ROBUST NONLINEAR UNMIXING OF HYPERSPECTRAL IMAGES WITH A LINEAR-MIXTURE/NONLINEAR-FLUCTUATION MODEL

Jie Chen[⋆]   *Cédric Richard*[†]

[⋆] CIAIC, School of Marine Science and Technology, Northwestern Polytechincal University, China
[†] Laboratoire Lagrange, Université de Nice Sophia-Antipolis, CNRS, France
E-mail: dr.jie.chen@ieee.org, cedric.richard@unice.fr

ABSTRACT

Hyperspectral data unmixing has attracted considerable attention in recent years. Hyperspectral data may however suffer from varying levels of signal-to-noise ratio over spectral bands. In this paper, we investigate a robust approach for nonlinear hyperspectral data unmixing. Each observed pixel is modeled as a linear mixing of endmember spectra with nonlinear fluctuations embedded in a reproducing kernel Hilbert space. Welsch M-estimator is considered for reducing the sensitivity of the unmixing process. Experimental results, with both synthetic and real data, illustrate the effectiveness of the proposed scheme.

Index Terms— Hyperspectral data analysis, nonlinear unmixing, robust unmixing, M-estimator

1. INTRODUCTION

Hyperspectral imagery has many applications and has received considerable attention from the community [1]. Pixel-vectors in hyperspectral images are usually mixtures of material signatures. Spectral unmixing aims at identifying endmembers and evaluating the fractions of abundance of the corresponding materials [2]. The linear mixing model is still widely used, with a variety of techniques proposed to both estimate endmembers and their abundances [3,4]. Nevertheless, it is recognized that nonlinear mixing effects can be prominent in real-world scenarios [1,2]. A variety of nonlinear mixing models and unmixing methods have been proposed in the literature. See [5] for an overview. In [6], and consecutively in [7–9], the authors propose a kernel-based unmixing framework where the mixing model consists of a linear mixing term, and additive nonlinear fluctuations depending on the endmembers and defined in a reproducing kernel Hilbert space (RKHS). This model is characterized by its ability to describe complex nonlinearities with moderate computational cost. Other works build on this principle, known as Residual Component Analysis [10,11].

The seminal work [6] and related publications use a quadratic error criterion for model fitting. This criterion implicitly assumes that the noise is i.i.d. Gaussian across spectral bands. It is however admitted that the noise level varies across bands, and some bands can be severely distorted. In this paper, we address this problem by considering a robust regression method. These techniques are less affected by outliers than the ordinary least squares method. M-estimation is a common framework for robust regression, among others such as L-estimation and R-estimation [12,13]. M-estimation methods cope with outlying observations by replacing the squared residuals in the mean square error criterion by another function of the residuals chosen to be less increasing than the square error function. We build on this idea to extend the framework [6] and address the problem of robust nonlinear unmixing of hyperspectral data. A convex problem is formulated based on Welsch M-estimation. Half-quadratic minimization is used to split the problem into two subproblems that can be addressed efficiently with existing tools.

Consider an hyperspectral image with L contiguous spectral bands. Suppose that the scene consists of R significant endmembers, each with a spectral signature $\boldsymbol{m}_i \in \mathbb{R}^L$. Let $\boldsymbol{r} = [r_1, \ldots, r_L] \in \mathbb{R}^L$ be an observation, and let $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_R]^\top \in \mathbb{R}^R$ be the vector of endmember abundances in $\boldsymbol{r}$. Let $\boldsymbol{M} = [\boldsymbol{m}_1, \ldots, \boldsymbol{m}_R] \in \mathbb{R}^{L \times R}$ be the matrix of endmember spectra. For the sake of convenience, the ℓ-th row of $\boldsymbol{M}$ is denoted by $\boldsymbol{m}_{\lambda_\ell}^\top \in \mathbb{R}^L$. Let $\mathbf{1}$ and $\mathbf{I}$ be the all-one vector and the identity matrix. Expression $\mathrm{diag}\{x_1, \ldots, x_L\}$ denotes a diagonal matrix with $x_1, \ldots, x_L$ its diagonal entries.

