PIXEL-WISE LINEAR/NONLINEAR NONNEGATIVE MATRIX FACTORIZATION FOR UNMIXING OF HYPERSPECTRAL DATA

Fei Zhu[(1)]*, Paul Honeine*[(2)] *and Jie Chen* [(3)]

[(1)]Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China
[(2)]LITIS, Normandie Univ, UNIROUEN, 76000 Rouen, France
[(3)]School of Marine Science and Technology, Northwestern Polytechnical University, 710072 Xi'an, China
fei.zhu@tju.edu.cn, paul.honeine@univ-rouen.fr, dr.jie.chen@ieee.org

ABSTRACT

Nonlinear spectral unmixing is a challenging and important task in hyperspectral image analysis. The kernel-based bi-objective nonnegative matrix factorization (Bi-NMF) has shown its usefulness in nonlinear unmixing; However, it suffers several issues that prohibit its practical application. In this work, we propose an unsupervised nonlinear unmixing method that overcomes these weaknesses. Specifically, the new method introduces into each pixel a parameter that adjusts the nonlinearity therein. These parameters are jointly optimized with endmembers and abundances, using a carefully designed objective function by multiplicative update rules. Experiments on synthetic and real datasets confirm the effectiveness of the proposed method.

Index Terms— Hyperspectral data analysis, nonlinear unmixing, unsupervised learning, kernel methods, nonnegative matrix factorization.

1. INTRODUCTION

Spectral unmixing is an important issue in hyperspectral data analysis, and it covers a wide range of applications such as land cover analysis, material sorting, water quality evaluation and mineral distribution analysis. In the unmixing problem, each observed pixel is a mixture of several spectral signatures of pure materials (endmembers) parameterized by their respective fractional abundances. We aim to extract the endmembers and estimate the abundances based on specific pre-assumed data models. The linear mixing model (LMM) is the most prevalent model to characterize the relation between the endmembers and an observed pixel. A variety of endmember extraction and abundance estimation methods have been studied with the LMM as reviewed in [1].

However, the LMM may not be appropriate in some practical situations where the light is scattered by multiple reflective or interacted materials. Nonlinear mixture models (NLMM) then provide an alternative to overcome the limitations of the LMM [2]. Physics-oriented nonlinear methods include the Hapke's model [3], bilinear-mixing models [4, 5], polynomial post-nonlinear mixing model (PPNMM) [6] and multilinear mixing model (MLM) [7]. Besides these parametric nonlinear models, kernel-based non-parametric methods have recently drawn increasing attention. In [8], the mixture model is defined by augmenting LMM with an addictive nonlinear fluctuation term defined in RKHS, and this model is extended in several means [9, 10, 11].

Based on the same principle, a bi-objective nonnegative matrix factorization (Bi-NMF) for nonlinear unmixing is proposed in [12]. Taking advantage of the NMF framework, the Bi-NMF is an unsupervised algorithm able to simultaneously estimate endmembers and abundances by jointly taking into account the nonlinear effects. Though useful, it suffers from the following drawbacks: i) The nonlinearity of the entire image is characterized in a global manner, ii) forming the Pareto front used in the algorithm requires to perform the proposed factorization a number of times, and iii) selecting the best factorization among Pareto solutions is difficult.

In this work, we present an unsupervised linear/nonlinear unmixing method that successfully bypasses the above drawbacks of the Bi-NMF. Specifically, each observed spectrum is modeled jointly by the linear and the kernel-based models, along with a pixel-wise variable adjusting their relative importance. The optimization problem is addressed by a block-coordinate descent (BCD) approach. Multiplicative algorithms are provided to update iteratively the endmembers, abundances, and nonlinearity variables. Experimental results on synthetic datasets with both linearly-mixed and nonlinearly-mixed pixels, and on real datasets show the effectiveness of the proposed method.

2. PROBLEM FORMULATION OF THE BI-NMF

This section briefly presents the LMM and the kernel-based mixture model in [13], both employed in the Bi-NMF method. The proposed unmixing approach will be developed based on these models.

