

ADMM for Maximum Correntropy Criterion

Fei Zhu, Abderrahim Halimi, Paul Honeine, Badong Chen, Nanning Zheng

▶ **To cite this version:**

Fei Zhu, Abderrahim Halimi, Paul Honeine, Badong Chen, Nanning Zheng. ADMM for Maximum Correntropy Criterion. Proc. 28th (INNS and IEEE-CIS) International Joint Conference on Neural Networks, 2016, Vancouver, Canada. pp.1420-1427, 10.1109/IJCNN.2016.7727365 . hal-01965916

HAL Id: hal-01965916
https://hal.science/hal-01965916

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.

ADMM for Maximum Correntropy Criterion

Fei Zhu[(1)], Abderrahim Halimi[(2)], Paul Honeine[(3)], Badong Chen[(4)], Nanning Zheng[(4)]

[(1)]Institut Charles Delaunay (CNRS), Université de Technologie de Troyes, France.
[(2)]School of Engineering and Physical Sciences, Heriot-Watt University, U.K.
[(3)]LITIS lab, Université de Rouen, France.
[(4)]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China.

fei.zhu@utt.fr ; a.halimi@hw.ac.uk ; paul.honeine@univ-rouen.fr ; chenbd@mail.xjtu.edu.cn ; nnzheng@mail.xjtu.edu.cn

***Abstract*—The correntropy provides a robust criterion for outlier-insensitive machine learning, and its maximisation has been increasingly investigated in signal and image processing. In this paper, we investigate the problem of unmixing hyperspectral images, namely decomposing each pixel/spectrum of a given image as a linear combination of other pixels/spectra called endmembers. The coefficients of the combination need to be estimated subject to the nonnegativity and the sum-to-one constraints. In practice, some spectral bands suffer from low signal-to-noise ratio due to acquisition noise and atmospheric effects, thus requiring robust techniques for the unmixing problem. In this work, we cast the unmixing problem as the maximization of a correntropy criterion, and provide a relevant solution using the alternating direction method of multipliers (ADMM) method. Finally, the relevance of the proposed approach is validated on synthetic and real hyperspectral images, demonstrating that the correntropy-based unmixing is robust to outlier bands.**

***Index Terms*—Correntropy, maximum correntropy estimation, alternating direction method of multipliers, hyperspectral image, unmixing problem.**

I. Introduction

Spectral unmixing is an essential issue in many disciplines, including machine learning and signal/image processing, with a wide range of applications, such as classification, segmentation, material identification and target detection. Typically, a hyperspectral image corresponds to a scene taken at many continuous and narrow bands across a certain wavelength range; namely, each pixel is a spectrum. Assuming that each spectrum is a mixture of several pure materials, the unmixing problem consists in identifying these pure materials (the so-called *endmembers*) and estimating their proportions (the so-called *abundances*) at each pixel [1]. While the endmember extraction is relatively easy from geometry (see [2] and references therein), the abundance estimation is still an open problem. Indeed, the abundances can be estimated using least-squares methods, geometric approaches [3], [2], or by including nonlinearities with kernel-based machines [4], [5], [6], [7], [8], [9]. In this paper, we consider the abundance estimation in unmixing hyperspectral images.

The linear mixture model (LMM) is the most investigated over the past decades [10], [11], [12]. Its underlying premise is that each pixel/spectrum is a linear combination of the endmembers. To be physically interpretable, two constraints are imposed: the abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC) for each pixel [13]. Considering both constraints, the fully-constrained least-squares method (FCLS) was developed in [10]. Another well-known algorithm is SUnSAL (for *Sparse Unmixing by variable Splitting and Augmented Lagrangian*). It addresses the same optimization problem by taking advantage of the alternating direction method of multipliers (ADMM) [14].

All these unmixing algorithms hugely suffer from noisy data and outliers within bands. Indeed, in real hyperspectral images for remote sensing, a considerable proportion (about 20%) of the spectral bands are noisy with low SNR, due to the atmospheric effect such as water absorption [15]. These bands need to be removed prior to applying any existing unmixing method; otherwise, the unmixing quality drastically decreases. Such sensitivity to outliers is due to the investigated ℓ_2-norm as a cost function in the FCLS and SUnSAL algorithms, as well as all unmixing algorithms that explore least-squares

solutions.

Information theoretic learning provides an elegant alternative to the conventional minimization of the ℓ_2-norm in least-squares problems, by considering the maximization of the so-called correntropy [16], [17], [18], [19]. Due to its stability and robustness to noise and outliers, the correntropy maximization is based on theoretical foundations and has been successfully applied to a wide class of applications, including cancer clustering [20], face recognition [21], and recently hyperspectral unmixing [22], to name a few. In these works, the resulting problem is optimized by the half-quadratic technique [23], either in a supervised manner [21] or as an unsupervised nonnegative matrix factorization [20], [22].