2. NONLINEAR UNMIXING IN RKHS WITH WELSCH M-ESTIMATOR

2.1. Nonlinear unmixing in RKHS

Consider the general unmixing process, acting between the entries r_ℓ of the observed reflectance vector, and the spectral signatures $\boldsymbol{m}_{\lambda_\ell}$ of the endmembers at each wavelength band λ_ℓ, defined as

$$r_\ell = \psi_{\boldsymbol{\alpha}}(\boldsymbol{m}_{\lambda_\ell}) + e_\ell$$

with $\psi_{\boldsymbol{\alpha}}$ the nonlinear function to identify between r_ℓ and $\boldsymbol{m}_{\lambda_\ell}$ for all ℓ, parameterized by $\boldsymbol{\alpha}$, and e_ℓ the estimation error at band λ_ℓ. Consider the general problem:

$$\psi_{\boldsymbol{\alpha}}^o = \arg\min_{\psi_{\boldsymbol{\alpha}}} \frac{1}{2}\|\psi_{\boldsymbol{\alpha}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu}\sum_{\ell=1}^{L}(r_{n,\ell} - \psi_{\boldsymbol{\alpha}}(\boldsymbol{m}_{\lambda_\ell}))^2 \quad (1)$$

where μ a positive parameter that controls the trade-off between structural error and misadjustment error. In [6], we define $\psi_{\boldsymbol{\alpha}}$ by a linear trend parameterized by the abundance vector $\boldsymbol{\alpha}$, combined with a nonlinear fluctuation function ψ, namely,

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{m}_{\lambda_\ell}) = \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} + \psi(\boldsymbol{m}_{\lambda_\ell}) \quad (2)$$

where ψ can be any real-valued function in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, endowed with the reproducing kernel κ such that $\psi(\boldsymbol{m}_{\lambda_\ell}) = \langle \psi, \kappa(\cdot, \boldsymbol{m}_{\lambda_\ell}) \rangle$. This semi-parametric model combines a linear mixing model and an additive nonlinear perturbation term. Considering RKHS leads to computationally efficient and accurate resolution methods for inverse problems of the form (1) by exploiting the idea, known as the *kernel trick*.

We propose K-Hype in [6] to perform unmixing by solving the following least-square regression problem regularized by the

ℓ_2-norm factor $\|\psi_{\boldsymbol{\alpha}}\|^2 = \|\boldsymbol{\alpha}\|^2 + \|\psi\|_{\mathcal{H}}^2$, namely,

$$\boldsymbol{\alpha}^o, \psi^o = \underset{\boldsymbol{\alpha}, \psi \in \mathcal{H}}{\arg\min} \frac{1}{2}\left(\|\boldsymbol{\alpha}\|^2 + \|\psi\|_{\mathcal{H}}^2 + \frac{1}{\mu}\|\boldsymbol{e}\|^2\right) \tag{3}$$

$$\text{subject to} \quad \boldsymbol{\alpha}_n \succeq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{\alpha} = 1,$$

with $\boldsymbol{e}$ the $L \times 1$ error vector with ℓ-th entry given by

$$e_\ell = r_\ell - (\boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} + \psi(\boldsymbol{m}_{\lambda_\ell})) \tag{4}$$

2.2. Robust regression

Robust regression provides an alternative to least squares regression when the assumption of Gaussian i.i.d. error terms is not fulfilled, particularly for heavy-tailed distributed errors. M-estimation introduced in [12] provides a common conceptual framework. It consists of minimizing the loss function:

$$\min_{\boldsymbol{\theta}} \sum_{\ell=1}^{L} \rho(e_\ell; \boldsymbol{\theta}) \tag{5}$$

where ρ is a function defining the contribution of each residual e_ℓ, and $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \psi\}$ is the set of unknown variables in (4). The function ρ should reasonably be differentiable, positive, even, non-decreasing on $\mathbb{R}_+$ and satisfies $\rho(0) = 0$. In most cases, no closed form solution exists for (5) and iterative optimization techniques are required. M-estimation can often be performed by solving iteratively the reweighted least-square problem:

$$\min_{\boldsymbol{\theta}} \sum_{\ell=1}^{L} w(e_\ell^{(k)}) \, e_\ell^2 \tag{6}$$

where w is a weighting function associated with ρ. Weights $w(e_\ell^{(k)})$ need to be reevaluated after iteration k to be used at iteration $k+1$.