F. Zhu was supported by the National Natural Science Foundation of China under Grant 61701337, the Natural Science Foundation of Tianjin City under Grand 18JCQNJC01600. P. Honeine was supported by the French National Research Agency through the project APi (ANR-18-CE23-0014).

Let $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T] \in \mathbb{R}_+^{L\times T}$ be the collection of T observed pixels over L spectral bands. Let $\boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_R] \in \mathbb{R}^{L\times R}$ be the endmember matrix composed by R endmembers $\{\boldsymbol{e}_i\}_{i=1}^R$, and $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_T] \in \mathbb{R}^{R\times T}$ the abundance matrix, where $\boldsymbol{a}_t \in \mathbb{R}^R$ is the abundance vector associated to the t-th pixel, with the scalar a_{rt} being the abundance fraction of the r-th endmember.

The linear mixing model [1] assumes that each observed pixel is a linear combination of several endmembers, weighted by their fractional abundances, namely

$$\boldsymbol{x}_t \approx \sum_{r=1}^{R} a_{rt}\, \boldsymbol{e}_r, \tag{1}$$

for all $t = 1, ..., T$. Let input space $\mathcal{X}$ be the subspace spanned by all the pixels $\boldsymbol{x}_t$ and endmembers $\boldsymbol{e}_r$. The linear unmixing problem aims to minimize the sum of squared residual errors between each input spectrum and its estimate in the input space $\mathcal{X}$, that is

$$J_{\mathcal{X}}(\boldsymbol{E}, \boldsymbol{A}) = \frac{1}{2}\sum_{t=1}^{T} \Big\| \boldsymbol{x}_t - \sum_{r=1}^{R} a_{rt}\, \boldsymbol{e}_r \Big\|^2, \tag{2}$$

where the nonnegativity constraints $\boldsymbol{E}, \boldsymbol{A} \succeq 0$ are imposed.

Let $\Phi(\cdot)$ be a nonlinear function that maps all pixels $\boldsymbol{x}_t$ and endmembers $\boldsymbol{e}_r$, from the input space $\mathcal{X}$ to some feature space $\mathcal{H}$, with the associated norm denoted by $\| \cdot \|_{\mathcal{H}}$. The inner product in $\mathcal{H}$ computes $\langle \Phi(\boldsymbol{x}_t), \Phi(\boldsymbol{x}_{t'}) \rangle_{\mathcal{H}} = \kappa(\boldsymbol{x}_t, \boldsymbol{x}_{t'})$, where $\kappa(\cdot, \cdot)$ is a kernel function. The kernel-based mixing model [13] defines the mixture process in the RKHS feature space as

$$\Phi(\boldsymbol{x}_t) \approx \sum_{r=1}^{R} a_{rt}\, \Phi(\boldsymbol{e}_r), \tag{3}$$

which allows to overcome the curse of the pre-image problem. Under the nonnegativity constraints on $\boldsymbol{E}$ and $\boldsymbol{A}$, the optimization problem is formulated to minimize the sum of the squared residual errors in the feature space $\mathcal{H}$, i.e.,

$$\mathcal{J}_{\mathcal{H}}(\boldsymbol{E}, \boldsymbol{A}) = \frac{1}{2}\sum_{t=1}^{T} \Big\| \Phi(\boldsymbol{x}_t) - \sum_{r=1}^{R} a_{rt}\, \Phi(\boldsymbol{e}_r) \Big\|_{\mathcal{H}}^2. \tag{4}$$

In the Bi-NMF method [12], the nonlinear unmixing is formulated as a bi-objective optimization problem. Using the sum-weighted method, the original problem breaks down to a series of single-objective subproblems in the form of

$$\min_{\boldsymbol{E}, \boldsymbol{A} \succeq 0} \ \alpha J_{\mathcal{X}}(\boldsymbol{E}, \boldsymbol{A}) + (1-\alpha) J_{\mathcal{H}}(\boldsymbol{E}, \boldsymbol{A}). \tag{5}$$

Employing a spread of α between 0 and 1, the linear and nonlinear models are fused at different levels. After solving each subproblem, a set of Pareto optimal solutions are obtained as candidates of the best unmixing result. Note that the major drawback of Bi-NMF is that it provides a global result of the nonlinearity detection, ignoring the fact that pixels in an image may have different nonlinearity levels.

