



HAL
open science

Correntropy Maximization via ADMM - Application to Robust Hyperspectral Unmixing

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

► **To cite this version:**

Fei Zhu, Abderrahim Halimi, Paul Honeine, Badong Chen, Nanning Zheng. Correntropy Maximization via ADMM - Application to Robust Hyperspectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55 (9), pp.1-12. hal-01965043

HAL Id: hal-01965043

<https://hal.science/hal-01965043v1>

Submitted on 24 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.

Correntropy Maximization via ADMM: Application to Robust Hyperspectral Unmixing

Fei Zhu, Abderrahim Halimi, *Member, IEEE*, Paul Honeine, *Member, IEEE*,
Badong Chen, *Senior Member, IEEE*, and Nanning Zheng, *Fellow, IEEE*

Abstract—In hyperspectral images, some spectral bands suffer from low signal-to-noise ratio due to noisy acquisition and atmospheric effects, thus requiring robust techniques for the unmixing problem. This paper presents a robust supervised spectral unmixing approach for hyperspectral images. The robustness is achieved by writing the unmixing problem as the maximization of the correntropy criterion subject to the most commonly used constraints. Two unmixing problems are derived: the first problem considers the fully constrained unmixing, with both the nonnegativity and sum-to-one constraints, while the second one deals with the nonnegativity and the sparsity promoting of the abundances. The corresponding optimization problems are solved using an alternating direction method of multipliers (ADMM) approach. Experiments on synthetic and real hyperspectral images validate the performance of the proposed algorithms for different scenarios, demonstrating that the correntropy-based unmixing with ADMM is particularly robust against highly noisy outlier bands.

Index Terms—Alternating direction method of multipliers (ADMM), correntropy, hyperspectral image, maximum correntropy estimation, unmixing problem.

I. INTRODUCTION

SPECTRAL unmixing is an essential issue in many disciplines, including signal and 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 narrowbands across a certain wavelength range, namely, each pixel is a spectrum. Assuming that

Manuscript received January 31, 2016; revised August 5, 2016 and January 28, 2017; accepted April 11, 2017. Date of publication June 30, 2017; date of current version August 25, 2017. The work of A. Halimi was supported by the Engineering and Physical Sciences Research Council under Grant EP/N003446/1. The work of F. Zhu and P. Honeine was supported by the French National Research Agency under Grant HYPANEMA: ANR-12BS03-0033. The work of F. Zhu and B. Chen was supported by the National Natural Science Foundation of China under Grant 11626250 and Grant 91648208, respectively. (*Corresponding author: Paul Honeine.*)

F. Zhu is with the Center for Applied Mathematics, Tianjin University, Tianjin 300072, China (e-mail: zhufei.fr@gmail.com).

A. Halimi is with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: a.halimi@hw.ac.uk).

P. Honeine is with the LITIS laboratory, Université de Rouen Normandie (Normandie Université), 76130 Mont-Saint-Aignan, France (e-mail: paul.honeine@univ-rouen.fr).

B. Chen and N. Zheng are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: chenbd@mail.xjtu.edu.cn; nnzheng@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2017.2696262

each spectrum is a mixture of several pure materials, the unmixing problem consists in two tasks: 1) identifying these pure materials (the so-called *endmembers*) and 2) estimating their proportions (the so-called *abundances*) at each pixel [1]. In practice, these two steps can be performed either sequentially or simultaneously [2]. Well-known end-member extraction algorithms include the pure-pixel-based ones, e.g., the vertex component analysis (VCA) [3] and the N-FINDR [4], as well as the minimum-volume-based ones, e.g., the minimum simplex analysis [5] and the minimum volume constrained nonnegative matrix factorization (NMF) [6]. While the endmember extraction is relatively easy from geometry, the abundance estimation remains an open problem. Indeed, the abundances can be estimated using least-squares methods, geometric approaches [2], or by tackling recently raised issues such as nonlinearity [7], [8]. In this paper, we consider the abundance estimation problem.

The linear mixture model (LMM) is the most investigated over the past decades [6], [9], [10]. Its underlying premise is that each pixel/spectrum is a linear combination of the endmembers. To be physically interpretable, two constraints are often enforced in the estimation problem: the abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC) for each pixel [11]. Considering both constraints, the fully constrained least-squares (FCLS) method was presented in [9]. A more recently proposed unmixing algorithm is the so-called *sparse unmixing by variable splitting and augmented Lagrangian* (SUnSAL) [12]. It addresses the same optimization problem by taking advantage of the alternating direction method of multipliers (ADMM) [13]. A constrained version of SUnSAL was also proposed to solve the constrained sparse regression problem, where the ASC constraint is relaxed and the ℓ_1 -norm regularizer is added.

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 signal-to-noise ratio (SNR), due to the atmospheric effect such as water absorption [14]. 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. It is worth

noting that nonlinear unmixing algorithms also suffer from this drawback, including the kernel-based FCLS (KFCLS) [15], nonlinear fluctuation methods [7], and postnonlinear methods [16].

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 [17], [18]. 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 [19], face recognition [20], and recently hyperspectral unmixing [21], to name a few. In these works, the resulting problem is optimized by the half-quadratic technique [22], either in a supervised manner [20] or as an unsupervised NMF [19], [21].

In this paper, we consider the hyperspectral unmixing problem by defining an appropriate correntropy-based criterion, thus taking advantage of its robustness to large outliers, as opposed to the conventional ℓ_2 -norm criteria. By including constraints commonly used for physical interpretation, we propose to solve the resulting constrained optimization problems with ADMM algorithms. Indeed, the ADMM approach splits a hard problem into a sequence of small and handful ones [13]. Its relevance to solve nonconvex problems was studied in [13, Sec. 9]. We show that ADMM provides a relevant framework for incorporating different constraints raised in the unmixing problem. We present the so-called *correntropy-based unmixing by variable splitting and augmented Lagrangian* (CUSAL), and study in particular two algorithms: CUSAL-FC to solve the fully constrained (ANC and ASC) correntropy based unmixing problem, and the CUSAL-SP to solve the sparsity-promoting correntropy-based unmixing problem. In the presence of highly noisy bands, the proposed ADMM method is more robust than classical half-quadratic methods for solving correntropy maximization [19], [21], by alleviating to some degree the parameter estimation problem in the latter (see Section IV-C for more details).

The rest of this paper is organized as follows. We first provide a succinct survey on the classical unmixing problems in Section II. In Section III, we propose the correntropy-based unmixing problems subject to the aforementioned constraints, and study the robustness. The resulting optimization problems are solved by the ADMM algorithms described in Section IV. Experiments on synthetic and real hyperspectral images are presented in Sections V and VI, respectively. Finally, Section VII provides some conclusions and future works.

II. CLASSICAL UNMIXING PROBLEMS

The 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 $\mathbf{Y} \in \mathbb{R}^{L \times T}$ denote the matrix of the T pixels/spectra of L spectral bands. Let \mathbf{y}_{*t} be its t th column and \mathbf{y}_{l*} its l th row, representing the l th band of all pixels. For notation simplicity,

we denote $\mathbf{y}_t = \mathbf{y}_{*t}$, for $t = 1, \dots, T$. The LMM can be written as

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

where $\mathbf{M} = [\mathbf{m}_1 \cdots \mathbf{m}_R] \in \mathbb{R}^{L \times R}$ is the matrix composed by the R endmembers with $\mathbf{m}_r = [m_{1r} \cdots m_{Lr}]^\top$, $\mathbf{x}_t = [x_{1t} \cdots x_{Rt}]^\top$ is the abundance vector associated with the t th pixel, and $\mathbf{n}_t \in \mathbb{R}^L$ is the additive noise. In matrix form for all pixels, we have $\mathbf{Y} = \mathbf{M}\mathbf{X} + \mathbf{N}$, where $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_T] \in \mathbb{R}^{R \times T}$ and \mathbf{N} is the noise matrix.