In this paper, we shall focus on Welsch M-estimator defined as:

$$\rho(e) = \frac{c^2}{2}\left[1 - \exp\left(-\frac{e^2}{c^2}\right)\right] \tag{7}$$

with c a parameter to be set. It is closely related to correntropy [14], which was used with NMF for the linear unmixing [15].

3. ROBUST NONLINEAR UNMIXING

3.1. Problem formulation

Instead of minimizing the square error loss function, we now consider the problem of minimizing $\sum_{\ell=1}^{L} -\exp(-e_\ell^2/c^2)$ derived from Welsch M-estimator. Similar to (3), we state the nonlinear unmixing problem as follows:

$$\boldsymbol{\alpha}^o, \psi^o = \underset{\boldsymbol{\alpha}, \psi \in \mathcal{H}}{\arg\min} J(\boldsymbol{\alpha}, \psi)$$

with

$$J(\boldsymbol{\alpha}, \psi) = \frac{1}{2\mu}\sum_{\ell=1}^{L}\left[-\exp\left(-\frac{e_\ell^2}{c^2}\right)\right] + \frac{1}{2}\|\boldsymbol{\alpha}\|^2 + \frac{1}{2}\|\psi\|_{\mathcal{H}}^2 \tag{8}$$

subject to

$$e_\ell = r_\ell - \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} - \psi(\boldsymbol{m}_{\lambda_\ell})$$

$$\boldsymbol{\alpha} \succeq 0 \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{\alpha} = 1.$$

Also, interested readers can refer to [7, 9] for including the spatial regularization into this nonlinear unmixing context.

3.2. Problem solving

We shall now present an alternating optimization algorithm to solve problem (8). Consider first the following result that is concerned with the convex conjugate of the exponential function.

Let $f(x) = \exp(x)$. Its convex conjugate $f^*(x)$ is given by:

$$f^*(x) = \max_t \{xt - f(t)\}$$

$$= \begin{cases} x\log(x) - x & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ \infty & \text{if} \quad x < 0 \end{cases} \tag{9}$$

On the other hand, since $f(x)$ is convex and continuous, we have:

$$f(x) = \max_t \{xt - f^*(t)\} \tag{10}$$

This implies that:

$$-f\left(-\frac{x^2}{c^2}\right) = \min_t\left\{\frac{x^2}{c^2}t + f^*(-t)\right\} \tag{11}$$

with c any non-zero constant. Using (9) we obtain that, given x, the minimum over t in (11) is reached for $t = \exp(-x^2/c^2)$. Note that the left hand term of (9) corresponds to the first term of the objective function $J(\boldsymbol{\alpha}, \psi)$ in (8).

Substituting (11) into (8), we get the following objective function in an augmented parameter space:

$$J_{\text{aug}}(\boldsymbol{\alpha}, \psi, \boldsymbol{t}) = \frac{1}{2\mu}\sum_{\ell=1}^{L}\left[\frac{e_\ell^2}{c^2}t_\ell + f^*(-t_\ell)\right] + \frac{1}{2}\|\boldsymbol{\alpha}\|^2 + \frac{1}{2}\|\psi\|_{\mathcal{H}}^2 \tag{12}$$

with $\boldsymbol{t} = [t_1, \ldots, t_L]$. For fixed $\boldsymbol{\alpha}$ and ψ, we have:

$$J(\boldsymbol{\alpha}, \psi) = \min_{\boldsymbol{t}} J_{\text{aug}}(\boldsymbol{\alpha}, \psi, \boldsymbol{t}) \tag{13}$$