3. PROPOSED UNSUPERVISED LINEAR/NONLINEAR UNMIXING BY KERNEL METHODS

Instead of characterizing the nonlinearity of all spectra by a single parameter, it is more reasonable to individually describe the nonlinearity level of each pixel. To this end, we define the objective function associated to each pixel $\boldsymbol{x}_t$ by:

$$\mathcal{J}_t(\boldsymbol{E}, \boldsymbol{a}_t, \mu_t) = \frac{\|\boldsymbol{x}_t - \boldsymbol{E}\boldsymbol{a}_t\|^2}{2\mu_t} + \frac{\|\Phi(\boldsymbol{x}_t) - \sum_{r=1}^{R} a_t\, \Phi(\boldsymbol{e}_r)\|_{\mathcal{H}}^2}{2(1-\mu_t)}, \tag{6}$$

where the unknown scalar $\mu_t \in (0, 1)$ balances the relative importance between the linear and kernel-based models at the t-th spectrum. Intuitively, a pixel tends to be linearly mixed when its nonlinearity level is close to 0, while a value closed to 1 indicates that the pixel is highly nonlinearly mixed. See [8] for a more thorough interpretation of the usefulness of the problem formulation in this form for hyperspectral unmixing.

Summing up the objective functions over all pixels and imposing the nonnegativity constraints on both endmembers and abundances, the optimization problem becomes

$$\begin{aligned} \min_{\boldsymbol{E}, \boldsymbol{A}, \boldsymbol{\mu}} \ & \sum_{t=1}^{T} \mathcal{J}_t(\boldsymbol{E}, \boldsymbol{a}_t, \mu_t) \\ \text{subject to } \ & \boldsymbol{E}, \boldsymbol{A} \succeq 0 \text{ and } \mu_t \in (0, 1), \end{aligned} \tag{7}$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2, \cdots, \mu_T]^\top$ is an unknown vector collecting nonlinearity parameters of all pixels, with each element satisfying the constraint $\mu_t \in (0, 1)$.

4. BLOCK-COORDINATE DESCENT ALGORITHM

To solve the problem (7), we apply a block-coordinate descent (BCD) optimization approach, by alternately updating each of the variables blocks $\boldsymbol{E}$, $\boldsymbol{A}$ and $\boldsymbol{\mu}$, while keeping the elements in the other two blocks fixed.

Recall that the inner product in $\mathcal{H}$ is given by $\langle \Phi(\boldsymbol{x}_t), \Phi(\boldsymbol{x}_{t'}) \rangle_{\mathcal{H}} = \kappa(\boldsymbol{x}_t, \boldsymbol{x}_{t'})$, expanding the objective function in (7) and removing all terms irrelevant to $\boldsymbol{e}_r$ and a_{rt}, yields:

$$\begin{aligned} \mathcal{J} = \sum_{t=1}^{T} \Bigg[& \frac{1}{2\mu_t} \Big(\sum_{r,s=1}^{R} a_{rt} a_{st} \boldsymbol{e}_r^\top \boldsymbol{e}_s - 2 \sum_{r=1}^{R} a_{rt} \boldsymbol{e}_r^\top \boldsymbol{x}_t \Big) \\ & + \frac{1}{2(1-\mu_t)} \Big(\sum_{r,s=1}^{R} a_{rt} a_{st} \kappa(\boldsymbol{e}_r, \boldsymbol{e}_s) - 2 \sum_{r=1}^{R} a_{rt} \kappa(\boldsymbol{e}_r, \boldsymbol{x}_t) \Big) \Bigg]. \end{aligned} \tag{8}$$

We now solve (8) with respect to the variables $\boldsymbol{E}$, $\boldsymbol{A}$.