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

$$\min_{\mathbf{x}_t} \|\mathbf{y}_t - \mathbf{M}\mathbf{x}_t\|_2^2 \quad (2)$$

for each $t = 1, \dots, T$, where $\|\cdot\|_2$ denotes the conventional ℓ_2 -norm. The solution to this conventional least-squares problem is given by the pseudoinverse of the (tall) endmember matrix, with $\mathbf{x}_t = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top \mathbf{y}_t$. The least-squares optimization problems (2), for all $t = 1, \dots, T$, are often written in a single optimization problem using the following matrix formulation:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_F^2 \quad (3)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Its solution is

$$\mathbf{X}_{\text{LS}} = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top \mathbf{Y}. \quad (4)$$

Finally, this optimization problem can also be tackled by considering all the image pixels at each spectral band, which yields the following least-squares optimization problem:

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

where $(\cdot)_{l*}$ denotes the l th row of its argument. While all these problem formulations have a closed-form solution, they suffer from two major drawbacks. The first one is that several constraints need to be imposed in order to have a physical meaning of the results. The second drawback is its sensitivity to noise and outliers, due to the use of the ℓ_2 -norm as a fitness measure. These two drawbacks are detailed in the following.

To be physically interpretable, the abundances should be nonnegative (ANC) and satisfy the sum-to-one constraint (ASC). Considering both constraints, the FCLS problem is formulated as, for each $t = 1, \dots, T$

$$\min_{\mathbf{x}_t} \|\mathbf{y}_t - \mathbf{M}\mathbf{x}_t\|_2^2, \quad \text{s.t. } \mathbf{x}_t \geq 0 \text{ and } \mathbf{1}^\top \mathbf{x}_t = 1$$

where $\mathbf{1} \in \mathbb{R}^{R \times 1}$ denotes the column vector of ones and ≥ 0 is the nonnegativity applied element-wise. In matrix form

$$\begin{aligned} & \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_F^2 \\ & \text{s.t. } \mathbf{X} \geq 0 \text{ and } \mathbf{1}^\top \mathbf{x}_t = 1, \quad \text{for } t = 1, \dots, T. \end{aligned}$$

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 [23], the multiplicative iterative strategies [24], and the FCLS technique [9]. More recently, the ADMM was applied with success for hyperspectral unmixing problem, with the SUnSAL algorithm [12].

Recent work in hyperspectral unmixing has advocated the sparsity of the abundance vectors [12], [25], [26]. In this case, each spectrum is fitted by a sparse linear mixture of endmembers, namely, only the abundances with respect to a small number of endmembers are nonzero. To this end, the sparsity-promoting regularization with the ℓ_1 -norm is included in the cost function, yielding the following constrained sparse regression problem [12], for each $t = 1, \dots, T$:

$$\min_{\mathbf{x}_t} \|\mathbf{y}_t - \mathbf{M}\mathbf{x}_t\|_2^2 + \lambda \|\mathbf{x}_t\|_1, \quad \text{s.t. } \mathbf{x}_t \geq 0$$

where the parameter λ balances the fitness of the least-squares solution and the sparsity level. It is worth noting that the ASC is relaxed when the ℓ_1 -norm is included. This problem is often considered using the following matrix formulation:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_F^2 + \lambda \sum_{t=1}^T \|\mathbf{x}_t\|_1, \quad \text{s.t. } \mathbf{X} \geq 0.$$

A. 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

$$\min_{\mathbf{X}} \sum_{l=1}^L \|\mathbf{y}_{l*} - (\mathbf{M}\mathbf{X})_{l*}\|_2^2 \quad (5)$$

subject to any of the aforementioned constraints. 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 $(\mathbf{M}\mathbf{X})_{l*}$ are far from actual observations \mathbf{y}_{l*} . Moreover, it is common for hyperspectral images to present up to 20% of unusable spectral bands due to low SNR essentially from atmospheric effects, such as water absorption. In the following section, we 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. CORRENTROPY-BASED UNMIXING PROBLEMS

In this section, we examine the correntropy and write the unmixing problems as correntropy maximization ones. Algorithms for solving these problems are derived in Section IV.

A. Correntropy

The correntropy, studied in [17] and [18], is a nonlinear local similarity measure. For two random variables, \mathcal{Y} and its

TABLE I
COMMONLY USED SHIFT-INVARIANT KERNELS

Kernel	$\kappa(u, v)$
Gaussian	$\exp\left(\frac{-1}{2\sigma^2} \ u - v\ ^2\right)$
Laplacian	$\exp\left(\frac{-1}{2\sigma^2} \ u - v\ \right)$
Inverse multiquadratic	$\frac{1}{\sqrt{\ u - v\ ^2 + c^2}}$

estimation $\hat{\mathcal{Y}}$ using some model/algorithm, it is defined by

$$\mathbb{E}[\kappa(\mathcal{Y}, \hat{\mathcal{Y}})] \quad (6)$$

where $\mathbb{E}[\cdot]$ is the expectation operator, and $\kappa(\cdot, \cdot)$ is a shift-invariant kernel satisfying the Mercer theorem [27]. In practice, while the joint distribution function of \mathcal{Y} and $\hat{\mathcal{Y}}$ is unavailable, the sample estimator of correntropy is adopted instead. Employing a finite number of data $\{(\mathbf{y}_{l*}, \hat{\mathbf{y}}_{l*})\}_{l=1}^L$, it is estimated by

$$\frac{1}{L} \sum_{l=1}^L \kappa(\mathbf{y}_{l*}, \hat{\mathbf{y}}_{l*}) \quad (7)$$

up to a normalization factor. Expressions of the most commonly-used shift-invariant Mercer kernels are presented in Table I. In the following, we restrict the presentation to the Gaussian kernel, since it is the mostly used one for correntropy [17], [20], [28]. It is worth noting that the proposed algorithms can be easily extended to other kernels. Replacing the expression of Gaussian kernel in (7) leads to the following expression for the correntropy:

$$\frac{1}{L} \sum_{l=1}^L \exp\left(\frac{-1}{2\sigma^2} \|\mathbf{y}_{l*} - \hat{\mathbf{y}}_{l*}\|_2^2\right) \quad (8)$$

where σ denotes the bandwidth of the Gaussian kernel.

The maximization of the correntropy, given by

$$\max_{\hat{\mathbf{y}}_{1*}, \dots, \hat{\mathbf{y}}_{L*}} \frac{1}{L} \sum_{l=1}^L \kappa(\mathbf{y}_{l*}, \hat{\mathbf{y}}_{l*})$$

is termed the maximum correntropy criterion [17]. It is noteworthy that well-known second-order statistics, such as the mean square error (MSE), depends heavily on the Gaussian and linear assumptions [17]. However, in the 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 [29]. By contrast, the maximization of the correntropy criterion is appropriate for non-Gaussian signal processing, and is robust in particular against large outliers, as shown next.

B. Underlying Robustness of the Correntropy Criterion

In this section, we study the sensitivity to outliers of the correntropy maximization principle, by showing the robustness of the underlying mechanism. To this end, we examine the behavior of the correntropy in terms of the residual error defined by $\epsilon_l = \|\mathbf{y}_{l*} - \hat{\mathbf{y}}_{l*}\|_2$. Thus, the correntropy (8) becomes

$$\frac{1}{L} \sum_{l=1}^L \exp\left(\frac{-1}{2\sigma^2} \epsilon_l^2\right).$$

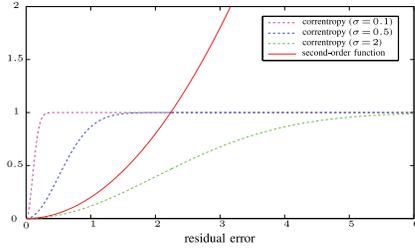


Fig. 1. Illustration of the second-order objective function (ϵ_l^2 , in red solid line) and the negative correntropy objective function $[1 - \exp(-1/(2\sigma^2))\epsilon_l^2]$, in dashed lines for several values of σ , in terms of the residual error (ϵ_l).

Compared with 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 increases, the second-order function keeps increasing dramatically. On the contrary, the correntropy is sensitive only within a region of small residual errors, this region being controlled by the kernel bandwidth. For large magnitudes of residual error, the correntropy falls to zero. Consequently, the correntropy criterion is robust to large outliers.