In this paper, we derive the the fully-constrained (ANC and ASC) correntropy maximization problem for unmixing hyperspectral images. We propose to solve this problem with the ADMM algorithm. Indeed, the ADMM approach splits a hard problem into a sequence of small and handful ones [24]. Its relevance to solve nonconvex problems was studied in [24, Section 9]. We show that ADMM provides a relevant framework for incorporating different constraints raised in unmixing problem. The resulting algorithm is called CUSAL for *Correntropy-based Unmixing by variable Splitting and Augmented Lagrangian*. The rest of the paper is organized as follows. We first revisit the classical unmixing problems with the linear model in Section II. In Section III, we present the proposed correntropy-based unmixing problem. The resulting optimization problem is solved by the ADMM in Section IV. Experiments on synthetic and real images are presented in Section V and VI, respectively. Finally, Section VII provides some conclusions and future works.

II. Classical Unmixing Problem

The linear mixture model (LMM) assumes that each spectrum can be expressed as a linear combination of a set of pure material spectra, termed endmembers [1]. Consider a hyperspectral image and let $\boldsymbol{Y} \in \mathbb{R}^{L\times T}$ denote the matrix of the T pixels/spectra of L spectral bands. Let $\boldsymbol{y}_{*t}$ be its t-th column and $\boldsymbol{y}_{l*}$ its l-th row, representing the l-th band of all pixels. For notation simplicity, we denote $\boldsymbol{y}_t = \boldsymbol{y}_{*t}$, for $t = 1, 2, \ldots, T$. The LMM can be written as

$$\boldsymbol{y}_t = \sum_{r=1}^{R} x_{rt}\, \boldsymbol{m}_r + \boldsymbol{n}_t = \boldsymbol{M}\boldsymbol{x}_t + \boldsymbol{n}_t, \qquad (1)$$

where $\boldsymbol{M} = [\boldsymbol{m}_1 \;\; \boldsymbol{m}_2 \;\; \cdots \;\; \boldsymbol{m}_R] \in \mathbb{R}^{L\times R}$ is the matrix composed by the R endmembers, $\boldsymbol{x}_t = [x_{1t} \;\; x_{2t} \;\; \cdots \;\; x_{Rt}]^\top$ is the abundance vector associated with the t-th pixel, and $\boldsymbol{n}_t \in \mathbb{R}^L$ is the additive noise. In matrix form for all pixels, we have

$$\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{X} + \boldsymbol{N},$$

where $\boldsymbol{X} = [\boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \cdots \;\; \boldsymbol{x}_T] \in \mathbb{R}^{R\times T}$ and $\boldsymbol{N}$ is the additive noise matrix.

In the following, the endmembers are assumed known, either from ground-truth information or by using any endmember extraction technique [2]. The spectral unmixing problem consists in estimating the abundances for each pixel, often by solving the least-squares optimization problem

$$\min_{\boldsymbol{x}_t} \|\boldsymbol{y}_t - \boldsymbol{M}\boldsymbol{x}_t\|_2^2,$$

for each $t = 1, 2, \ldots, T$, where $\|\cdot\|_2$ denotes the conventional ℓ_2-norm. To be physically interpretable as emphasized in [1], the fractional abundances should be nonnegative (ANC) and satisfy the sum-to-one constraint (ASC). Considering both constraints, the fully-constrained least-squares problem [10] is formulated as

$$\begin{aligned} &\min_{\boldsymbol{x}_t} \|\boldsymbol{y}_t - \boldsymbol{M}\boldsymbol{x}_t\|_2^2, \\ &\text{subject to } \boldsymbol{x}_t \geq 0 \text{ and } \mathbf{1}^\top \boldsymbol{x}_t = 1, \text{for } t = 1, \ldots, T, \end{aligned}$$

where $\mathbf{1} \in \mathbb{R}^{R\times 1}$ denotes the column vector of ones. Since there is no closed-form solution when dealing with the nonnegativity constraint, several iterative techniques have been proposed, such as the *active set* scheme with the Lawson and Hanson's algorithm [25], the multiplicative iterative strategies [26], and the fully-constrained least-squares (FCLS) technique [27]. More recently, the alternating direction method of multipliers (ADMM) has been applied with success for hyperspectral unmixing problem, with the so-called SUnSAL algorithm [14].

On sensitivity to outliers

All the aforementioned algorithms rely on solving a (constrained) least-squares optimization problem, thus inheriting the drawbacks of using the ℓ_2-norm as the fitness measure. A major drawback is its sensitivity to outliers, where outliers are some spectral bands that largely deviate from the rest of the bands. Indeed, considering all the image pixels, the least-squares optimization problems take the form

$$\min_{\boldsymbol{X}} \sum_{l=1}^{L} \|\boldsymbol{y}_{l*} - (\boldsymbol{M}\boldsymbol{X})_{l*}\|_2^2, \tag{2}$$

subject to any of the aforementioned constraints. In this expression, $(\cdot)_{l*}$ denotes the l-th row of its argument. From this formulation, it is easy to see how the squared ℓ_2-norm gives more weight to large residuals, namely to outliers in which predicted values $(\boldsymbol{M}\boldsymbol{X})_{l*}$ are far from actual observations $\boldsymbol{y}_{l*}$. Moreover, it is common for hyperspectral images to present up to 20% of unusable spectral bands due to low signal-to-noise ratio essentially from atmospheric effects, such as water absorption. In the following section, we propose to overcome this difficulty by considering the correntropy maximization principle from the information theoretic learning, which yields an optimization problem that is robust to outliers.