4.1. Optimization with respect to $\boldsymbol{E}$

We first fix $\boldsymbol{A}$ and $\boldsymbol{\mu}$ to minimize the objective function (8) in terms of endmember matrix $\boldsymbol{E}$. The gradient of (8) with respect to $\boldsymbol{e}_r$ is given by

$$\begin{aligned} \nabla_{\boldsymbol{e}_r} \mathcal{J} = \sum_{t=1}^{T} a_{nt} \Big[& \frac{1}{\mu_t} (\boldsymbol{E}\boldsymbol{a}_t - \boldsymbol{x}_t) \\ & + \frac{1}{1-\mu_t} \Big(\sum_{m=1}^{R} a_{mt} \nabla_{\boldsymbol{e}_r} \kappa(\boldsymbol{e}_r, \boldsymbol{e}_m) - \nabla_{\boldsymbol{e}_r} \kappa(\boldsymbol{e}_r, \boldsymbol{x}_t) \Big) \Big]. \end{aligned} \tag{9}$$

In this work, we consider to use the Gaussian RBF kernel defined by $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\frac{1}{2\sigma^2} \|\boldsymbol{x} - \boldsymbol{y}\|^2)$ due to its good performance in a wide-range of applications. Consequently, we have

$$\nabla_{\boldsymbol{e}_r} \kappa(\boldsymbol{e}_r, \boldsymbol{z}) = -\frac{1}{\sigma^2} \kappa(\boldsymbol{e}_r, \boldsymbol{z})(\boldsymbol{e}_r - \boldsymbol{z}). \tag{10}$$

Using the split gradient method [14] to appropriately set the stepsize parameter η in the gradient descent updating rule $\boldsymbol{e}_r = \boldsymbol{e}_r - \eta \nabla_{\boldsymbol{e}_r} \mathcal{J}$, a multiplicative update rule of endmember matrix $\boldsymbol{E}$ is achieved as

$$\boldsymbol{e}_r = \boldsymbol{e}_r \otimes \frac{\sum_{t=1}^{T} \frac{\sigma^2 a_{rt}}{\mu_t} \boldsymbol{x}_t + \frac{a_{rt}}{1-\mu_t} \Big(\kappa(\boldsymbol{e}_r, \boldsymbol{x}_t) \boldsymbol{x}_t + \sum_{m=1}^{R} a_{mt} \kappa(\boldsymbol{e}_r, \boldsymbol{e}_m) \boldsymbol{e}_r \Big)}{\sum_{t=1}^{T} \frac{\sigma^2 a_{rt}}{\mu_t} \boldsymbol{E}\boldsymbol{a}_t + \frac{a_{rt}}{1-\mu_t} \Big(\kappa(\boldsymbol{e}_r, \boldsymbol{x}_t) \boldsymbol{e}_r + \sum_{m=1}^{R} a_{mt} \kappa(\boldsymbol{e}_r, \boldsymbol{e}_m) \boldsymbol{e}_m \Big)}, \tag{11}$$

where $\otimes$ represents the Hadamard multiplication operator.

4.2. Optimization with respect to $\boldsymbol{A}$

We then fix $\boldsymbol{E}$ and $\boldsymbol{\mu}$, to derive the update rule for the abundance matrix $\boldsymbol{A}$. The derivative of objective function (8) with respect to a_{rt} is given by:

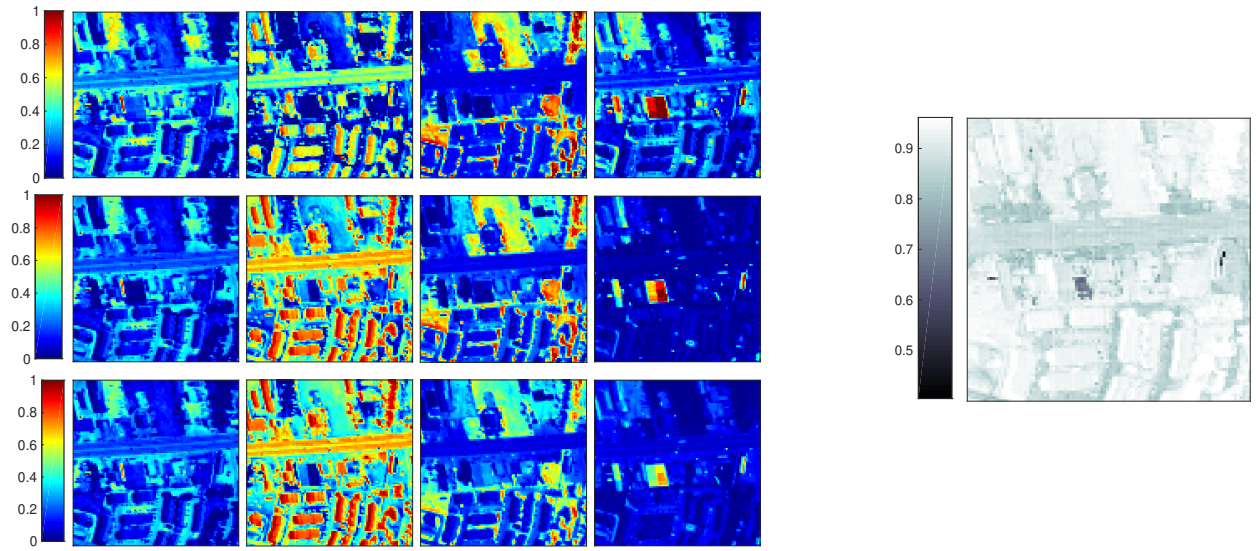
$$\begin{aligned} \nabla_{a_{rt}} \mathcal{J} = & \frac{1}{\mu_t} \Big[\sum_{m=1}^{R} a_{mt}\, \boldsymbol{e}_r^\top \boldsymbol{e}_m - \boldsymbol{e}_r^\top \boldsymbol{x}_t \Big] \\ & + \frac{1}{1-\mu_t} \Big[\sum_{m=1}^{R} a_{mt}\, \kappa(\boldsymbol{e}_r, \boldsymbol{e}_m) - \kappa(\boldsymbol{e}_r, \boldsymbol{x}_t) \Big]. \end{aligned} \tag{12}$$

Again, applying the split gradient method leads to the multiplicative update rule for $\boldsymbol{A}$ as follows:

$$a_{rt} = a_{rt} \times \frac{(1-\mu_t)\, \boldsymbol{e}_r^\top \boldsymbol{x}_t + \mu_t\, \kappa(\boldsymbol{e}_r, \boldsymbol{x}_t)}{(1-\mu_t) \sum_{m=1}^{R} a_{mt} \boldsymbol{e}_r^\top \boldsymbol{e}_m + \mu_t \sum_{m=1}^{R} a_{mt}\, \kappa(\boldsymbol{e}_r, \boldsymbol{e}_m)}. \tag{13}$$

Algorithm 1 The BCD algorithm for solving problem (7)

Input: $k = 0$, Initialize $\boldsymbol{E}^0$ and $\boldsymbol{A}^0$
1: **repeat**
2: update $\boldsymbol{E}^{k+1}$ with (11)
3: update $\boldsymbol{A}^{k+1}$ with (13)
4: update $\boldsymbol{\mu}^{k+1}$ with (15)
5: $k = k + 1$
6: **until** stopping criterion
Output: $\boldsymbol{E}$, $\boldsymbol{A}$ and $\boldsymbol{\mu}$



(a) (b)

Fig. 1: Experimental results on Urban image: (a) Abundance maps for Grass, Asphalt, Tree, Roof obtained by MinDisCo, MVCNMF and the proposed method. (b)μ-map obtained by the proposed method.