C. Correntropy-Based Unmixing Problems

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

$$\mathcal{C}(\mathbf{X}) = - \sum_{l=1}^L \exp \left(\frac{-1}{2\sigma^2} \|\mathbf{y}_{l*} - (\mathbf{M}\mathbf{X})_{l*}\|_2^2 \right) \quad (9)$$

where the Gaussian kernel was considered, or equivalently

$$\mathcal{C}(\mathbf{X}) = - \sum_{l=1}^L \exp \left(\frac{-1}{2\sigma^2} \sum_{t=1}^T \left(y_{lt} - \sum_{r=1}^R x_{rt} m_{lr} \right)^2 \right). \quad (10)$$

Considering both the ANC and ASC constraints, the fully constrained correntropy unmixing problem becomes

$$\begin{aligned} \min_{\mathbf{X}} \mathcal{C}(\mathbf{X}) \\ \text{s.t. } \mathbf{X} \geq 0 \text{ and } \mathbf{1}^\top \mathbf{x}_t = 1, \quad \text{for } t = 1, \dots, T. \end{aligned} \quad (11)$$

To promote the sparsity of the abundances, the objective function (9), (10) can be augmented by the ℓ_1 -norm penalty on the abundance matrix \mathbf{X} , leading to the following problem:

$$\min_{\mathbf{X}} \mathcal{C}(\mathbf{X}) + \lambda \sum_{t=1}^T \|\mathbf{x}_t\|_1, \quad \text{s.t. } \mathbf{X} \geq 0. \quad (12)$$

IV. ADMM FOR SOLVING THE CORRENTROPY-BASED UNMIXING PROBLEMS

We first briefly review the ADMM, following the expressions in [13, Ch. 3]. Consider an optimization problem of the form

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x})$$

where the functions f and g are closed, proper, and convex. The ADMM solves the equivalent constrained problem

$$\min_{\mathbf{x}, \mathbf{z}} f(\mathbf{x}) + g(\mathbf{z}) \text{ s.t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \quad (13)$$

such as having the particular constraint $\mathbf{x} = \mathbf{z}$ for instance. While this formulation may seem trivial, the optimization problem can now be tackled using the augmented Lagrangian method where the objective function is separable in \mathbf{x} and \mathbf{z} . By alternating on each variable separately, the ADMM repeats a direct update of the dual variable. In its scaled form, the ADMM algorithm is summarized in Algorithm 1. Assuming that the (unaugmented, namely, $\rho = 0$ in Algorithm 1) Lagrangian associated with problem (13) has a saddle point, the ADMM iterates, as given in Algorithm 1, satisfy the following: 1) the objective function convergence; 2) the primal residual convergence, i.e., $\|\mathbf{x}_{k+1} - \mathbf{z}_{k+1}\|_2 \rightarrow 0$; and 3) the dual residual convergence, i.e., $\rho \|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2 \rightarrow 0$ (see [13] for more details).

Algorithm 1 ADMM Algorithm [13]

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

- 1: Initialize $k = 0$, \mathbf{x}_0 , \mathbf{z}_0 , and \mathbf{u}_0
 - 2: **repeat**
 - 3: $\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}_k - \mathbf{c} + \mathbf{u}_k\|_2^2$;
 - 4: $\mathbf{z}_{k+1} = \arg \min_{\mathbf{z}} g(\mathbf{z}) + \frac{\rho}{2} \|\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z} - \mathbf{c} + \mathbf{u}_k\|_2^2$;
 - 5: $\mathbf{u}_{k+1} = \mathbf{u}_k + \mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}$;
 - 6: $k = k + 1$;
 - 7: **until** stopping criterion
-

A. Correntropy-Based Unmixing With Full Constraints

In the following, we apply the ADMM algorithm to solve the correntropy-based unmixing problem in the fully constrained case, presented in (11). The main steps are summarized in Algorithm 2. Rewrite the variables to be optimized in a vector $\mathbf{x} \in \mathbb{R}^{RT \times 1}$, which is stacked by the columns of the matrix \mathbf{X} , namely, $\mathbf{x} = [\mathbf{x}_1^\top \cdots \mathbf{x}_T^\top]^\top$. Also rewrite the following vectors in $\mathbb{R}^{RT \times 1}$: $\mathbf{z} = [\mathbf{z}_1^\top \cdots \mathbf{z}_T^\top]^\top$ and $\mathbf{u} = [\mathbf{u}_1^\top \cdots \mathbf{u}_T^\top]^\top$, where, for $t = 1, \dots, T$, $\mathbf{z}_t = [z_{1t} \cdots z_{Rt}]^\top$ and $\mathbf{u}_t = [u_{1t} \cdots u_{Rt}]^\top$. By following the formulation of the ADMM in Algorithm 1, we set:

$$\begin{aligned} f(\mathbf{x}) &= \mathcal{C}(\mathbf{x}) + \sum_{t=1}^T \iota_{\{1\}}(\mathbf{1}^\top \mathbf{x}_t) \\ g(\mathbf{z}) &= \iota_{\mathbb{R}_+^{RT}}(\mathbf{z}) \\ \mathbf{A} &= -\mathbf{I}, \quad \mathbf{B} = \mathbf{I} \text{ and } \mathbf{c} = \mathbf{0} \end{aligned} \quad (14)$$

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

$$\iota_{\mathcal{S}}(u) = \begin{cases} 0 & \text{if } u \in \mathcal{S} \\ \infty & \text{otherwise.} \end{cases}$$

In this case, the subproblem of the \mathbf{x} -update (in line 3 of Algorithm 1) addresses a nonconvex problem without any 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 $t = 1, \dots, T$. Let $\bar{\mathbf{x}} \in \mathbb{R}^{(R-1)T \times 1}$ be the reduced vector of $(R-1)$ unknowns to be estimated, stacked by

$$\bar{\mathbf{x}}_t = [x_{1t} \ \dots \ x_{(R-1)t}]^\top$$

for $t = 1, \dots, T$. By this means, the objective function in (14) is transformed from (10) into the reduced form

$$f_1(\bar{\mathbf{x}}) = - \sum_{l=1}^L \exp\left(\frac{-1}{2\sigma^2} \sum_{t=1}^T \epsilon_l(\bar{\mathbf{x}}_t)^2\right) \quad (15)$$

where $\epsilon_l(\bar{\mathbf{x}}_t) = y_{lt} - m_{lR} - \sum_{p=1}^{R-1} (m_{lp} - m_{lR})x_{pt}$, for $l = 1, \dots, L$. The gradient of (15) with respect to $\bar{\mathbf{x}}$ is stacked as

$$\frac{\partial f_1}{\partial \bar{\mathbf{x}}} = \left[\frac{\partial f_1}{\partial \bar{\mathbf{x}}_1} \ \dots \ \frac{\partial f_1}{\partial \bar{\mathbf{x}}_T} \right]^\top \in \mathbb{R}^{(R-1)T \times 1}$$

where $(\partial f_1 / \partial \bar{\mathbf{x}}_t) = [(\partial f_1 / \partial \bar{x}_{1t}) \ \dots \ (\partial f_1 / \partial \bar{x}_{(R-1)t})]^\top$, with the entries given by

$$\frac{\partial f_1(\bar{\mathbf{x}})}{\partial \bar{x}_{rt}} = \frac{1}{\sigma^2} \sum_{l=1}^L (m_{lR} - m_{lr}) \exp\left(\frac{-1}{2\sigma^2} \sum_{s=1}^T \epsilon_l(\bar{\mathbf{x}}_s)^2\right) \epsilon_l(\bar{\mathbf{x}}_t)$$

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

$$\begin{aligned} \phi(\bar{\mathbf{x}}) = & \frac{\rho}{2} \sum_{t=1}^T \left(1 - \sum_{p=1}^{R-1} x_{pt} - z_{Rt,k} - u_{Rt,k} \right)^2 \\ & + \frac{\rho}{2} \sum_{p=1}^{R-1} (x_{pt} - z_{pt,k} - u_{pt,k})^2 \end{aligned}$$

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

$$\frac{\partial \phi(\bar{\mathbf{x}})}{\partial \bar{x}_{rt}} = \rho \left(x_{rt} + \sum_{p=1}^{R-1} x_{pt} - 1 + z_{Rt,k} - z_{rt,k} + u_{Rt,k} - u_{rt,k} \right) \quad (16)$$

for all $r = 1, \dots, R-1$ and $t = 1, \dots, T$.