Then a two-step alternating minimization scheme can be used for solving (8) [16, 17]:

$$\boldsymbol{t}^{(k)} = \underset{\boldsymbol{t}}{\arg\min} J_{\text{aug}}(\boldsymbol{\alpha}^{(k-1)}, \psi^{(k-1)}, \boldsymbol{t}) \tag{14}$$

$$\boldsymbol{\alpha}^{(k)}, \psi^{(k)} = \underset{\boldsymbol{\alpha}, \psi \in \mathcal{H}}{\arg\min} J_{\text{aug}}(\boldsymbol{\alpha}, \psi, \boldsymbol{t}^{(k)}). \tag{15}$$

The resulting minimization method is called half-quadratic and is equivalent to the gradient linearization iteration. The subproblems associated with the two steps can be expressed as follows, where the iteration index (k) is dropped for clearance:

- Subproblem I: For fixed $\boldsymbol{\alpha}$ and ψ, minimize $J_{\text{aug}}(\boldsymbol{\alpha}, \psi, \boldsymbol{t})$ with respect to $\boldsymbol{t}$. Equations (9) and (11) provide a closed form solution for $\ell = 1, \ldots, L$:

$$t_\ell = \exp\left(-\frac{(r_\ell - \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} - \psi(\boldsymbol{m}_{\lambda_\ell}))^2}{c^2}\right). \tag{16}$$

- Subproblem II: For fixed $\boldsymbol{t}$, minimize $J_{\text{aug}}(\boldsymbol{\alpha}, \psi, \boldsymbol{t})$ with respect to $\boldsymbol{\alpha}$ and ψ, namely, solve the following problem:

$$\boldsymbol{\alpha}^o, \psi^o = \underset{\boldsymbol{\alpha}, \psi \in \mathcal{H}}{\arg\min} \frac{1}{2\mu}\sum_{\ell=1}^{L}\frac{t_\ell}{c^2}e_\ell^2 + \frac{1}{2}\|\boldsymbol{\alpha}\|^2 + \frac{1}{2}\|\psi\|_{\mathcal{H}}^2$$

$$\text{subject to} \quad e_\ell = r_\ell - \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} - \psi(\boldsymbol{m}_{\lambda_\ell})$$

$$\boldsymbol{\alpha} \succeq 0 \quad \text{and} \quad \boldsymbol{1}^\top \boldsymbol{\alpha} = 1 \tag{17}$$

Related solving method is described in the next subsection.

Subproblems I and II are solved iteratively until the convergence is achieved. Specifically, problem (17) is a weighted version of (3) where the fitting error at band λ_ℓ is weighted by the quantity t_ℓ/c^2 that is related to the error magnitude. The algorithm assigns lower weights for bands with significant fitting errors caused either by high-level noise or by model mismatch. This property improves the robustness of the algorithm.

3.3. Solving subproblem II

We now present the method of solving subproblem II via its Lagrange duality. By introducing the Lagrange multipliers β_ℓ, γ_r and λ with $\ell = 1, \ldots, L$ and $r = 1, \ldots, R$, the Lagrange function associated with (17) is given by

$$\mathcal{L} = \frac{1}{2}\Big[\frac{1}{\mu}\sum_{\ell=1}^{L}\frac{t_\ell}{c^2}e_\ell^2 + \|\boldsymbol{\alpha}\|^2 + \|\psi\|_{\mathcal{H}}^2\Big] - \sum_{\ell=1}^{L}\beta_\ell(e_\ell - r_\ell$$
$$+ \boldsymbol{\alpha}^\top \boldsymbol{m}_{\lambda_\ell} + \psi(\boldsymbol{m}_{\lambda_\ell})) - \sum_{r=1}^{R}\gamma_r\alpha_r + \lambda(\mathbf{1}^\top\boldsymbol{\alpha} - 1) \quad (18)$$