III. The Correntropy-based Unmixing Problem

In this section, we examine the correntropy and rewrite the hyperspectral unmixing problem as a correntropy maximization one.

A. Correntropy

The correntropy, proposed in [16], [17], is a nonlinear, local similarity measure. For two random variables U and V, the correntropy is defined by

$$\mathbb{E}[\kappa(U, V)], \tag{3}$$

where $\mathbb{E}[\cdot]$ is the expectation operator, and $\kappa(\cdot, \cdot)$ is a shift-invariant kernel satisfying the Mercer theorem [28]. In practice, while the joint distribution function of U and V is often unavailable, the sample estimator of correntropy is adopted instead. Employing a finite number of data $\{(u_l, v_l)\}_{l=1}^{L}$, the latter is estimated by

$$\frac{1}{L}\sum_{l=1}^{L} \kappa(u_l, v_l), \tag{4}$$

up to a normalization factor. Regarding the kernel function, the Gaussian kernel is the most commonly-used kernel for correntropy [16], [21], [29]. It is defined by $\kappa(u, v) = \exp(\frac{-1}{2\sigma^2}\|u - v\|^2)$, where σ denotes the kernel bandwidth. This leads to the following expression for the correntropy:

$$\frac{1}{L}\sum_{l=1}^{L} \exp\left(\tfrac{-1}{2\sigma^2}\|u_l - v_l\|_2^2\right). \tag{5}$$

The maximization of the similarity measured by the correntropy, given by $\max \frac{1}{L}\sum_{l=1}^{L} \kappa(u_l, v_l)$, is termed the maximum correntropy criterion (MCC) [16], [30]. It is noteworthy that well-known second-order statistics, such as the mean square error (MSE) depends heavily on the Gaussian and linear assumptions [16]. However, in presence of non-Gaussian noise and in particular large outliers, *i.e.*, observations greatly deviated from the data bulk, the effectiveness of the MSE-based algorithms will significantly deteriorate [31]. By contrast, the MCC-based algorithms is appropriate for non-Gaussian signal processing, and is robust in particular against large outliers, as shown next.

We examine the behavior of the correntropy in terms of the residual error defined by $\epsilon_l = \|u_l - v_l\|$. Therefore, the correntropy (5) becomes $\frac{1}{L}\sum_{l=1}^{L} \exp\left(\frac{-1}{2\sigma^2}\epsilon_l^2\right)$. Compared with the second-order statistics, *e.g.* MSE, the correntropy is more robust with respect to the outliers, as shown in Fig. 1 illustrating the second-order and the correntropy objective functions in terms of the residual error. As the residual error augments, the second-order function keeps increasing dramatically. On the contrary, the correntropy is only sensitive within a region of small residual errors, this region being controlled by the kernel bandwidth. For large magnitudes of residual error out of the central range, the correntropy grows extremely gently and tends to a constant. Consequently, the correntropy criterion is robust to large outliers, especially for small kernel bandwidths.

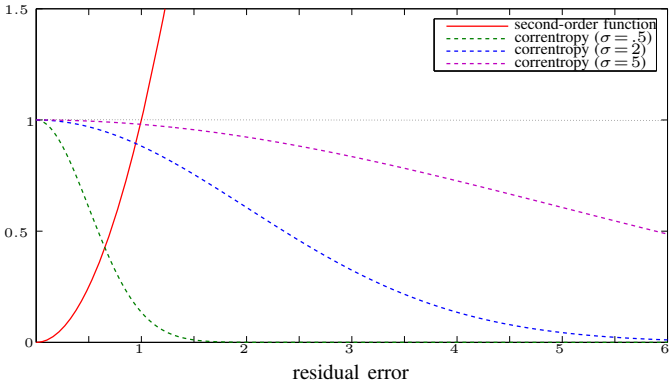


Fig. 1. Illustration of second-order objective function (solid line) and the correntropy objective function (dashed lines) with different values of the kernel bandwidth.

B. Correntropy-based unmixing problem

The correntropy-based unmixing problem consists in estimating the unknown abundance matrix $\boldsymbol{X}$, by minimizing the objective function $\mathcal{C}$ (the negative of correntropy), given by

$$\mathcal{C}(\boldsymbol{X}) = -\sum_{l=1}^{L} \exp\left(\tfrac{-1}{2\sigma^2}\|\boldsymbol{y}_l - (\boldsymbol{M}\boldsymbol{X})_{l*}\|_2^2\right), \quad (6)$$

where the Gaussian kernel is considered and the nonnegative constraint is enforced on all the entries of $\boldsymbol{X}$. Considering both the ANC and ASC constraints, the fully-constrained case for problem (6) becomes

$$\begin{aligned} &\min_{\boldsymbol{X}} \ \mathcal{C}(\boldsymbol{X}), \\ &\text{subject to } \boldsymbol{X} \geq 0 \text{ and } \mathbf{1}^\top \boldsymbol{x}_t = 1, \text{for } t = 1, \ldots, T. \end{aligned} \quad (7)$$