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

B. Sparsity-Promoting Unmixing Algorithm

In order to apply the ADMM algorithm, we express the constrained optimization problem (12) as follows:

$$\begin{aligned} f(\mathbf{x}) &= \mathcal{C}(\mathbf{x}) \\ g(\mathbf{z}) &= \iota_{\mathbb{R}_+^{\text{RT}}}(\mathbf{z}) + \lambda \|\mathbf{z}\|_1 \\ \mathbf{A} &= -\mathbf{I}, \quad \mathbf{B} = \mathbf{I} \text{ and } \mathbf{c} = \mathbf{0}. \end{aligned} \quad (17)$$

Algorithm 2 Correntropy-Based Unmixing With Full Constraints (CUSAL-FC)

-
- 1: Initialize $k = 0$, $\rho > 0$, $\eta > 0$, $\sigma > 0$; \mathbf{x}_0 , \mathbf{z}_0 , and \mathbf{u}_0 ;
 - 2: **repeat**
 - 3: **repeat**
 - 4: $\bar{\mathbf{x}}_{k+1} = \bar{\mathbf{x}}_{k+1} - \eta \left(\frac{\partial f_1}{\partial \bar{\mathbf{x}}_{k+1}} + \frac{\partial \phi}{\partial \bar{\mathbf{x}}_{k+1}} \right)$;
 - 5: **until** convergence
 - 6: reform \mathbf{x}_{k+1} using $\bar{\mathbf{x}}_{k+1}$;
 - 7: $\mathbf{z}_{k+1} = \max(\mathbf{0}, \mathbf{x}_{k+1} - \mathbf{u}_k)$;
 - 8: $\mathbf{u}_{k+1} = \mathbf{u}_k - (\mathbf{x}_{k+1} - \mathbf{z}_{k+1})$;
 - 9: $k = k + 1$;
 - 10: **until** stopping criterion
-

By analogy with the previous case, the \mathbf{x} -update in line 3 of Algorithm 1 is solved iteratively with the gradient descent method and is given in lines 3–5 of Algorithm 3. The gradient of (17) with respect to \mathbf{x} is stacked by $(\partial f / \partial \mathbf{x}_t)$, where

$$\frac{\partial f}{\partial \mathbf{x}_t} = -\frac{1}{\sigma^2} \sum_{l=1}^L \epsilon_l(\mathbf{x}_t) \exp\left(\frac{-1}{2\sigma^2} \sum_{s=1}^T (\epsilon_l(\mathbf{x}_s))^2\right) \mathbf{m}_{l*}^\top$$

for $t = 1, \dots, T$, where $\epsilon_l(\mathbf{x}_t) = y_{lt} - \sum_{r=1}^R x_{rt} m_{lr}$. The \mathbf{z} -update in line 4 Algorithm 1 involves solving

$$\mathbf{z}_{k+1} = \arg \min_{\mathbf{z}} \iota_{\mathbb{R}_+^{\text{RT}}}(\mathbf{z}) + (\lambda/\rho) \|\mathbf{z}\|_1 + \frac{1}{2} \|\mathbf{z} - \mathbf{x}_{k+1} - \mathbf{u}_k\|_2^2. \quad (18)$$

In [13], the ADMM has been applied to solve various ℓ_1 -norm problems, including the well-known LASSO [30]. The only difference between (18) and the \mathbf{z} -update in LASSO is that in the latter, no nonnegativity term $\iota_{\mathbb{R}_+^{\text{RT}}}(\mathbf{z})$ is enforced. In this case, the \mathbf{z} -update in LASSO is the element-wise soft thresholding operation

$$\mathbf{z}_{k+1} = S_{\lambda/\rho}(\mathbf{x}_{k+1} - \mathbf{u}_k)$$

where the soft thresholding operator [13] is defined by

$$S_b(\zeta) = \begin{cases} \zeta - b & \text{if } \zeta > b \\ 0 & \text{if } \|\zeta\| < b \\ \zeta + b & \text{if } \zeta < -b. \end{cases}$$

Following [12], it is straightforward to project the result onto the nonnegative orthant in order to include the nonnegativity constraint, thus yielding:

$$\mathbf{z}_{k+1} = \max(\mathbf{0}, S_{\lambda/\rho}(\mathbf{x}_{k+1} - \mathbf{u}_k))$$

where the maximum function is element-wise. All these results lead to the correntropy-based unmixing algorithm with sparsity promoting, as summarized in Algorithm 3.

C. On the Bandwidth Determination and Convergence

We apply a threefold stopping criterion for Algorithms 2 and 3, according to [12] and [13].

- 1) The primal and dual residuals are small enough, namely, $\|\mathbf{x}_{k+1} - \mathbf{z}_{k+1}\|_2 \leq \epsilon_1$ and $\rho \|\mathbf{z}_{k+1} - \mathbf{z}_k\|_2 \leq \epsilon_2$, where $\epsilon_1 = \epsilon_2 = \sqrt{\text{RT}} \times 10^{-5}$ as in [12].

Algorithm 3 Correntropy-Based Unmixing With Sparsity Promoting (CUSAL-SP)

```

1: Initialize  $k = 0$ ,  $\rho > 0$ ,  $\sigma > 0$ ,  $\eta > 0$ ,  $\lambda > 0$ ;  $\mathbf{x}_0$ ,  $\mathbf{z}_0$ , and  $\mathbf{u}_0$ ;
2: repeat
3:   repeat
4:      $\mathbf{x}_{k+1} = \mathbf{x}_{k+1} - \eta \left( \frac{\partial f}{\partial \mathbf{x}_{k+1}} + \rho(\mathbf{x}_{k+1} - \mathbf{z}_k - \mathbf{u}_k) \right)$ ;
5:   until convergence
6:    $\mathbf{z}_{k+1} = \max(\mathbf{0}, S_{\lambda/\rho}(\mathbf{x}_{k+1} - \mathbf{u}_k))$ ;
7:    $\mathbf{u}_{k+1} = \mathbf{u}_k - (\mathbf{x}_{k+1} - \mathbf{z}_{k+1})$ ;
8:    $k = k + 1$ ;
9: until stopping criterion
  
```

- 2) The primal residual starts to increase, i.e., $\|\mathbf{x}_{k+1} - \mathbf{z}_{k+1}\|_2 > \|\mathbf{x}_k - \mathbf{z}_k\|_2$.
- 3) The maximum iteration number is attained.

The bandwidth σ in the Gaussian kernel should be well tuned. Note that a small value for this parameter punishes harder the outlier bands, thus increasing the robustness of the algorithm to outliers [20]. Note that, in this paper, the ADMM is applied to address a nonconvex objective function, and thus no convergence is guaranteed theoretically, according to [13]. Considering these issues, we propose to fix the bandwidth empirically as summarized in Algorithm 4 and described next.

Algorithm 4 Tuning the Bandwidth Parameter σ

```

1: Initialize  $\sigma = \sigma_0$  using (19);  $p = 1$ ;
2: Do CUSAL with Algorithm 2 or 3;
3: if stopping criterion 1) or 3) is satisfied then
4:   if condition  $\frac{\|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_F}{\|\mathbf{Y} - \mathbf{M}\mathbf{X}_{LS}\|_F} < 2$  is satisfied, then
5:      $\sigma^* = \sigma$  (optimal value)
6:   else
7:     increase  $\sigma = 1.2\sigma$ , and go to line 2
8:   end if
9: else
10:  if  $\sigma > 1000\sigma_0$  (due to the overestimated  $\sigma_0$ ) then
11:     $p = p + 1$ ;
12:    decrease  $\sigma = \sigma_0/p$ , and go to line 2
13:  else
14:    increase  $\sigma = 1.2\sigma$ , and go to line 2
15:  end if
16: end if
  
```

Following [20] and [21], we first initialize the bandwidth parameter as a function of the reconstruction error, given by

$$\sigma_0^2 = \frac{\theta}{\text{LT}} \|\mathbf{Y} - \mathbf{M}\mathbf{X}_{LS}\|_F^2 \quad (19)$$

where the parameter θ is chosen as $\theta = (\text{RT}/2)$ in this paper and \mathbf{X}_{LS} is the least-squares solution (4). In the case of a result too apart from that of least-squares solution, the parameter is augmented by $\sigma = 1.2\sigma$, until that the condition $(\|\mathbf{Y} - \mathbf{M}\mathbf{X}\|_F) / (\|\mathbf{Y} - \mathbf{M}\mathbf{X}_{LS}\|_F) < 2$ is satisfied. The algorithm divergence occurs if the stopping criterion 2) is satisfied, namely, the primal residual increases over iterations. In this case, either the parameter is too large due to an

overestimated initialization or it is too small. Accordingly, we either decrease it by $\sigma = \sigma_0/p$ or increase it by $\sigma = 1.2\sigma$, until that the ADMM converges.