4.3. Optimization with respect to $\boldsymbol{\mu}$

Regarding the optimization with respect to $\boldsymbol{\mu}$, it has a closed-form expression for the optimum solution $\boldsymbol{\mu}^*$, thanks to the elegant problem formulation in (6) [1]. It is observed that the variable μ_t only appears in $\mathcal{J}_t$, thus, the gradient of (7) with respect to $\boldsymbol{\mu}$ can be expressed as

$$\nabla_{\boldsymbol{\mu}} \mathcal{J} = \left[\frac{\partial \mathcal{J}_1}{\partial \mu_1}, \frac{\partial \mathcal{J}_2}{\partial \mu_2}, \cdots, \frac{\partial \mathcal{J}_T}{\partial \mu_T}\right]^\top. \tag{14}$$

Setting $\nabla_{\boldsymbol{\mu}} \mathcal{J} = 0$ yields the closed-form expression of the optimum $\boldsymbol{\mu}^* = [\mu_1^*, \mu_2^*, ..., \mu_T^*]^\top$, where the t-th element μ_t takes the form

$$\mu_t^* = \left(1 + \sqrt{\frac{\|\Phi(\boldsymbol{x}_t) - \sum_{r=1}^R a_{rt}\, \Phi(\boldsymbol{e}_r)\|_{\mathcal{H}}^2}{\|\boldsymbol{x}_t - \boldsymbol{E}\boldsymbol{a}_t\|^2}}\right)^{-1}. \tag{15}$$

Here, the residual errors in feature space $\mathcal{H}$ can be easily computed via the kernel function by

$$\begin{aligned} &\|\Phi(\boldsymbol{x}_t) - \sum_{r=1}^R a_{rt}\Phi(\boldsymbol{e}_r)\|_{\mathcal{H}}^2 \\ &= \sum_{r,s=1}^R a_{st}a_{rt}\kappa(\boldsymbol{e}_r, \boldsymbol{e}_s) - 2\sum_{r=1}^R a_{rt}\kappa(\boldsymbol{x}_t, \boldsymbol{e}_r) + 1. \end{aligned} \tag{16}$$

Finally, we summarize the proposed algorithm in Algorithm 1. Its total complexity is $\mathcal{O}(kR^2LT)$ after k iterations.

[1] Consider the optimization of $f(x) = \frac{a}{x} + \frac{b}{1-x}$ over the interval $x \in (0, 1)$, where $a, b \geq 0$ are constants. This function is convex on $(0, 1)$, and its optimum is achieved at $x^* = (1 + \sqrt{b/a})^{-1}$ [15].

Table 1: RMSE of abundances on synthetic data ($\times 10^{-2}$)

Mix. / Unmix.	80% LMM & 20% GBM			50% LMM & 50% GBM			20%LMM & 80 %GBM		
	25dB	30dB	35dB	25dB	30dB	35dB	25dB	30dB	35dB
FCLS	6.55	4.74	4.30	5.86	4.38	5.91	6.96	7.23	7.54
GBM	6.15	4.69	4.24	5.47	3.81	4.83	6.68	7.08	7.38
PPNMM	5.85	4.81	4.28	5.52	3.53	4.49	6.47	7.06	7.22
MinDisCo	8.27	9.31	8.20	8.19	8.17	9.46	10.40	11.00	10.13
MVCNMF	**4.44**	4.02	**3.02**	4.70	3.57	**4.06**	11.76	12.15	11.81
Proposed	5.65	**3.98**	3.54	**4.58**	**3.42**	4.48	**5.38**	**5.84**	**5.78**

Table 2: $\mathrm{SAD}^{\mathrm{E}}$ of endmembers on synthetic data ($\times 10^{-2}$)

Mix. / Unmix.	80% LMM & 20% GBM			50% LMM & 50% GBM			20% LMM & 80% GBM		
	25dB	30dB	35dB	25dB	30dB	35dB	25dB	30dB	35dB
VCA	3.31	3.14	2.89	4.49	3.76	4.04	4.79	5.22	4.67
MinDisCo	6.53	6.10	5.97	7.66	7.26	6.09	7.69	7.81	7.45
MVCNMF	3.35	3.08	2.64	4.00	3.74	3.85	28.19	28.99	28.70
Proposed	**2.98**	**2.66**	**2.36**	**3.59**	**3.15**	**3.13**	**3.91**	**4.49**	**3.77**

5. EXPERIMENTAL RESULTS

In this section, experiments were conducted with both synthetic and real datasets to validate the proposed algorithm. Five state-of-the-art unmixing algorithms were compared, including three supervised methods, namely, FCLS [16], GBM [17], and PPNMM [18], and two unsupervised ones, namely, MinDisCo [19] and MVCNMF [20]. For the supervised unmixing approaches, VCA [21] was applied to extract endmembers.