IV. ADMM for Solving the Correntropy-based Unmixing Problem

We first briefly review the alternating direction method of multipliers (ADMM), following the expressions in [24, Chap. 3]. Assuming that the functions f and g are closed, proper and convex, the ADMM solves the problem

$$\begin{aligned} &\min \quad f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ &\text{subject to} \quad A\boldsymbol{x} + B\boldsymbol{z} = \boldsymbol{c}, \end{aligned} \quad (8)$$

with variables $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{z} \in \mathbb{R}^m$, where $A \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{p\times m}$ and $\boldsymbol{c} \in \mathbb{R}^p$. The algorithm ADMM, in its scaled form, is summarized in Algorithm 1.

Algorithm 1 The Alternating Direction Method of Multipliers (ADMM) [24]

Input: functions f and g, matrices A and B, vector $\boldsymbol{c}$, parameter ρ

1: Initialize $\boldsymbol{x}_0$, $\boldsymbol{z}_0$ and $\boldsymbol{u}_0$
2: **repeat**
3: $\quad \boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\rho}{2}\|A\boldsymbol{x} + B\boldsymbol{z}_k - \boldsymbol{c} + \boldsymbol{u}_k\|_2^2$;
4: $\quad \boldsymbol{z}_{k+1} = \arg\min_{\boldsymbol{z}} g(\boldsymbol{z}) + \frac{\rho}{2}\|A\boldsymbol{x}_{k+1} + B\boldsymbol{z} - \boldsymbol{c} + \boldsymbol{u}_k\|_2^2$;
5: $\quad \boldsymbol{u}_{k+1} = \boldsymbol{u}_k + A\boldsymbol{x}_{k+1} + B\boldsymbol{z}_{k+1} - \boldsymbol{c}$;
6: **until** stopping criterion

In the following, we apply the ADMM algorithm to solve the correntropy-based unmixing problem in the fully-constrained case, presented in (7). The main steps are summarized in Algorithm 2. Rewrite the variables to be optimized in a vector $\boldsymbol{x} \in \mathbb{R}^{RT\times 1}$, which is stacked by the columns of the matrix $\boldsymbol{X}$, namely $\boldsymbol{x} = [\boldsymbol{x}_1^\top \ \boldsymbol{x}_2^\top \ \cdots \ \boldsymbol{x}_T^\top]^\top$. Rewrite the vectors $\boldsymbol{z} = [\boldsymbol{z}_1^\top \ \boldsymbol{z}_2^\top \ \cdots \ \boldsymbol{z}_T^\top]^\top \in \mathbb{R}^{RT\times 1}$ and $\boldsymbol{u} = [\boldsymbol{u}_1^\top \ \boldsymbol{u}_2^\top \ \cdots \ \boldsymbol{u}_T^\top]^\top \in \mathbb{R}^{RT\times 1}$, where $\boldsymbol{z}_t = [z_{1t} \ z_{2t} \ \cdots \ z_{Rt}]^\top$ and $\boldsymbol{u}_t = [u_{1t} \ u_{2t} \ \cdots \ u_{Rt}]^\top$, for $t = 1, 2, \ldots, T$. Following the formulation in Algorithm 1, we set

$$f(\boldsymbol{x}) = \mathcal{C}(\boldsymbol{x}) + \sum_{t=1}^{T} \iota_{\{1\}}(\mathbf{1}^\top \boldsymbol{x}_t) \quad (9)$$

$$g(\boldsymbol{x}) = \iota_{\mathbb{R}_+^{RT}}(\boldsymbol{x}) \quad (10)$$

$$A = -\boldsymbol{I}, B = \boldsymbol{I} \ \text{ and } \ \boldsymbol{c} = \mathbf{0}, \quad (11)$$

where $\boldsymbol{I}$ is the identity matrix, $\mathbf{0} \in \mathbb{R}^{RT\times 1}$ is the zero vector and $\iota_S(\boldsymbol{x})$ denotes the indicator function of the set S defined by

$$\iota_S(\boldsymbol{x}) = \begin{cases} 0, & \text{if } \boldsymbol{x} \in S \\ \infty, & \text{elsewise.} \end{cases}$$

In this case, the subproblem of the x-update in line 3 of Algorithm 1 addresses a nonconvex problem without closed-form solution. To overcome this difficulty, we apply an inexact ADMM variant in lines 3-5 of Algorithm 2, which solves the subproblem iteratively using the gradient descent method, instead of solving it exactly and explicitly. Before that, we eliminate the T equality constraints,

i.e., the sum-to-one constraints, by replacing x_{Rt} with $x_{Rt} = 1 - \sum_{r=1}^{R-1} x_{rt}$, for $t = 1, 2, ..., T$. Let $\overline{\boldsymbol{x}} \in \mathbb{R}^{(R-1)T\times 1}$ be the reduced vector of variables, stacked by $\overline{\boldsymbol{x}}_t = [x_{1t} \;\; x_{2t} \;\; \cdots \;\; x_{(R-1)t}]^\top$, $t = 1, 2, ..., T$. By this means, the function (9) is transformed into the reduced-form