The nonconvexity of the problem requires a loop for tuning outside the ADMM. This leads to an algorithm more time consuming than the half-quadratic methods that transform the original problem to a reweighted NMF [19], [20]. However, for half-quadratic with the expectation conditional maximization method (see details in [19]), although the parameter is updated over iterations, it is around the initialization value that is estimated from a formula similar to (19). When dealing with high noise levels, this value is often not small enough to account for the influence brought by outliers. The parameter tuning step in ADMM alleviates the parameter estimation problem to some degree. This phenomenon is observed in the experiments in the next sections.

Although the convergence is not proved theoretically, we will provide evidence in the experiments that the proposed ADMM algorithm converges to stationary points and that the results have robustness with respect to the initializations.

V. EXPERIMENTS WITH SYNTHETIC DATA

In this section, the performance of the proposed fully constrained (CUSAL-FC) and sparsity-promoting (CUSAL-SP) algorithms is evaluated on synthetic data. A comparative study is performed considering six state-of-the-art methods proposed for linear and nonlinear unmixing models.

- 1) *FCLS* [9]: 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.
- 2) *SUnSAL* [12]: This method is based on the ADMM. Several variants are developed by including different constraints, with the fully constrained SUnSAL-FCLS and the sparsity-promoting SUnSAL-SP.
- 3) *The Bayesian Algorithm for Generalized Bilinear Model (BayGBM)* [31], [32]: This method estimates the abundances with the generalized bilinear model (GBM), which adds second-order interactions between endmembers to the linear model, yielding the model

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

where $0 \leq \gamma_{ij,t} \leq 1$ controls the interactions between endmembers \mathbf{m}_i and \mathbf{m}_j , and \odot is the element-wise product. The BayGBM considers both ANC and ASC.

- 4) *KFCLS* [15]: This method generalizes FCLS, by replacing the inner product with a kernel function. In the following, the Gaussian kernel is applied for simulation.
- 5) *NMF With a Maximum Correntropy Criterion (NMF-MCC)* [19]: This NMF-based method maximizes the correntropy between the input matrix and the product of the two matrices to estimate in an unsupervised manner. The resulting optimization problem is solved using the half-quadratic technique and the expectation conditional maximization method.

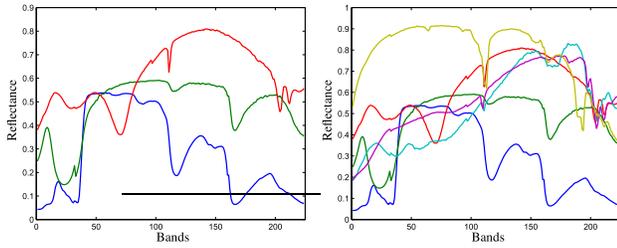


Fig. 2. (Left) $R = 3$ and (right) 6 USGS signatures chosen for simulation.

TABLE II

COMPARISON OF RMSE ($\times 10^{-2}$) WITH DIFFERENT $\overline{\text{SNR}}$, FOR $R = 3$

	$\overline{\text{SNR}} = 10$	$\overline{\text{SNR}} = 20$	$\overline{\text{SNR}} = 30$	$\overline{\text{SNR}} = 40$	$\overline{\text{SNR}} = 50$
FCLS	10.18	② 3.86	② 1.20	② 0.41	① 0.12
SUnSAL-FCLS	10.17	② 3.86	② 1.20	② 0.41	① 0.12
BayGBM	10.23	3.88	② 1.20	0.42	0.16
KFCLS	12.64	4.43	1.42	0.55	② 0.13
NMF-MCC	13.73	5.53	1.95	① 0.38	0.15
ℓ_1 -CENMF	② 9.91	4.42	1.28	0.42	0.15
CUSAL-FC	① 7.92	① 3.03	① 1.15	② 0.41	① 0.12

- 6) *Correntropy-Based NMF Promoted by ℓ_1 -Norm (ℓ_1 -CENMF)* [21]: This method improves the NMF-MCC by including the ℓ_1 -norm of the second unknown matrix to the objective function, for the sake of a sparse representation. Similar to the NMF-MCC, the half-quadratic strategy is applied to solve the optimization problem.
- 7) For fair comparison, the supervised cases of NMF-MCC and ℓ_1 -CENMF are considered in the experiments, by setting the endmember matrices as the actual ones.

A. Performance of CUSAL-FC (Fully Constrained Algorithm)

We first compare the fully constrained CUSAL-FC, presented in Section IV-A, with the state-of-the-art methods. Two sets of experiments are performed, mainly considering the influence of the noise level and the number of endmembers.

Each image, of 50×50 pixels, is generated using the linear mixing model (1). The Gaussian noise is added as in [21] where the noise levels, represented by SNR, vary across the bands to simulate the cases of real hyperspectral images. The $R \in \{3, 6\}$ endmembers, as shown in Fig. 2, are drawn from the USGS digital spectral library [33]. These endmembers are defined over $L = 244$ continuous bands with the wavelength ranging from 0.2 to $3.0 \mu\text{m}$. The abundance vectors \mathbf{x}_t are uniformly generated using a Dirichlet distribution as in [33] and [34].

The unmixing performance is evaluated using the abundance root MSE (RMSE) [31], [35], defined by

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

where $\hat{\mathbf{x}}_t$ is the estimated abundance vector.

In the first set of experiments, the SNRs of bands are generated using the normal distribution $\text{SNR} \sim \mathcal{N}(\overline{\text{SNR}}, \epsilon^2)$, with $\overline{\text{SNR}} \in \{10, 20, 30, 40, 50\}$ and $\epsilon = 5$ according to [21]. Tables II and III illustrate the average of RMSE

TABLE III

COMPARISON OF RMSE ($\times 10^{-2}$) WITH DIFFERENT $\overline{\text{SNR}}$, FOR $R = 6$

	$\overline{\text{SNR}} = 10$	$\overline{\text{SNR}} = 20$	$\overline{\text{SNR}} = 30$	$\overline{\text{SNR}} = 40$	$\overline{\text{SNR}} = 50$
FCLS	② 9.04	5.14	2.05	① 0.70	① 0.24
SUnSAL-FCLS	② 9.04	5.14	2.05	① 0.70	① 0.24
BayGBM	9.05	5.14	2.04	① 0.70	① 0.24
KFCLS	9.53	6.06	2.20	0.76	② 0.25
NMF-MCC	9.49	5.28	2.08	② 0.74	0.33
ℓ_1 -CENMF	9.45	② 5.06	① 1.98	② 0.74	0.32
CUSAL-FC	① 7.87	① 4.63	② 2.02	① 0.70	① 0.24

TABLE IV

COMPARISON OF RMSE ($\times 10^{-2}$) WITH $\overline{\text{SNR}}_1 = 30$ AND DIFFERENT $\overline{\text{SNR}}_2$, FOR $R = 3$

	$\overline{\text{SNR}}_2 = 5$	$\overline{\text{SNR}}_2 = 10$	$\overline{\text{SNR}}_2 = 15$
FCLS	7.66	4.86	2.99
SUnSAL-FCLS	7.66	4.85	2.99
BayGBM	7.70	4.90	3.00
KFCLS	10.45	5.85	4.45
NMF-MCC	7.50	4.82	2.91
ℓ_1 -CENMF	② 5.33	② 3.08	② 2.17
CUSAL-FC	① 1.75	① 1.66	① 1.73

over 10 Monte Carlo realizations, respectively, for $R = 3$ and $R = 6$. It is easy to see that, when the average noise level is relatively high with $\overline{\text{SNR}} = 10$ and 20, the proposed CUSAL-FC algorithm outperforms all the comparing methods in terms of RMSE, for different numbers of endmembers. When the average noise level is relatively low, namely, $\overline{\text{SNR}} = 40$ and 50, the proposed CUSAL-FC is able to provide comparable results as the least squares approaches, e.g., FCLS and SUnSAL-FCLS. It is also shown that the performance of the proposed algorithm improves when increasing the SNR.

