Correntropy Maximization via ADMM

– Application to Robust Hyperspectral Unmixing –

Fei Zhu, Abderrahim Halimi, *Member, IEEE,* Paul Honeine, *Member, IEEE*, Badong Chen, *Senior Member, IEEE,* 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 non-negativity and sum-to-one constraints, while the second one deals with the non-negativity 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

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 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 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 two tasks: (i) identifying these pure materials (the so-called *endmembers*); (ii) 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 endmember extraction algorithms include the pure-pixel-based ones, *e.g.*, the vertex component analysis (VCA) [3]

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

A. Halimi is with the School of Engineering and Physical Sciences, Heriot-Watt University, U.K. (a.halimi@hw.ac.uk)

P. Honeine is with the LITIS lab, Université de Rouen Normandie, Normandie Université, France. (paul.honeine@univ-rouen.fr)

Badong Chen and Nanning Zheng are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, China. (chenbd; nnzheng@mail.xjtu.edu.cn)

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 [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 non-negativity constraint (ANC) and the abundance sum-to-one constraint (ASC) for each pixel [11]. Considering both constraints, the fully-constrained least-squares method (FCLS) was presented in [9]. A more recently proposed unmixing algorithm is the so-called SUnSAL, for *Sparse Unmixing by variable Splitting and Augmented Lagrangian* [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 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 fully-constrained least-squares (KFCLS) [15], nonlinear fluctuation methods [7] and post-nonlinear 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 nonnegative matrix factorization [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 alternating direction method of multipliers (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, Section 9]. We show that ADMM provides a relevant framework for incorporating different constraints raised in the unmixing problem. We present the so-called CUSAL (for *Correntropy-based Unmixing by variable Splitting and Augmented Lagrangian*), 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 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 the 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 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 $t = 1, \ldots, T$. The LMM can be written as

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

where $\boldsymbol{M} = [\boldsymbol{m}_1 \ \cdots \ \boldsymbol{m}_R] \in \mathbb{R}^{L\times R}$ is the matrix composed by the R endmembers with $\boldsymbol{m}_r = [m_{1r} \ \cdots \ m_{Lr}]^\top$, $\boldsymbol{x}_t = [x_{1t} \ \cdots \ x_{Rt}]^\top$ 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 $\boldsymbol{X} = [\boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_T] \in \mathbb{R}^{R\times T}$ and $\boldsymbol{N}$ 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

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

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

$$\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_F^2, \tag{3}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. Its solution is

$$\boldsymbol{X}_{\mathrm{LS}} = (\boldsymbol{M}^\top\boldsymbol{M})^{-1}\boldsymbol{M}^\top\boldsymbol{Y}. \tag{4}$$

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

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

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 fully-constrained least-squares problem is formulated as, for each $t = 1, \ldots, T$,

$$\min_{\boldsymbol{x}_t} \|\boldsymbol{y}_t - \boldsymbol{M}\boldsymbol{x}_t\|_2^2, \text{ subject to } \boldsymbol{x}_t \succeq 0 \text{ and } \mathbf{1}^\top \boldsymbol{x}_t = 1,$$

where $\mathbf{1} \in \mathbb{R}^{R\times 1}$ denotes the column vector of ones and $\succeq 0$ is the non-negativity applied element-wise; In matrix form:

$$\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_F^2, \text{ subject to } \boldsymbol{X} \succeq 0 \text{ and } \mathbf{1}^\top \boldsymbol{x}_t = 1, \text{ for } t = 1, \ldots, T.$$

Since there is no closed-form solution when dealing with the non-negativity 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 fully-constrained least-squares (FCLS) technique [9]. More recently, the alternating direction method of multipliers (ADMM) was applied with success for hyperspectral unmixing problem, with the SUnSAL algorithm [12].

Recent work in hyperspectral unmixing have 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 $t = 1, \ldots, T$,

$$\min_{\boldsymbol{x}_t} \|\boldsymbol{y}_t - \boldsymbol{M}\boldsymbol{x}_t\|_2^2 + \lambda\|\boldsymbol{x}_t\|_1, \text{ subject to } \boldsymbol{x}_t \succeq 0,$$

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 by using the following matrix formulation

$$\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}\|_F^2 + \lambda \sum_{t=1}^{T} \|\boldsymbol{x}_t\|_1, \text{ subject to } \boldsymbol{X} \succeq 0.$$

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

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

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 $(\boldsymbol{M}\boldsymbol{X})_{l*}$ 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 overcome this difficulty by considering the correntropy maximization principle from the information theoretic learning, which yields an optimization problem that is robust to outliers.