with $\gamma_r \geq 0$. The conditions for optimality of $\mathcal{L}$ are

$$\begin{cases} \boldsymbol{\alpha}^o = \sum_{\ell=1}^{L}\beta_\ell^o\,\boldsymbol{m}_{\lambda_\ell} + \boldsymbol{\gamma}^o - \lambda^o\mathbf{1} \\ \psi^o = \sum_{\ell=1}^{L}\beta_\ell^o\,\kappa(\cdot,\,\boldsymbol{m}_{\lambda_\ell}) \\ e_\ell^o = \mu\,\beta_\ell^o/(\frac{t_\ell}{c^2}) \end{cases} \quad (19)$$

where κ denotes the reproducing kernel of $\mathcal{H}$. By substituting (19) into (18), we get the dual problem (20) (see top of the next page), where $\boldsymbol{K}$ is the Gram matrix defined as $[\boldsymbol{K}]_{\ell p} = \kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_p})$. Problem (20) is a concave quadratic programming problem with respect to the dual variables. Finally, with the optimal dual variables β_ℓ^o, $\boldsymbol{\gamma}_r^o$ and λ^o, the vector of fractional abundances is evaluated by

$$\boldsymbol{\alpha}^o = \boldsymbol{M}\boldsymbol{\beta}^o + \boldsymbol{\gamma}^o - \lambda^o\mathbf{1}, \quad (21)$$

where $\boldsymbol{\beta}^o = [\beta_1^o, \ldots, \beta_L^o]^\top$ and $\boldsymbol{\gamma} = [\gamma_1^o, \ldots, \gamma_R^o]^\top$. Each pixel can consequently be reconstructed via $\boldsymbol{r}^o = \boldsymbol{M}\boldsymbol{\alpha}^o + \boldsymbol{K}\boldsymbol{\beta}^o$.

4. EXPERIMENTAL RESULTS

4.1. Experiments with synthetic images

First, we validated the algorithm with synthetic data. Five endmembers were selected from the ENVI software library. These spectra consist of 210 contiguous bands, from 0.3951 to 2.56 micrometers. The reflectance vectors were generated with two nonlinear mixture models. The first one was the bilinear model defined as $\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \sum_{i=1}^{R}\sum_{j=i+1}^{R}\alpha_i\,\alpha_j\,\boldsymbol{m}_i \otimes \boldsymbol{m}_j + \boldsymbol{v}$, with $\otimes$ denoting the Hadamard product. The second one was a post-nonlinear model (PNMM) given by $\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^{0.7} + \boldsymbol{v}$. The data were corrupted by zero-mean Gaussian noises $\boldsymbol{v}$ with band dependent standard deviations depicted in Figs. 1(a)-1(c). The first case corresponds to an i.i.d. Gaussian noise. The second case corresponds to a noise with large standard deviations at 30th, 100th and 200th bands. The last case corresponds to bell-shaped band-dependent standard deviations. The following typical algorithms were compared in order to illustrate the performance of the proposed algorithm. Their tuning parameters were set by preliminary experiments: 1) The linear unmixing method [3] (FCLS); 2) The bilinear unmixing method with extended endmember matrix [3] (BExtM); 3) The nonlinear unmixing algorithm K-Hype [6]. Parameter μ was set to 0.01, and the Gaussian kernel was used with $\sigma = 2$; 4) The proposed robust K-Hype algorithm where c was set to 0.5. Parameter μ was set to the same value as K-Hype (see above) divided by c^2

for equivalence. For each scenario, $N = 2500$ pixels were generated to evaluate the Root Mean Square Error (RMSE) defined by RMSE $= \sqrt{\frac{1}{NR}\sum_{n=1}^{N}\|\boldsymbol{\alpha}_n - \boldsymbol{\alpha}_n^o\|^2}$. The results are reported in Table 1. One can observe that FCLS has large estimation errors because the mixture models are nonlinear. BExtM improves the estimation accuracy specifically in the bilinear case. As it is based on a generic kernel-based framework, K-Hype is a flexible algorithm that allows to efficiently address a number of nonlinear unmixing problems. In these experiments, it works significantly better than BExtM. However, its performance is degraded by non-i.i.d. noises. Finally, the proposed algorithm is robust against non-i.i.d. noises and outperforms K-Hype, specially for the 2nd case where some bands are corrupted by high-level noise. The averaged values of $1/t_\ell$ over 2500 pixels for these three cases are provided in Figs. 1(d)-1(f). These error-related reweighing factors are consistent with noise standard deviations.