$$f_1(\overline{\boldsymbol{x}}) = \sum_{l=1}^{L} -\exp\left(\frac{-\sum_{t=1}^{T} h_l(\overline{\boldsymbol{x}}_t)^2}{2\sigma^2}\right), \qquad (12)$$

where $h_l(\overline{\boldsymbol{x}}_t) = y_{lt} - m_{lR} - \sum_{p=1}^{R-1}(m_{lp} - m_{lR})x_{pt}$, for $l = 1, 2, ..., L$. The gradient of (12) with respect to $\overline{\boldsymbol{x}}$ is stacked as $\frac{\partial f_1}{\partial \overline{\boldsymbol{x}}} = [\frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_1}^\top \;\; \cdots \;\; \frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_T}^\top]^\top \in \mathbb{R}^{(R-1)T\times 1}$, where $\frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_t} = [\frac{\partial f_1}{\partial \overline{x}_{1t}} \;\; \cdots \;\; \frac{\partial f_1}{\partial \overline{x}_{(R-1)t}}]^\top$, with the entries given by

$$\frac{\partial f_1(\overline{\boldsymbol{x}})}{\partial \overline{x}_{rt}} = \frac{1}{\sigma^2}\sum_{l=1}^{L}(m_{lR} - m_{lr}) \exp\Big(\frac{-\sum_{s=1}^{T} h_l(\overline{\boldsymbol{x}}_s)^2}{2\sigma^2}\Big) h_l(\overline{\boldsymbol{x}}_t),$$

for all $r = 1, 2, ..., (R-1)$ and $t = 1, 2, ..., T$. Similarly, the function $\frac{\rho}{2}\|\boldsymbol{x} - \boldsymbol{z}_k - \boldsymbol{u}_k\|_2^2$ is expressed with respect to $\overline{\boldsymbol{x}}$ as

$$\phi(\overline{\boldsymbol{x}}) = \frac{\rho}{2}\sum_{t=1}^{T}\sum_{r=1}^{R-1}(x_{rt} - z_{rt,k} - u_{rt,k})^2 + \Big(1 - \sum_{r=1}^{R-1} x_{rt} - z_{Rt,k} - u_{Rt,k}\Big)^2,$$

with the entries in its gradient $\frac{\partial \phi}{\partial \overline{\boldsymbol{x}}}$ given by

$$\frac{\partial \phi(\overline{\boldsymbol{x}})}{\partial \overline{x}_{rt}} = \rho(x_{rt} + \sum_{p=1}^{R-1} x_{pt} - 1 + z_{Rt,k} - z_{rt,k} + u_{Rt,k} - u_{rt,k}), \qquad (13)$$

for all $r = 1, 2, ..., (R-1)$ and $t = 1, 2, ..., T$.

The solution of the z-update in line 4 Algorithm 1 is the projection of $\boldsymbol{x}_{k+1} - \boldsymbol{u}_k$ onto the first orthant as shown in line 7 of Algorithm 2.

Algorithm 2 Correntropy unmixing (**CUSAL**)

1: Initialize $k = 0$, $\rho > 0$, $\eta > 0$, $\sigma > 0$, norms of the primal and dual residuals $r_0 = 10^5$ and $s_0 = 10^5$; $\boldsymbol{x}_0$, $\boldsymbol{z}_0$ and $\boldsymbol{u}_0$;
2: **repeat**
3: **repeat**
4: $\overline{\boldsymbol{x}}_{k+1} = \overline{\boldsymbol{x}}_{k+1} - \eta(\frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_{k+1}} + \frac{\partial \phi}{\partial \overline{\boldsymbol{x}}_{k+1}})$;
5: **until** convergence
6: reform $\boldsymbol{x}_{k+1}$ using $\overline{\boldsymbol{x}}_{k+1}$;
7: $\boldsymbol{z}_{k+1} = \max(\boldsymbol{0}, \boldsymbol{x}_{k+1} - \boldsymbol{u}_k)$;
8: $\boldsymbol{u}_{k+1} = \boldsymbol{u}_k - (\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1})$;
9: $r_{k+1} = \|\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1}\|_2$, $s_{k+1} = \rho\|\boldsymbol{z}_{k+1} - \boldsymbol{z}_k\|_2$;
10: $k = k + 1$;
11: **until** stopping criterion

V. Experiments with Synthetic Data

In this section, the performance of the proposed CUSAL is evaluated on synthetic data. A comparative study is performed considering five state-of-the-art methods proposed for both linear and nonlinear unmixing models. We apply a three-fold stopping criterion for Algorithm 2 according to [24], [14]: (i) the primal and dual residuals are small enough, (ii) the primal residual starts to increase, or (iii) the predefined maximum iteration number is attained.