The second set of experiments is conducted in order to examine the performance of the proposed method in presence of highly noisy bands, which is a common phenomenon for real hyperspectral images. To this end, the data are similarly generated as previously described, where two normal distributions $\text{SNR} \sim \mathcal{N}(\overline{\text{SNR}}_1, \epsilon^2)$ and $\text{SNR} \sim \mathcal{N}(\overline{\text{SNR}}_2, \epsilon^2)$ are used, with $\epsilon = 5$. While most bands have a common average noisy level with $\overline{\text{SNR}}_1 = 30$, there are 40 out of 224 bands randomly chosen to be severely corrupted by high-level average noise with $\overline{\text{SNR}}_2 \in \{5, 10, 15\}$. Tables IV and V report the average of RMSE over 10 Monte Carlo realizations, respectively, for $R = 3$ and $R = 6$. We observe that the proposed CUSAL-FC algorithm is the most effective among all the comparing methods when the data contain highly noisy bands, regardless of number of endmembers.

We investigate the convergence property of the proposed ADMM algorithm with CUSAL-FC, and examine its robustness with respect to the initialization. To this end, a toy image of 50×50 pixels is generated using the linear mixing model (1), where the Gaussian noise with $\text{SNR} = 30$ is added. The $R = 3$ endmembers, as shown in Fig. 2 (left), is considered while the abundance vectors \mathbf{x}_t are uniformly generated using a Dirichlet distribution. The changes of the objective function value and the primal and dual residuals

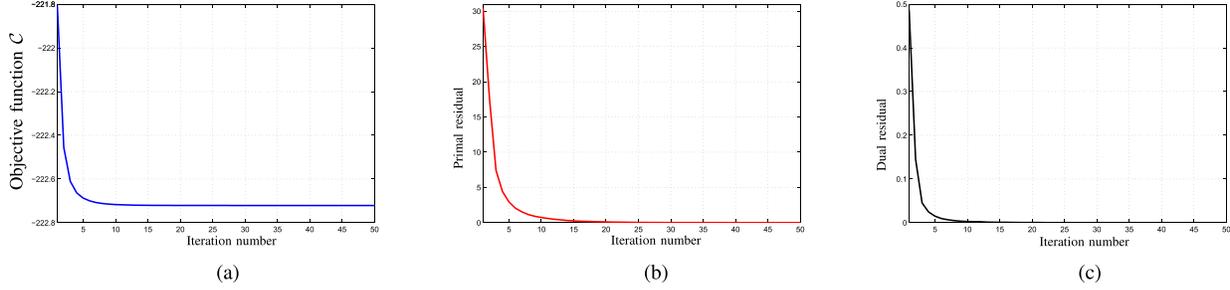


Fig. 3. Illustration of the convergence property of CUSAL-FC, by plotting the objective function value and the primal and dual residuals over the first 50 iterations, using a fixed value of $\mu = 0.01$. (a) Objective function value. (b) Primal residual. (c) Dual residual.

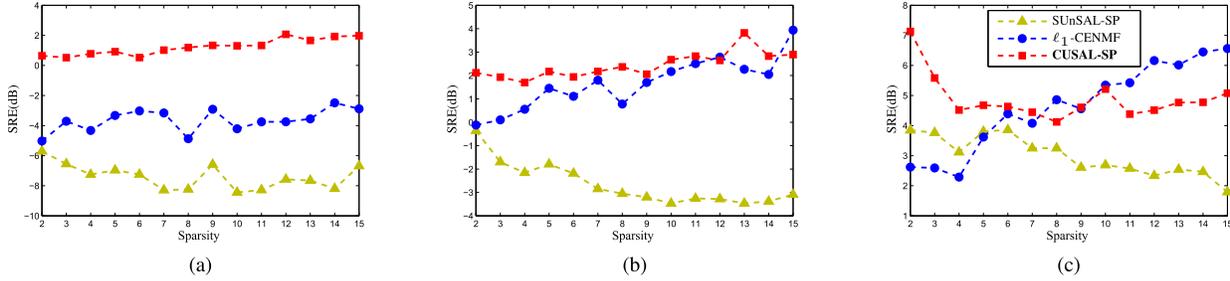


Fig. 4. Averaged SRE with respect to the sparsity level K , averaged over 10 Monte Carlo realizations. Comparison for various average noise level SNR. (a) $\overline{\text{SNR}} = 10$. (b) $\overline{\text{SNR}} = 20$. (c) $\overline{\text{SNR}} = 30$.

TABLE V
COMPARISON OF RMSE ($\times 10^{-2}$) WITH $\overline{\text{SNR}}_1 = 30$
AND DIFFERENT $\overline{\text{SNR}}_2$, FOR $R = 6$

	$\overline{\text{SNR}}_2 = 5$	$\overline{\text{SNR}}_2 = 10$	$\overline{\text{SNR}}_2 = 15$
FCLS	8.00	6.27	4.37
SUNSAL-FCLS	8.00	6.27	4.37
BayGBM	8.02	6.29	4.38
KFCLS	8.56	7.81	4.44
NMF-MCC	8.08	6.22	4.31
ℓ_1 -CENMF	② 7.15	② 5.21	② 3.63
CUSAL-FC	① 3.98	① 3.73	① 3.35

TABLE VI
AVERAGES AND DEVIATIONS OF RMSE ($\times 10^{-2}$) WITH
RESPECT TO DIFFERENT INITIALIZATIONS

	CUSAL-FC	NMF-MCC
$x_{rt} \sim \mathcal{U}(0, 1)$	$1.16 \pm 2.5 \times 10^{-2}$	$1.27 \pm 3.5 \times 10^{-2}$
$x_{rt} \sim \mathcal{N}(0.5, 0.1)$	$0.09 \pm 0.2 \times 10^{-2}$	$0.11 \pm 0.1 \times 10^{-2}$
$x_{rt} \sim \mathcal{N}(0.5, 0.2)$	$0.09 \pm 0.5 \times 10^{-2}$	$4.70 \pm 4.0 \times 10^{-2}$

over the first 50 iterations are given in Fig. 3. A rapid drop of objective function and residuals for the first 20 iterations is observed, signifying that the proposed ADMM algorithm converges to stationary points with modest accuracy. We examine the robustness of CUSAL-FC with respect to the initialization, and compare it with the half-quadratic approach proposed for maximum correntropy criterion, namely, NMF-MCC [19]. The elements in X are identically initialized for each method, using: 1) uniform distribution $x_{rt} \sim \mathcal{U}(0, 1)$; 2) normal distribution $x_{rt} \sim \mathcal{N}(0.5, 0.1)$ combined with rounding all the negative values up to zero; and 3) normal distribution $x_{rt} \sim \mathcal{N}(0.5, 0.2)$ combined with rounding all the negative values up to zero. Ten Monte Carlo realizations are performed, leading to the averages and deviations of RMSE given in Table VI. In contrast to NMF-MCC, these results show that the proposed ADMM algorithm provides good results even when considering different initializations.¹

¹It is not reasonable to directly compare the objective function values of CUSAL-FC and NMF-MCC, since these two methods address distinct optimization problems involving different values of parameter σ and different constraints.

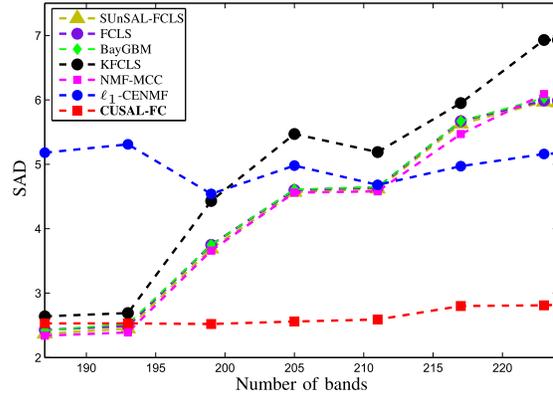


Fig. 5. *Cuprite* image: the averaged SAD using a different number of bands, computed without the noisy bands 1–3, 105–115, 150–170, and 223–224.