TABLE I

COMMONLY-USED SHIFT-INVARIANT KERNELS

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

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], [18], is a nonlinear local similarity measure. For two random variables, $\mathcal{Y}$ and its estimation $\widehat{\mathcal{Y}}$ using some model/algorithm, it is defined by

$$\mathbb{E}[\kappa(\mathcal{Y},\widehat{\mathcal{Y}})], \tag{6}$$

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 $\widehat{\mathcal{Y}}$ is unavailable, the sample estimator of correntropy is adopted instead. Employing a finite number of data $\{(\boldsymbol{y}_{l*},\widehat{\boldsymbol{y}}_{l*})\}_{l=1}^{L}$, it is estimated by

$$\frac{1}{L}\sum_{l=1}^{L}\kappa(\boldsymbol{y}_{l*},\widehat{\boldsymbol{y}}_{l*}), \tag{7}$$

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

$$\frac{1}{L}\sum_{l=1}^{L}\exp\left(\tfrac{-1}{2\sigma^2}\|\boldsymbol{y}_{l*}-\widehat{\boldsymbol{y}}_{l*}\|_2^2\right), \tag{8}$$

where σ denotes the bandwidth of the Gaussian kernel.

The maximization of the correntropy, given by

$$\max_{\widehat{\boldsymbol{y}}_{1*},\ldots,\widehat{\boldsymbol{y}}_{L*}}\frac{1}{L}\sum_{l=1}^{L}\kappa(\boldsymbol{y}_{l*},\widehat{\boldsymbol{y}}_{l*}),$$

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 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.

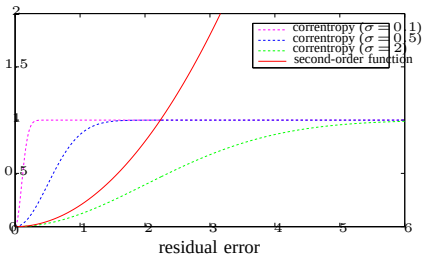


Fig. 1. Illustration of the second-order objective function (ϵ_l^2, in solid red line) and the negative correntropy objective function ($1-\exp(\frac{-1}{2\sigma^2}\epsilon_l^2)$, in dashed lines for several values of σ), in terms of the residual error (ϵ_l).

B. The 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 $\epsilon_l = \|\boldsymbol{y}_{l*} - \widehat{\boldsymbol{y}}_{l*}\|_2$. Thus, the correntropy (8) becomes

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

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 only sensitive 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 $\boldsymbol{X}$, by minimizing the objective function $\mathcal{C}$ (the negative of correntropy), given by

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

where the Gaussian kernel was considered, or equivalently

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

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

$$\begin{aligned}\min_{\boldsymbol{X}}\ \mathcal{C}(\boldsymbol{X}),\ \text{subject to}\ \boldsymbol{X}\succeq 0\\ \text{and}\ \mathbf{1}^\top\boldsymbol{x}_t = 1,\ \text{for}\ t=1,\ldots,T.\end{aligned} \tag{11}$$

To promote sparse representations, the objective function (9)-(10) can be augmented by the ℓ_1-norm penalty on the abundance matrix $\boldsymbol{X}$, leading to the following problem

$$\min_{\boldsymbol{X}}\mathcal{C}(\boldsymbol{X}) + \lambda\sum_{t=1}^{T}\|\boldsymbol{x}_t\|_1,\ \text{subject to}\ \boldsymbol{X}\succeq 0. \tag{12}$$

IV. ADMM FOR SOLVING THE CORRENTROPY-BASED UNMIXING PROBLEMS

We first briefly review the alternating direction method of multipliers (ADMM), following the expressions in [13, Chap. 3]. Consider an optimization problem of the form

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) + g(\boldsymbol{x}),$$

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

$$\min_{\boldsymbol{x},\boldsymbol{z}} f(\boldsymbol{x}) + g(\boldsymbol{z}) \text{ subject to } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}, \tag{13}$$

such as having the particular constraint $\boldsymbol{x} = \boldsymbol{z}$ 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 $\boldsymbol{x}$ and $\boldsymbol{z}$. 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 the problem (13) has a saddle point, the ADMM iterates, as given in Algorithm 1, satisfy the following: (i) the objective function convergence; (ii) the primal residual convergence, *i.e.*, $\|\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1}\|_2 \to 0$; and (iii) the dual residual convergence, *i.e.*, $\rho\|\boldsymbol{z}_{k+1} - \boldsymbol{z}_k\|_2 \to 0$. See [13] for more details.