5.1. Experiments on synthetic data

The performance of the proposed method was firstly studied on a series of synthetic images of size 20×20 pixels. To ensure the existence of both linear and nonlinear mixed pixels, two mixing models, namely LMM and GBM were used for data generation. The GBM is defined by

$$\boldsymbol{x}_t = \sum_{n=1}^{N} a_{nt}\, \boldsymbol{e}_n + \sum_{n=1}^{N-1} \sum_{m=n+1}^{N} \gamma_{nm}\, a_{nt}\, a_{mt}\, (\boldsymbol{e}_n \otimes \boldsymbol{e}_m) + \boldsymbol{n}, \tag{17}$$

where $\gamma_{nm} \in [0, 1]$ was drawn from a uniform distribution and $\boldsymbol{n} \in \mathbb{R}^{L\times 1}$ is the additive noise [17]. The $N = 3$ endmembers were randomly drawn from the candidate spectra set consisting of 19 spectra from the United States Geological Survey (USGS) digital spectral library [22]. The abundance vectors were generated using a uniform distribution by enforcing both the nonnegativity and the sum-to-one constraints [22]. Each image was composed by pixels generated by the LMM and GBM with three proportion settings, namely, $(80\%, 20\%)$, $(50\%, 50\%)$ and $(20\%, 80\%)$. Gaussian noise with the signal-to-noise ratio (SNR) of 25 dB, 30 dB and 35 dB were added to the generated data.

The bandwidth parameter σ in the Gaussian kernel was tuned using the candidate values $\{1, 2, 5, 10, 15, 20, 25, 30, 35\}$ on one synthetic image, and then was set to $\sigma = 25$ in all experiments. The maximum iteration number was set to 1500 in all tested methods. Unmixing performance was evaluated by both the root mean square error (RMSE) [17]

$$\mathrm{RMSE} = \sqrt{\frac{1}{NT} \sum_{t=1}^{T} \|\boldsymbol{a}_t - \overline{\boldsymbol{a}}_t\|^2} \tag{18}$$

for abundance estimation, and the averaged spectral angle distance ($\mathrm{SAD}^{\mathrm{E}}$) [19]

$$\mathrm{SAD}^{\mathrm{E}} = \frac{1}{N} \sum_{n=1}^{N} \arccos\left(\frac{\langle \boldsymbol{e}_n, \overline{\boldsymbol{e}_n} \rangle}{\|\boldsymbol{e}_n\| \|\overline{\boldsymbol{e}}_n\|} \right). \tag{19}$$

for endmember estimation. Here, $\overline{\boldsymbol{e}}_n$ and $\overline{\boldsymbol{a}}_t$ represent the actual endmember and abundance vector.

For each setting, ten Monte-Carlo simulations were conducted. The average RMSE and $\mathrm{SAD}^{\mathrm{E}}$ are reported in Table 1 and Table 2, respectively. For abundance estimation, the proposed method provides the best estimation in most cases with varying proportions of nonlinearly-mixed pixels at different noise levels. For the cases with 80% LMM and 20% GBM pixels at SNR $= 25,\ 35$ dB, and with 50% LMM and 50% GBM pixels at SNR $= 35$ dB, the proposed method yields second best RMSE. For endmember estimation, the proposed method is superior to the other algorithms in all experiments.