B. Performance of CUSAL-SP (Sparsity-Promoting Algorithm)

The performance of the proposed sparsity-promoting CUSAL-SP, presented in IV-B, is compared with two sparsity-promoting methods: SUNSAL-SP and ℓ_1 -CENMF, on a series of data with sparse abundance matrices. We study the influence of: 1) the noise level over bands, namely, $\overline{\text{SNR}}$ and 2) the sparsity level of the abundances. Each image, of 15×15 pixels, is generated by the LMM. The endmember matrix is composed by $R = 62$ USGS signatures, where the angle between

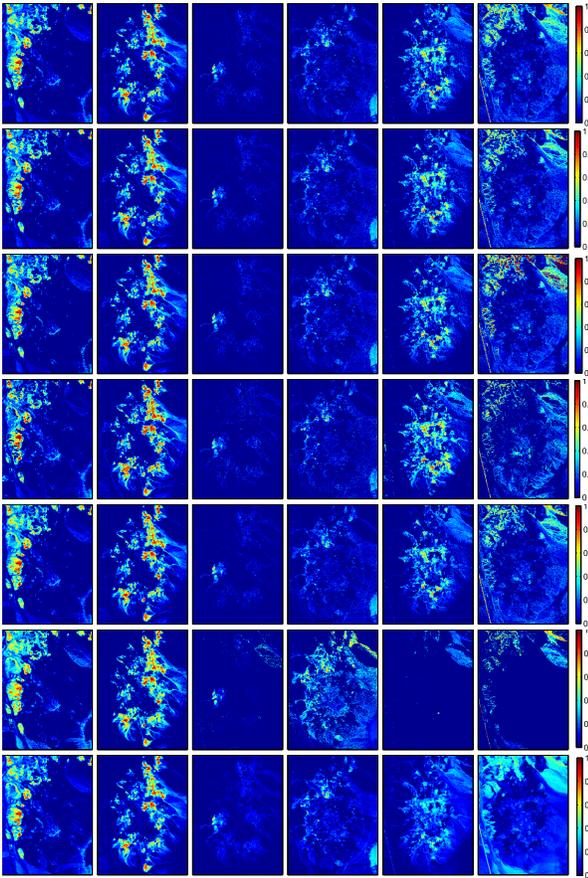


Fig. 6. Cuprite image. Estimated abundance maps using 187 clean bands. (Left to right) Sphene, alunite, buddingtonite, kaolinite, chalcedony, and highway. (Top to bottom) SUnSAL-FCLS, FCLS, BayGBM, KFCLS, NMF-MCC, ℓ_1 -CENMF, and CUSAL-FC.

any two different endmembers is larger than 10° [25]. The K nonzero entries in each abundance vector \mathbf{x}_t are generated by a Dirichlet distribution. The value of K (i.e., the indicator of sparsity level) ranges from 2 to 15, while the average level of noise $\overline{\text{SNR}} \in \{10, 20, 30\}$. A rough measure of the sparsity level of the unknown abundance matrix from the input spectra [36] takes the form

$$\hat{s} = \frac{1}{\sqrt{L}} \sum_{l=1}^L \frac{\sqrt{T} - \|\mathbf{y}_{l*}\|_1 / \|\mathbf{y}_{l*}\|_2}{\sqrt{T} - 1}.$$

For all the algorithms, the sparsity-promoting parameter λ is adjusted using the set $\hat{s} \times \{10^{-5}, 5 \cdot 10^{-5}, 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}\}$.

The unmixing performance with the sparsity-promoting algorithms is evaluated using the signal-to-reconstruction error (SRE), measured in decibels, according to [12] and [25]. It is defined by

$$\text{SRE} = 10 \log_{10} \left(\frac{\sum_{t=1}^T \|\mathbf{x}_t\|_2^2}{\sum_{t=1}^T \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2} \right).$$

The results, averaged over 10 Monte Carlo realizations, are illustrated in Fig. 4. Considering that the abundance matrix under estimation is sparse at different levels, we conclude the following: CUSAL-SP always outperforms SUnSAL-SP. When dealing with high noise levels, namely, $\overline{\text{SNR}} = 10$ and 20, CUSAL-SP outperforms ℓ_1 -CENMF in

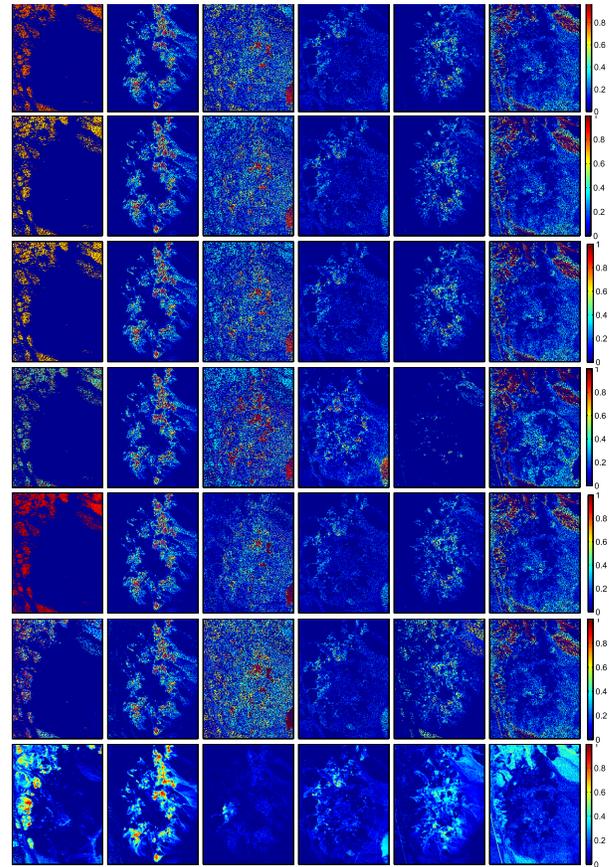


Fig. 7. Cuprite image. Estimated abundance maps using 205 bands, with 187 clean bands. The same legend as Fig. 6.

most cases (except for $K = 12$ and 15 with $\overline{\text{SNR}} = 20$). When the noise level is relatively low with $\overline{\text{SNR}} = 30$, the CUSAL-SP provides the best unmixing quality with the highest SRE value for $K < 7$, while ℓ_1 -CENMF leads to the best unmixing quality, especially for $K > 10$. Still, the proposed CUSAL-SP always outperforms SUnSAL-SP, which is not the case of ℓ_1 -CENMF.

VI. EXPERIMENTS WITH REAL DATA

This section presents the performance of the proposed algorithms on a real hyperspectral image. We consider a 250×190 subimage 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 [7], [25]. The raw data contain $L = 224$ bands, covering a wavelength range $0.4\text{--}2.5 \mu\text{m}$. 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 up to 14 minerals [3]. Neglecting the similar signatures, we consider 12 endmembers as often investigated in the literature [7], [37]. The VCA technique is first applied to extract these endmembers on the clean image with $L = 187$ bands. 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. Therefore, the experiments

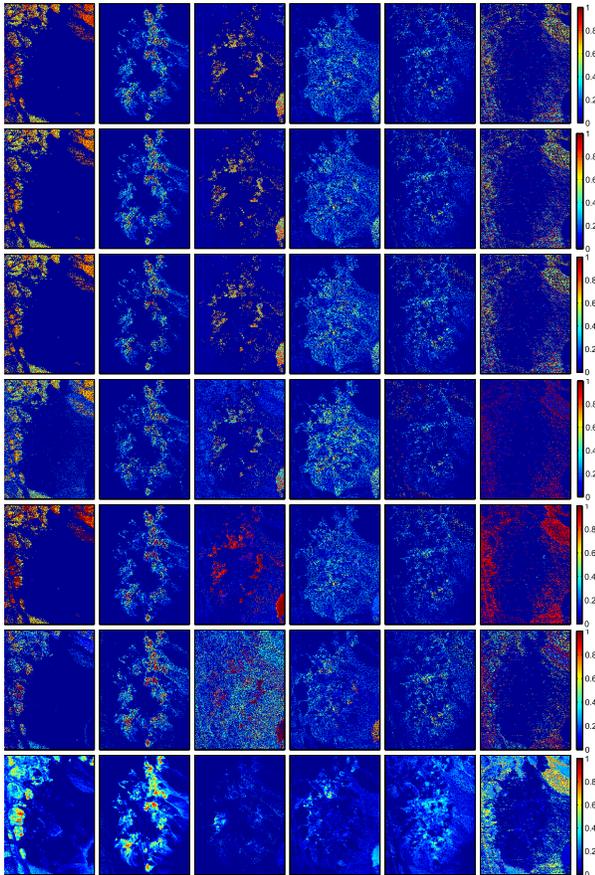


Fig. 8. Cuprite image. Estimated abundance maps using all the 224 bands, with 187 clean bands. The same legend as Fig. 6.