Algorithm 1 The ADMM algorithm [13]

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

1: Initialize $k = 0$, $\boldsymbol{x}_0$, $\boldsymbol{z}_0$ and $\boldsymbol{u}_0$
2: **repeat**
3: $\quad \boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z}_k - \boldsymbol{c} + \boldsymbol{u}_k\|_2^2$;
4: $\quad \boldsymbol{z}_{k+1} = \arg\min_{\boldsymbol{z}} g(\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c} + \boldsymbol{u}_k\|_2^2$;
5: $\quad \boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \boldsymbol{A}\boldsymbol{x}_{k+1} + \boldsymbol{B}\boldsymbol{z}_{k+1} - \boldsymbol{c}$;
6: $\quad k = k + 1$;
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 $\boldsymbol{x} \in \mathbb{R}^{RT\times 1}$, which is stacked by the columns of the matrix $\boldsymbol{X}$, namely $\boldsymbol{x} = [\boldsymbol{x}_1^\top \ \cdots \ \boldsymbol{x}_T^\top]^\top$. Rewrite also the following vectors in $\mathbb{R}^{RT\times 1}$: $\boldsymbol{z} = [\boldsymbol{z}_1^\top \ \cdots \ \boldsymbol{z}_T^\top]^\top$ and $\boldsymbol{u} = [\boldsymbol{u}_1^\top \ \cdots \ \boldsymbol{u}_T^\top]^\top$, where, for $t = 1, \ldots, T$, $\boldsymbol{z}_t = [z_{1t} \ \cdots \ z_{Rt}]^\top$ and $\boldsymbol{u}_t = [u_{1t} \ \cdots \ u_{Rt}]^\top$. By following the formulation of the ADMM in Algorithm 1, we set

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

where $\boldsymbol{I}$ 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 $\boldsymbol{x}$-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 $t = 1, \ldots, T$. Let $\overline{\boldsymbol{x}} \in \mathbb{R}^{(R-1)T\times 1}$ be the reduced vector of $(R-1)$ unknowns to be estimated, stacked by

$$\overline{\boldsymbol{x}}_t = \begin{bmatrix} x_{1t} & \cdots & x_{(R-1)t} \end{bmatrix}^\top,$$

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

$$f_1(\overline{\boldsymbol{x}}) = -\sum_{l=1}^{L} \exp\left(\frac{-1}{2\sigma^2} \sum_{t=1}^{T} \epsilon_l(\overline{\boldsymbol{x}}_t)^2\right), \tag{15}$$

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

$$\frac{\partial f_1}{\partial \overline{\boldsymbol{x}}} = \begin{bmatrix} \frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_1}^\top & \cdots & \frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_T}^\top \end{bmatrix}^\top \in \mathbb{R}^{(R-1)T\times 1},$$

where $\frac{\partial f_1}{\partial \overline{\boldsymbol{x}}_t} = \begin{bmatrix} \frac{\partial f_1}{\partial \overline{x}_{1t}} & \cdots & \frac{\partial f_1}{\partial \overline{x}_{(R-1)t}} \end{bmatrix}^\top$, with the entries given by

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

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

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

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

$$\frac{\partial \phi(\overline{\boldsymbol{x}})}{\partial \overline{x}_{rt}} = \rho\Big(x_{rt} + \sum_{p=1}^{R-1} x_{pt} - 1 + z_{Rt,k} - z_{rt,k} + u_{Rt,k} - u_{rt,k}\Big), \tag{16}$$

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

The solution of the $\boldsymbol{z}$-update in line 4 Algorithm 1 becomes the projection of $\boldsymbol{x}_{k+1} - \boldsymbol{u}_k$ 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

$$f(\boldsymbol{x}) = \mathcal{C}(\boldsymbol{x}) \tag{17}$$

$$g(\boldsymbol{z}) = \iota_{\mathbb{R}_+^{RT}}(\boldsymbol{z}) + \lambda\|\boldsymbol{z}\|_1$$

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

Algorithm 2 Correntropy-based unmixing with full-constraints (**CUSAL-FC**)

1: Initialize $k = 0$, $\rho > 0$, $\eta > 0$, $\sigma > 0$; $\boldsymbol{x}_0$, $\boldsymbol{z}_0$ 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: $\boldsymbol{z}_{k+1} = \max(\mathbf{0}, \boldsymbol{x}_{k+1} - \boldsymbol{u}_k)$;
8: $\boldsymbol{u}_{k+1} = \boldsymbol{u}_k - (\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1})$;
9: $k = k + 1$;
10: **until** stopping criterion

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

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

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

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

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 $\boldsymbol{z}$-update in LASSO is that in the latter, no non-negativity term $\iota_{\mathbb{R}_+^{RT}}(\boldsymbol{z})$ is enforced. In this case, the $\boldsymbol{z}$-update in LASSO is the element-wise soft thresholding operation

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

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 non-negativity constraint, thus yielding

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

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.

