the estimated abundance vector is given by

$$\boldsymbol{\alpha}^* = \frac{\boldsymbol{M}^\top\boldsymbol{\beta}^* + \boldsymbol{\gamma}^*}{\boldsymbol{1}^\top(\boldsymbol{M}^\top\boldsymbol{\beta}^* + \boldsymbol{\gamma}^*)}. \tag{14}$$

3. EXPERIMENTS

We shall now conduct some simulations to validate the proposed unmixing algorithm, and to compare it with state-of-the-art methods, using both synthetic and real images.

3.1. Experiments on synthetic images

Let us first report some experimental results on synthetic images. The materials considered in the following scenes are epidote, kaolinite, buddingtonite, alunite, calcite, almandine, jarosite and lepidolite, whose spectra consist of 420 contiguous bands. In the first scene, only the first three materials were selected to generate images. In the second scene, the first five materials were used. In the third scene, the eight materials were considered. For each scene, three 50-by-50 images were generated with different mixture models, each providing $N = 2500$ pixels for evaluating and comparing performances. These three models were the linear model, the generalized bilinear mixture model with attenuation factors γ_{ij} equal to 1, and a post-nonlinear mixing model (PNMM) defined by $\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^{0.7} + \boldsymbol{n}$, where $(\,\cdot\,)^{0.7}$ denotes the exponential value 0.7 applied to each entry of the vector. All these images were corrupted with an additive white Gaussian noise $\boldsymbol{n}$ with SNR of 30 dB, except in the case illustrated by Figure 1 that is described hereafter. For our algorithm, the Gaussian kernel

$$\kappa_{\text{nlin}}(\boldsymbol{m}_{\lambda_p}, \boldsymbol{m}_{\lambda_\ell}) = \exp(-\|\boldsymbol{m}_{\lambda_p} - \boldsymbol{m}_{\lambda_\ell}\|^2/2\sigma^2)$$

was selected to be the reproducing kernel of $\mathcal{H}_{\text{nlin}}$.

For illustration purpose only, Figure 1 depicts the evolution of u and $1-u$ that was observed during a run of the two-stage optimization algorithm derived from (5)-(6). These preliminary simulations were conducted with a noiseless mixed pixel composed of 5 endmembers. The step sizes were adjusted in order to provide a clear demonstration. It can be observed that, in the case of the linear mixing, the weighting coefficient u converged towards 1. This means that, as expected, the ummixing algorithm only captured the linear component with ψ_{lin}. It converged to other values when confronted with the bilinear and post-nonlinear mixing models.

Our approach was compared with the fully constrained least square method (FCLS) [9], the kernel fully constrained least square method (KFCLS) [8], the extended endmember-matrix method (ExtM) [5], and the Bayesian algorithm derived for generalized bilinear model (BilBay) [4]. The root mean square error was used to compare these five algorithms. Preliminary runs were performed using independent data to tune their parameters. The results are reported in Table 1. Our method was the most efficient one in these experiments.

3.2. Experiments with AVIRIS images

Let us now illustrate the performance of the proposed algorithm, when applied to a well known image captured on the Cuprite mining district (NV, USA) by AVIRIS. A sub-image of 30×101 pixels was chosen to evaluate and compare the performance of the algorithms that have been previously enumerated. This area of interest has $L = 189$ spectral bands. VCA

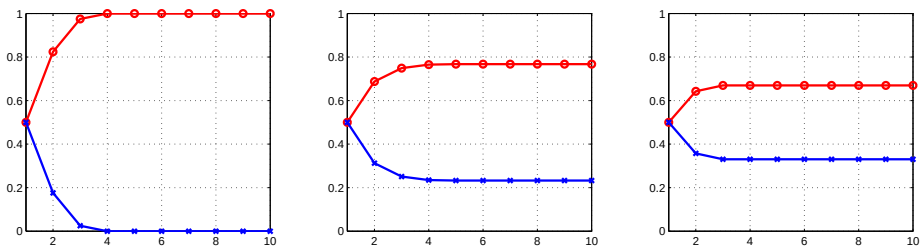

Fig. 1. Evolution of u (curves with 'o') and $1-u$ (curves with 'x') for the three models. From left to right: linear, bilinear, PNMM

Table 1. RMSE comparison

	$R=3$			$R=5$			$R=8$		
	linear	bilinear	PNMM	linear	bilinear	PNMM	linear	bilinear	PNMM
FCLS	0.0037	0.0758	0.0604	0.0134	0.1137	0.1427	0.0148	0.0930	0.1079
ExtM	0.0079	0.0312	0.0601	0.0157	0.0575	0.1427	0.0173	0.0560	0.1126
KFCLS	0.0054	0.2711	0.2371	0.0200	0.2051	0.1955	0.0216	0.1431	0.1274
BilBay	0.0384	0.0285	0.1158	0.0585	0.0441	0.1741	0.0448	0.0369	0.1159
Proposed	0.0104	0.0315	0.0230	0.0196	0.0288	0.0346	0.0185	0.0221	0.0291

algorithm was used to extract four endmembers. To evaluate the performance, the averaged spectral angle between the original $\boldsymbol{r}$ and the reconstructed $\boldsymbol{r}^*$ pixel vectors was used

$$\mathrm{SA} = \frac{1}{N}\sum_{n=1}^{N}\theta(\boldsymbol{r}_n, \boldsymbol{r}_n^*)$$

with N is the number of pixels. The performance of each approach is reported in Table 2. Note that KFCLS was not considered in these tests as there is no possible direct reconstruction of pixels. Clearly, in this example, our algorithm had lower reconstruction errors than the other approaches.

Table 2. Spectral angles comparison

	FCLS	ExtM	BilBay	Proposed
SA	0.0206	0.0126	0.0154	0.0110

4. REFERENCES

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, 2002.

[2] K. J. Guilfoyle, M. L. Althouse, and C.-I. Chang, "A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis function neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2314–2318, 2001.

[3] J. Plaza, P. Martínez, R. Pérez, and A. Plaza, "Nonlinear neural network mixture models for fractional abundance estimation in AVIRIS hyperspectral images," in *Proc. AVIRIS workshop*, Pasadena, CA, 2004.

[4] A. Halimi, Y. Altman, 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.

[5] N. Raksuntorn and Q. Du, "Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 836–840, 2010.

[6] J. M. P. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," in *Proc. SPIE*, 2009, vol. 7477.

[7] J. M. P. Nascimento and J. M. Bioucas-Dias, "Unmixing hyperspectral intimate mixtures," in *Proc. SPIE*, 2010, vol. 7830.

[8] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE IGARSS*, 2007, pp. 4041–4044.

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

[10] J. Chen, C. Richard, and P. Honeine, "A novel kernel-based nonlinear unmixing scheme of hyperspectral images," in *Proc. 45th Asilomar Conference on Signals, Systems, and Computers (Asilomar)*, Pacific Grove (CA), USA, November 2011.

[11] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Granvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, University Press, Cambridge, 2004.

[13] J. F. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM review*, vol. 40, no. 2, pp. 207–227, 1998.