The proposed CUSAL algorithm is compared to the following methods:

- Fully-Constrained Least-Squares (**FCLS**) [10]: The FCLS is developed for the linear model. Enforcing both ANC and ASC constraints, this technique yields the optimal abundance matrix in the least-squares sense. See also [2].
- Sparse Unmixing by variable Splitting and Augmented Lagrangian (**SUnSAL**) [14]: This method is based on the ADMM. Under the linear assumption, several variants are developed by including different constraints. In the following, we consider the fully-constrained variant, termed SUnSAL-FCLS.
- The Bayesian algorithm for Generalized Bilinear Model (**BayGBM**) [32], [33]: The BayGBM is a nonlinear abundance estimation method based on the generalized bilinear model (GBM), which adds the second-order interactions between endmembers to the linear model and is defined as

  $$\boldsymbol{y}_t = \boldsymbol{M}\boldsymbol{x}_t + \sum_{i=1}^{R-1}\sum_{j=i+1}^{R} \gamma_{ij,t} x_{it} x_{jt}(\boldsymbol{m}_i \odot \boldsymbol{m}_j) + \boldsymbol{n}_t,$$

  where $0 \leq \gamma_{ij,t} \leq 1$ controls the interactions between endmembers $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, and $\odot$ is the Hadamard product (element-wise product) operator. The BayGBM is designed considering both ANC and ASC.

- The Bayesian algorithm for Polynomial Postnonlinear Mixing Model (**BayPPNMM**) [34]: The BayPPNMM is a Bayesian-based algorithm proposed to estimate the parameters involved in the polynomial postnonlinear mixing model (PPNMM). This model assumes that the pixel reflectances are nonlinear functions of endmembers, and is defined by

$$\boldsymbol{y}_t = \boldsymbol{M}\boldsymbol{x}_t + b_t(\boldsymbol{M}\boldsymbol{x}_t) \odot (\boldsymbol{M}\boldsymbol{x}_t) + \boldsymbol{n}_t, \quad (14)$$

  where the nonlinear terms are characterized by $b_t \in \mathbb{R}$, and both ANC and ASC are required.
- Kernel Fully-Constrained Least-Squares (**KFCLS**) [4]: This method generalizes FCLS, by replacing the inner product with a kernel function. In the following, the Gaussian kernel is applied for simulation. A preliminary test on 100 randomly selected pixels is performed for each data under study, using the set $\{1, 2, 3, 4, 5\}$ for the parameter σ and the set $\{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ for the parameter that controls the relaxation of the sum-to-one constraint.

A series of experiments are performed, mainly considering the influences of three aspects: (i) noise level, (ii) number of corrupted bands and (iii) number of endmembers.

In the following simulations, each synthetic image has a fixed size of 50×50 pixels. First, it is generated using the linear mixture model (1), where Gaussian noise $\boldsymbol{n}_t$ of SNR $\in \{15, 35\}$dB is added. The $R \in \{3, 6, 9\}$ endmembers are randomly drawn from 20 spectra selected from the United States Geological Survey (USGS) digital spectral library [35]. These candidate endmembers are measured over $L = 244$ continuous bands with the wavelength ranging from $0.2\mu m$ to $3.0\mu m$. The abundance vectors $\boldsymbol{x}_t$ are uniformly generated using a Dirichlet distribution as in [35], [36]. To imitate the noisy bands in the real hyperspectral images, several bands in the generated data are corrupted by replacing the corresponding rows $\boldsymbol{y}_{l*}$ with random values within $[0, 1]$. The number of corrupted bands varies in the set $\{0, 20, 40, 60\}$.

The unmixing performance is evaluated using the abundance root mean square error (RMSE) [32], [37], defined by

$$\text{RMSE} = \sqrt{\frac{1}{RT}\sum_{t=1}^{T} \|\boldsymbol{x}_t - \widehat{\boldsymbol{x}}_t\|^2},$$

where $\widehat{\boldsymbol{x}}_t$ is the estimated abundance vector. Fig. 2 illustrates the average of RMSE over 10 Monte-Carlo realizations. It is easy to see that, in presence of outlier bands, the proposed CUSAL outperforms all the comparing methods in terms of RMSE, for different noise levels and numbers of endmembers. It is also shown that the performance of the proposed algorithm improves when increasing the SNR.