TABLE VII
COMPUTATIONAL TIME (ms/pixel) ON UNMIXING
THE CUPRITE IMAGE USING 224 BANDS

SUnSAL-FCLS	0.89
FCLS	0.42
BayGBM	34.80
KFCLS	0.39
NMF-MCC	3.14
ℓ_1 -CENMF	5.54
CUSAL-FC	20.94

are conducted with $L = 187, 193, 199, 205, 211, 217, 223,$ and 224 bands.

Since ground-truth abundances are unknown, the performance is measured with the averaged spectral angle distance (SAD) between the input spectra \mathbf{y}_t and the reconstructed ones $\hat{\mathbf{y}}_t$, as illustrated in Fig. 5, where the SAD is defined by

$$\text{SAD} = \frac{1}{T} \sum_{t=1}^T \arccos \left(\frac{\mathbf{y}_t^\top \hat{\mathbf{y}}_t}{\|\mathbf{y}_t\| \|\hat{\mathbf{y}}_t\|} \right).$$

The estimated abundance maps using 187, 205, and 224 bands are given in Figs. 6–8, respectively. In the absence of noisy bands (i.e., $L = 187$ bands), all the methods lead to satisfactory abundance maps, with NMF-MCC 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 algorithms and their robustness in the presence of corrupted spectral bands. The MATLAB (R2010) average implementation times per pixel in milliseconds are shown in Table VII, when experiments are performed with all $L = 224$ bands. The estimated time for CUSAL-FC includes the estimation of the parameter σ .

VII. CONCLUSION

This paper presented a supervised unmixing algorithm based on the correntropy maximization principle. Two correntropy-based unmixing problems were addressed, the first with the nonnegativity and sum-to-one constraints and the second with the nonnegativity constraint and a sparsity-promoting term. The ADMM was investigated in order to solve the correntropy-based unmixing problems. The robustness of the proposed unmixing method was validated on synthetic and real hyperspectral images. Future works include the generalization of the correntropy criterion to account for the multiple reflection phenomenon [31], [38], [39], as well as incorporating nonlinear models [40].

REFERENCES

- [1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, 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 Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2185–2195, Jun. 2012.
- [3] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
- [4] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, no. 10, p. 266, 1999.
- [5] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.
- [6] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] A. Huck, M. Guillaume, and J. Blanc-Talon, "Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2590–2602, Jun. 2010.
- [11] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
- [12] J. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2010, pp. 1–4.

- [13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
- [14] A. Zelinski and V. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp. (IGARSS)*, Jul./Aug. 2006, pp. 387–390.
- [15] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2007, pp. 4041–4044.
- [16] J. Chen, C. Richard, and P. Honeine, "Estimating abundance fractions of materials in hyperspectral images by fitting a post-nonlinear mixing model," in *Proc. IEEE Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2013, pp. 1–4.
- [17] W. Liu, P. P. Pokharel, and J. C. Príncipe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
- [18] J. C. Príncipe, *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*. New York, NY, USA: Springer, 2010.
- [19] J. J.-Y. Wang, X. Wang, and X. Gao, "Non-negative matrix factorization by maximizing correntropy for cancer clustering," *BMC Bioinform.*, vol. 14, no. 1, p. 107, 2013.
- [20] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.
- [21] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy-based metric," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4027–4040, Nov. 2015.
- [22] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, 2005.
- [23] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Classics in Applied Mathematics). Philadelphia, PA, USA: SIAM, 1987.
- [24] 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 Process.*, vol. 81, pp. 945–974, May 2001.
- [25] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [26] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [27] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
- [28] B. Chen and J. C. Príncipe, "Maximum correntropy estimation is a smoothed map estimation," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 491–494, Aug. 2012.
- [29] 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.
- [30] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
- [31] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.
- [32] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, "Unmixing hyperspectral images using the generalized bilinear model," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Jul. 2011, pp. 1886–1889.
- [33] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [34] A. Halimi, N. Dobigeon, and J.-Y. Tourneret, "Unsupervised unmixing of hyperspectral images accounting for endmember variability," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4904–4917, Dec. 2015.
- [35] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.
- [36] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.
- [37] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.
- [38] W. Fan, B. Hu, J. Miller, and M. Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *Int. J. Remote Sens.*, vol. 30, no. 11, pp. 2951–2962, Jun. 2009.
- [39] A. Halimi, J. M. Bioucas-Dias, N. Dobigeon, G. S. Buller, and S. McLaughlin, "Fast hyperspectral unmixing in presence of nonlinearity or mismodeling effects," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 146–159, Jun. 2017.
- [40] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, Oct. 2016.



and hyperspectral image analysis.

Fei Zhu was born in Liaoning, China, in 1988. She received the B.S. degree in mathematics, applied mathematics, and economics from Xi'an Jiaotong University, Xi'an, China, in 2011, and the M.S. and Ph.D. degrees in systems optimization and security from the University of Technology of Troyes, Troyes, France, in 2013 and 2016, respectively.

Since 2016, she has been an Assistant Professor with the Center for Applied Mathematics, Tianjin University, Tianjin, China. Her research interests include kernel methods, nonlinear signal processing, and hyperspectral image analysis.



Research Associate with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, U.K.

Abderrahim Halimi (S'11–M'14) received the Engineering Degree in electronics from the National Polytechnic School of Algiers, El Harrach, Algeria, in 2009, and the M.Sc. and Ph.D. degrees in signal processing from the Institut National Polytechnique de Toulouse, University of Toulouse, Toulouse, France, in 2010 and 2013, respectively.

From 2013 to 2015, he was a Post-Doctoral Research Associate with the University of Toulouse and the University of Technology of Troyes, Troyes, France. Since 2015, he has been a Post-Doctoral

Research Associate with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, U.K. His research interests include statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to hyperspectral imaging, satellite altimetry, and single photon depth imaging.



Paul Honeine (M'07) was born in Beirut, Lebanon, in 1977. He received the Dipl.-Ing. degree in mechanical engineering and the M.Sc. degree in industrial control from the Faculty of Engineering, Lebanese University, Beirut, in 2002 and 2003, respectively, and the Ph.D. degree in systems optimization and security from the University of Technology of Troyes, Troyes, France, in 2007.

He was a Post-Doctoral Research Associate with the Systems Modeling and Dependability Laboratory, University of Technology of Troyes, from 2007 to 2008, where he was also an Assistant Professor from 2008 to 2015. Since 2015, he has been a Full Professor with the LITIS Laboratory, University of Rouen (Normandie Université), Mont-Saint-Aignan, France. He has authored more than 100 peer-reviewed papers. His research interests include non-stationary signal analysis and classification, nonlinear and statistical signal processing, sparse representations, machine learning, applications to (wireless) sensor networks, biomedical signal and image processing, hyperspectral imagery, and nonlinear adaptive system identification.

Dr. Honeine was a recipient of the 2009 Best Paper Award from the IEEE Workshop on Machine Learning for Signal Processing for a paper which he had co-authored with C. Richard.



Badong Chen (M'10–SM'13) received the B.S. and M.S. degrees in control theory and engineering from Chongqing University, Chongqing, China, in 1997 and 2003, respectively, and the Ph.D. degree in computer science and technology from Tsinghua University, Beijing, China, in 2008.

He was a Post-Doctoral Researcher with Tsinghua University from 2008 to 2010. He was a Post-Doctoral Associate with the Computational NeuroEngineering Laboratory, University of Florida, Gainesville, FL, USA, from 2010 to 2012. He was

a Visiting Research Scientist with Nanyang Technological University, Singapore, in 2015. He is currently a Professor with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China. He has authored two books, three chapters, and over 100 papers in various journals and conference proceedings. His research interests include signal processing, information theory, machine learning, and their applications in cognitive science and engineering.

Dr. Chen is an Associate Editor of the *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS* and the *Journal of the Franklin Institute*. He has been on the editorial board of *Entropy*.



Nanning Zheng (SM'93–F'06) received the bachelor's degree from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1975, the M.S. degree in information and control engineering from Xi'an Jiaotong University, in 1981, and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985.

He is currently a Professor and the Director with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University. His research interests include computer vision, pattern recognition and image processing, and hardware implementation of intelligent systems.

Prof. Zheng was a member of the Chinese Academy of Engineering in 1999, and he is the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He serves as an Executive Deputy Editor of the *Chinese Science Bulletin*.