4.2. Experiments with AVIRIS data

Finally, we applied the proposed algorithm to the well-known hyperspectral image captured on the Cuprite mining district by AVIRIS. A sub-image of 250×191 pixels was chosen to evaluate the algorithm. Spectra consist of $L = 188$ spectral bands. Twelve endmembers were used as in [6]. Parameter μ was fixed to $2 \cdot 10^{-3}$. The other parameters were the same as for the experiments on synthetic data. The averaged spectral angle $\Theta = \frac{1}{N}\sum_{n=1}^{N}\theta(\boldsymbol{r}_n, \boldsymbol{r}_n^o)$ between original $\boldsymbol{r}_n$ and reconstructed $\boldsymbol{r}_n^o$ pixel vectors was used to evaluate the performance, where N denotes the number of processed pixels and $\theta(\boldsymbol{r}, \boldsymbol{r}^o) = \cos^{-1}(\boldsymbol{r}^\top\boldsymbol{r}^o/\|\boldsymbol{r}\|\|\boldsymbol{r}^o\|)$. The results are shown in Tab. 2, and the abundance maps of selected materials are illustrated in Fig. 2. The proposed approach achieves better performance than FCLS and BExtM, but similar one as K-Hype. The proposed approach however provides an additional information on the level of noise in each band, which is rather constant as the range of $1/t_\ell$ is small. This justifies the similarity of results for this scene.

5. CONCLUSION

In this paper, we proposed a robust algorithm for nonlinear unmixing of hyperspectral images based on Welsch M-estimator. The solving process is based on an efficient half-quadratic optimization framework, which marginally modifies K-Hype algorithm. Extensions to other M-estimators such as log-cosh, Fair and Huber will be considered in future works.

6. REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Sig. Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

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

[4] 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 approches," *IEEE J. Sel. Topic Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

$$\max_{\boldsymbol{\beta},\boldsymbol{\gamma},\lambda} \mathcal{L}'(\boldsymbol{\beta},\boldsymbol{\gamma},\lambda) = -\frac{1}{2} \left(\begin{array}{c} \boldsymbol{\beta} \\ \hline \boldsymbol{\gamma} \\ \hline \lambda \end{array} \right)^{\top} \left(\begin{array}{c|c|c} \boldsymbol{K}_\psi & \boldsymbol{M} & -\boldsymbol{M}\mathbf{1} \\ \hline \boldsymbol{M}^{\top} & \mathbf{I} & -\mathbf{1} \\ \hline -\mathbf{1}^{\top}\boldsymbol{M}^{\top} & -\mathbf{1}^{\top} & R \end{array} \right) \left(\begin{array}{c} \boldsymbol{\beta} \\ \hline \boldsymbol{\gamma} \\ \hline \lambda \end{array} \right) + \left(\begin{array}{c} \boldsymbol{r} \\ \hline 0 \\ \hline -1 \end{array} \right)^{\top} \left(\begin{array}{c} \boldsymbol{\beta} \\ \hline \boldsymbol{\gamma} \\ \hline \lambda \end{array} \right) \tag{20}$$

$$\text{subject to} \quad \boldsymbol{\gamma} \succeq \mathbf{0}$$

$$\text{with} \quad \boldsymbol{K}_\psi = \left(\boldsymbol{K} + \mu \operatorname{diag} \left\{ c^2/t_1, \ldots, c^2/t_L \right\} \right) + \boldsymbol{M}\boldsymbol{M}^{\top}$$

Table 1. RMSE comparison with the synthetic data.