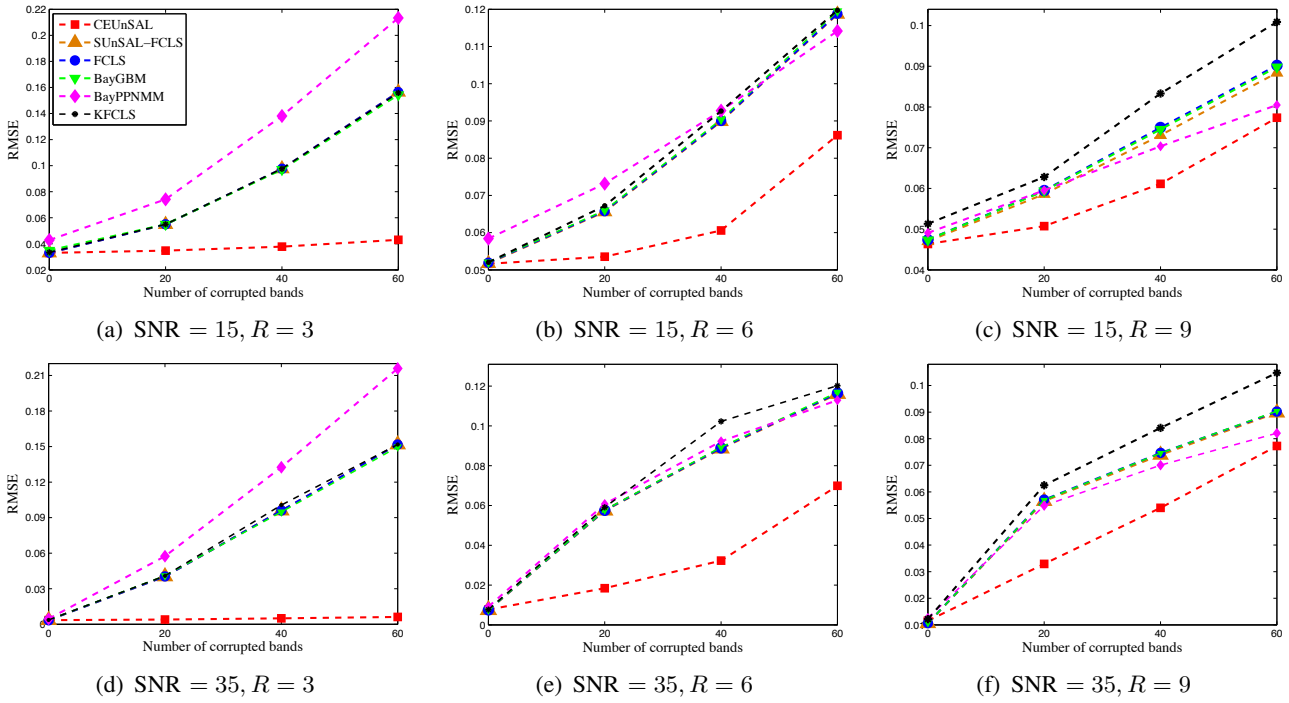
VI. Experiments with Real Data

This section presents the performance of the proposed algorithm on a real hyperspectral image. We consider a 250×190 sub-image taken from the Cuprite mining image, acquired by the AVIRIS sensor when flying over Las Vegas, Nevada, USA. The image has been widely investigated in the literature [5], [38]. The raw data contains $L = 224$ bands, covering a wavelength range $0.4 - 2.5\mu m$. Among, there are 37 relatively noisy ones with low SNR, namely the bands $1-3$, $105-115$, $150-170$, and $223 - 224$. The geographic composition of this area is estimated to include 14 minerals [39]. Neglecting the similar signatures, a reduced number of 12 endmembers is considered [5], [40]. The VCA technique is first applied to extract the endmembers on the clean image with $L = 187$ bands. The resulting 12 pixel indices are retained. Starting from $L = 187$ bands, the noisy bands, randomly chosen from the bands $1 - 3$, $105 - 115$, $150 - 170$, and $223 - 224$, are gradually included to form a series of input data. In practice, the experiments are conducted with $L = 187, 193, 199, 205, 211, 217, 223$ and 224 bands.

The unmixing results in terms of the averaged spectral angle distance (SAD) between the input spectra $\boldsymbol{y}_t$ and the reconstructed ones $\widehat{\boldsymbol{y}}_t$ are illustrated in Fig. 3. The SAD is defined by

$$\text{SAD} = \frac{1}{T}\sum_{t=1}^{T} \arccos\Big(\frac{\boldsymbol{y}_t^\top \widehat{\boldsymbol{y}}_t}{\|\boldsymbol{y}_t\|\|\widehat{\boldsymbol{y}}_t\|}\Big).$$

In absence of noisy bands (*i.e.,* $L = 187$ bands), all the considered methods lead to satisfactory abun-



(a) SNR $= 15, R = 3$ (b) SNR $= 15, R = 6$ (c) SNR $= 15, R = 9$

(d) SNR $= 35, R = 3$ (e) SNR $= 35, R = 6$ (f) SNR $= 35, R = 9$

Fig. 2. The averaged root mean square error (RMSE) with respect to the number of corrupted bands (10 Monte-Carlo realizations). **Left to right:** Various number of endmembers. **Top to bottom:** Different SNR.

dance maps, with BayPPNMM providing the smallest SAD. As the number of noisy bands increases, especially from $L = 199$ to $L = 224$, the unmixing performance of the state-of-the-art methods deteriorates drastically, while the proposed CUSAL yields stable SAD. The obtained results confirm the good behavior of the proposed CUSAL and its robustness in presence of corrupted spectral bands.

VII. Conclusion

This paper presented a supervised unmixing algorithm based on the correntropy maximization principle. The alternating direction method of multipliers (ADMM) was investigated in order to maximize the correntropy criterion under nonnegativity and sum-to-one contraints. The effectiveness and robustness of the proposed method were validated on both synthetic and real hyperspectral images. Future works include the generalization of the correntropy criterion to account for the multiple reflection phenomenon, as well as a sparsity-aware version of the proposed method.