Table 3: RE and SAD on Urban image ($\times 10^{-2}$)

	FCLS	GBM	PPNMM	MinDisCo	MVCNMF	Proposed
RE	8.64	8.64	1.37	1.13	0.79	**0.78**
SAD	31.44	31.46	13.79	7.71	6.19	**5.98**

5.2. Experiments on real data

The performance of the proposed method was also examined on a sub-image from the Urban data, captured by HYDICE sensor. The raw image contains 307×307 pixels, with 210 spectral bands ranging from $0.4\mu m$ to $2.5\mu m$. The field has been known mainly composed of four endmembers, namely, asphalt, grass, tree and roof [12, 23]. The top left corner with 100×100 pixels were utilized, and 162 bands with high SNR were retained for analysis [12, 23].

Without information on actual abundance vectors and endmembers, the unmixing performance is evaluated using the reconstruction error (RE) and the average spectral angle distance (SAD) between the observed and reconstructed pixels, as reported in Table 3. We observe that the proposed method leads to the smallest RE and SAD among all the comparing methods. The abundance maps estimated by different methods are illustrated in Figure 1(a). The μ-map obtained by the proposed method is shown in Figure 1(b), from which the distribution of nonlinearity levels of pixels is clearly illustrated.

6. CONCLUSION

This paper presented an unsupervised unmixing strategy based on the linear/nonlinear mixture assumption, by taking advantage of the NMF framework. The nonlinearity level was estimated at each pixel, which allows to achieve a more accurate modeling capacity than previously proposed methods, such as [12]. The formulated optimization problem was solved by the BCD strategy. Future works include incorporating proper spatial regularizations on these newly introduced variables.

7. REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, April 2012.

[2] N. Dobigeon, J.-Y. Tourneret, C. Richard, J.C.M. Bermudez, S. McLaughlin, and A.O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 82–94, Jan 2014.

[3] B. Hapke, "Bidirectional reflectance spectroscopy: 1. theory," *Journal of Geophysical Research: Solid Earth*, vol. 86, no. B4, pp. 3039–3054, 1981.

[4] A. Halimi, Y. Altmann, N. Dobigeon, and J. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," in *2011 IEEE Statistical Signal Processing Workshop (SSP)*, June 2011, pp. 413–416.

[5] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1430–1437, Feb 2014.

[6] Y. Altmann, A. Halimi, N. Dobigeon, and J. Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 3017–3025, June 2012.

[7] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 240–251, Jan 2016.

[8] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 480–492, Jan. 2013.

[9] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances of hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2654–2665, May 2014.

[10] J. Chen and C. Richard, "Robust nonlinear unmixing of hyperspectral images with a linear-mixture/nonlinear-fluctuation model," in *Geoscience and Remote Sensing Symposium*, 2016.

[11] Z. Li, J. Chen, and S. Rahardja, "A graph regularized multilinear mixing model for nonlinear hyperspectral unmixing," *Remote Sens.*, vol. 11, no. 529, pp. 1–17, 2019.

[12] F. Zhu and P. Honeine, "Bi-objective nonnegative matrix factorization: Linear versus kernel-based models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4012–4022, 2016.

[13] F. Zhu, P. Honeine, and M. Kallas, "Kernel non-negative matrix factorization without the pre-image problem," in *IEEE workshop on Machine Learning for Signal Processing*, Reims, France, Sep. 2014.

[14] H. Lanteri, C. Theys, C. Richard, and D. Mary, "Regularized split gradient method for nonnegative matrix factorization," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1133–1136.

[15] J. Chen, C. Richard, A. Ferrari, and P. Honeine, "Nonlinear unmixing of hyperspectral data with partially linear least-squares support vector regression," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 2174–2178.

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

[17] 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, Nov. 2011.

[18] Y. Altmann, A. Halimi, N. Dobigeon, and J-Y Tourneret, "Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery," *Image Processing, IEEE Transactions on*, vol. 21, no. 6, pp. 3017–3025, 2012.

[19] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.

[20] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[21] J.M.P. Nascimento and J.M. Bioucas Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[22] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, Aug 2008.

[23] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3867–3879, Dec. 2007.