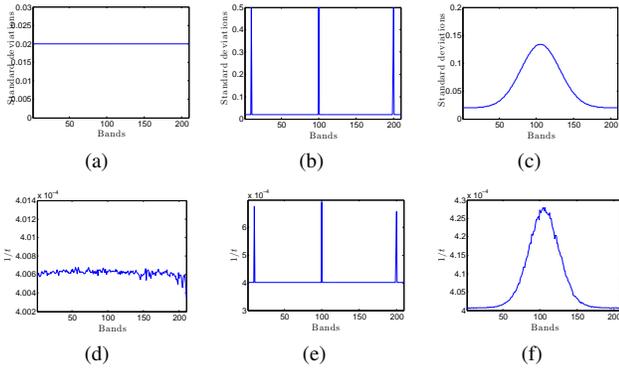|	Bilinear			PNMM		
	Noise 1	Noise 2	Noise 3	Noise 1	Noise 2	Noise 3
FCLS	0.2395±0.0192	0.2417±0.0202	0.2398±0.0195	0.1872±0.0060	0.1909±0.0119	0.1886±0.0084
BExtM	0.0581±0.0025	0.1027±0.0089	0.0724±0.0041	0.1561±0.0046	0.1686±0.0136	0.1619±0.0081
K-Hype	0.0279±0.0006	0.0614±0.0045	0.0595±0.0033	0.0476±0.0016	0.0741±0.0057	0.0752±0.0050
Proposed	0.0274±0.0006	0.0334±0.0009	0.0572±0.0032	0.0474±0.0017	0.0509±0.0019	0.0730±0.0048

(a)   (b)   (c)

(d)   (e)   (f)

Fig. 1. Noise standard deviations (first row) and values of $1/t$ averaged over all pixels (second row). From left to right: three types of noise standard deviation distribution.

Table 2. Spectral angles between original and reconstructed pixels.

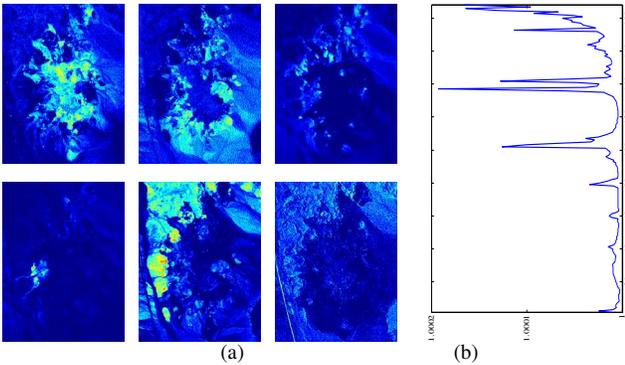|	FCLS	BExtM	K-Hype	Proposed
Θ	0.0136	0.0123	0.0070	0.0070

(a)   (b)

Fig. 2. (a) Abundances maps of selected materials. From top to bottom, left to right: chalcedony, alunite, kaolinite, buddingtonite, sphene, US highway 95. (b) Values of $1/t$ averaged over pixels.

[5] 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," *IEEE Sig. Process. Mag. Process.*, vol. 31, no. 1, pp. 82–94, Jan. 2014.

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

[7] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, May 2014.

[8] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral images with multi-kernel learning," in *Proc. IEEE WHISPERS*, Shanghai, China, Jun. 2012, pp. 1–4.

[9] J. Chen, C. Richard, A. Ferrari, and P. Honeine, "Nonlinear unmixing of hyperspectral data with partially linear least-squares support vector regression," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013, pp. 2174–2178.

[10] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

[11] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Residual component analysis of hyperspectral images–Application to joint nonlinear unmixing and nonlinearity detection," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2148–2158, May 2014.

[12] P. Huber, *Robust statistics*, Wiley, New York, 1981.

[13] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York, 1986.

[14] J. C. Principe, P. P. Pokharel, I. Santamaria, J. Xu, K. H. Jeong, and W. Liu, "Correntropy for random processes: properties, and applications in signal processing," in *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives (Information Science and Statistics)*. Springer Verlag, 2010.

[15] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy based metric," *IEEE Trans. Image Process.*, 2015, (to appear).

[16] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comp.*, vol. 27, no. 3, pp. 937–966, Oct. 2005.

[17] M. Nikolova and R. H. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1623–1627, Jun. 2007.