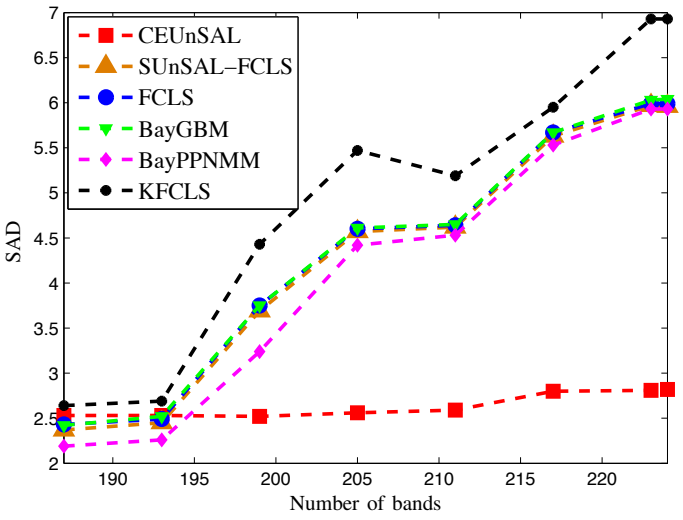


Fig. 3. The SAD values using different numbers of bands on the Cuprite image.

Acknowledgment

This work was supported by the French ANR, grant HYPANEMA: ANR-12BS03-0033.

REFERENCES

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

[2] P. Honeine and C. Richard, "Geometric unmixing of large hyperspectral images: a barycentric coordinate approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2185–2195, Jun. 2012.

[3] ——, "A simple scheme for unmixing hyperspectral data based on the geometry of the n-dimensional simplex," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, Honolulu (Hawaii), USA, 25 - 30 July 2010, pp. 2271–2274.

[4] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*. IEEE, 2007, pp. 4041–4044.

[5] 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.

[6] N. H. Nguyen, J. Chen, C. Richard, P. Honeine, and C. Theys, "Supervised nonlinear unmixing of hyperspectral images using a pre-image method," in *New Concepts in Imaging: Optical and Statistical Models, In Eds. D. Mary, C. Theys, and C. Aime*, ser. EAS Publications Series. EDP Sciences, 2013, vol. 59, pp. 417–437.

[7] 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.

[8] ——, "Estimating abundance fractions of materials in hyperspectral images by fitting a post-nonlinear mixing model," in *Proc. IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing*, Jun. 2013.

[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. 38th IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, May 2013, pp. 2174–2178.

[10] D. Heinz and C. 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.

[11] 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.

[12] 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, March 2007.

[13] 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, 2012.

[14] J. M. Bioucas-Dias and M. A. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*. IEEE, 2010, pp. 1–4.

[15] A. Zelinski and V. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *IEEE International Conference on Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006*, July 2006, pp. 387–390.

[16] W. Liu, P. Pokharel, and J. C. Príncipe, "Correntropy: properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[17] J. C. Principe, *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer Science & Business Media, 2010.

[18] Z. Wu, S. Peng, B. Chen, and H. Zhao, "Robust hammerstein adaptive filtering under maximum correntropy criterion," *Entropy*, vol. 17, no. 10, pp. 7149–7166, 2015.

[19] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Principe, "Convergence of a fixed-point algorithm under maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1723–1727, 2015.

[20] J. J. Wang, X. Wang, and X. Gao, "Non-negative matrix factorization by maximizing correntropy for cancer clustering," *BMC bioinformatics*, vol. 14, no. 1, p. 107, 2013.

[21] R. He, W. Zheng, and B. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1561–1576, 2011.

[22] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy-based metric," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4027–4040, Nov 2015.

[23] M. Nikolova and M. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM Journal on Scientific computing*, vol. 27, no. 3, pp. 937–966, 2005.

[24] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[25] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems (Classics in Applied Mathematics)*. Society for Industrial Mathematics, 1987.

[26] H. Lantéri, M. Roche, O. Cuevas, and C. Aime, "A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints," *Signal Processing*, vol. 81, pp. 945–974, May 2001.

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

[28] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[29] B. Chen and J. C. Príncipe, "Maximum correntropy estimation is a smoothed map estimation," *IEEE Signal Processing Letters*, vol. 19, no. 8, pp. 491–494, 2012.

[30] B. Chen, L. Xing, J. Liang, N. Zheng, and J. C. Príncipe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 880–884, July 2014.

[31] Z. Wu, S. Peng, B. Chen, and H. Zhao, "Robust hammerstein adaptive filtering under maximum correntropy criterion," *Entropy*, vol. 17, no. 10, p. 7149, 2015.

[32] 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.

[33] ——, "Unmixing hyperspectral images using the generalized bilinear model." in *IGARSS*, 2011, pp. 1886–1889.

[34] Y. Altmann, A. Halimi, N. Dobigeon, and J. Y. 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, 2012.

[35] 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.

[36] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4904–4917, december 2015.

[37] 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.

[38] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, June 2011.

[39] J. 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.

[40] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse nmf for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2815–2826, 2013.