Algorithm 3 Correntropy-based unmixing with sparsity-promoting (**CUSAL-SP**))

Initialize $k = 0$, $\rho > 0$, $\sigma > 0$, $\eta > 0$, $\lambda > 0$; $\boldsymbol{x}_0$, $\boldsymbol{z}_0$ and $\boldsymbol{u}_0$;
repeat
  repeat
    $\boldsymbol{x}_{k+1} = \boldsymbol{x}_{k+1} - \eta\big(\frac{\partial f}{\partial \boldsymbol{x}_{k+1}} + \rho(\boldsymbol{x}_{k+1} - \boldsymbol{z}_k - \boldsymbol{u}_k)\big)$;
  until convergence
  $\boldsymbol{z}_{k+1} = \max(\boldsymbol{0}, S_{\lambda/\rho}(\boldsymbol{x}_{k+1} - \boldsymbol{u}_k))$;
  $\boldsymbol{u}_{k+1} = \boldsymbol{u}_k - (\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1})$;
  $k = k + 1$;
until stopping criterion

C. On the bandwidth determination and convergence

We apply a three-fold stopping criterion for Algorithms 2 and 3, according to [13], [12]: (i) the primal and dual residuals are small enough, namely $\|\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1}\|_2 \leq \epsilon_1$ and $\rho\|\boldsymbol{z}_{k+1} - \boldsymbol{z}_k\|_2 \leq \epsilon_2$, where $\epsilon_1 = \epsilon_2 = \sqrt{RT} \times 10^{-5}$ as in [12], (ii) the primal residual starts to increase, *i.e.*, $\|\boldsymbol{x}_{k+1} - \boldsymbol{z}_{k+1}\|_2 > \|\boldsymbol{x}_k - \boldsymbol{z}_k\|_2$, or (iii) 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 study, the ADMM is applied to address a nonconvex objective function, 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.

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

$$\sigma_0^2 = \frac{\theta}{LT}\|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{X}_{\text{LS}}\|_F^2, \tag{19}$$

where the parameter θ is chosen as $\theta = \frac{RT}{2}$ in this study, and $\boldsymbol{X}_{\text{LS}}$ 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 $\frac{\|\boldsymbol{Y}-\boldsymbol{M}\boldsymbol{X}\|_F}{\|\boldsymbol{Y}-\boldsymbol{M}\boldsymbol{X}_{\text{LS}}\|_F} < 2$ is satisfied. The algorithm divergence occurs if the stopping criterion (ii) 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 which transforms the original problem to a re-weighted 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.

Algorithm 4 Tuning the bandwidth parameter σ

1: Initialize $\sigma = \sigma_0$ using (19); $p = 1$;
2: Do CUSAL with Algorithm 2 or Algorithm 3;
3: **if** stopping criterion (i) or (iii) is satisfied **then**
4: **if** condition $\frac{\|\boldsymbol{Y}-\boldsymbol{M}\boldsymbol{X}\|_F}{\|\boldsymbol{Y}-\boldsymbol{M}\boldsymbol{X}_{\text{LS}}\|_F} < 2$ 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 σ_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**

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.

- Fully-Constrained Least-Squares (**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.
- Sparse Unmixing by variable Splitting and Augmented Lagrangian (**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-sparse.
- 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

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

  where $0 \leq \gamma_{ij,t} \leq 1$ controls the interactions between endmembers $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, and $\odot$ is the element-wise product. The BayGBM considers both ANC and ASC.
- Kernel Fully-Constrained Least-Squares (**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.

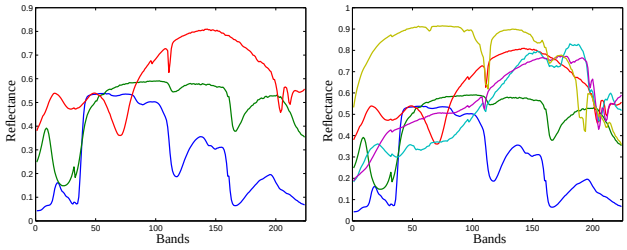


Fig. 2. The $R = 3$ (left) and 6 (right) USGS signatures chosen for simulation.

- Nonnegative Matrix Factorization 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.
- 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.
- 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 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 $0.2\mu m$ to $3.0\mu m$. The abundance vectors $\boldsymbol{x}_t$ are uniformly generated using a Dirichlet distribution as in [33], [34].

The unmixing performance is evaluated using the abundance root mean square error (RMSE) [31], [35], 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.

In the first set of experiments, the SNR 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]. TABLE II and III illustrate the average of RMSE 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

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**

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**

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 is similarly generated as previously described, where two normal distributions SNR $\sim \mathcal{N}(\overline{\text{SNR}_1}, \epsilon^2)$ and 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 $\overline{\text{SNR}_2} \in \{5, 10, 15\}$. TABLE 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 contains 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 SNR $= 30$ is added. The $R = 3$ endmembers, as shown in Fig. 2 (left), is considered while the abundance vectors $\boldsymbol{x}_t$ are uniformly generated using a Dirichlet distribution. The changes of the objective function value, the primal and dual residuals 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

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**

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**

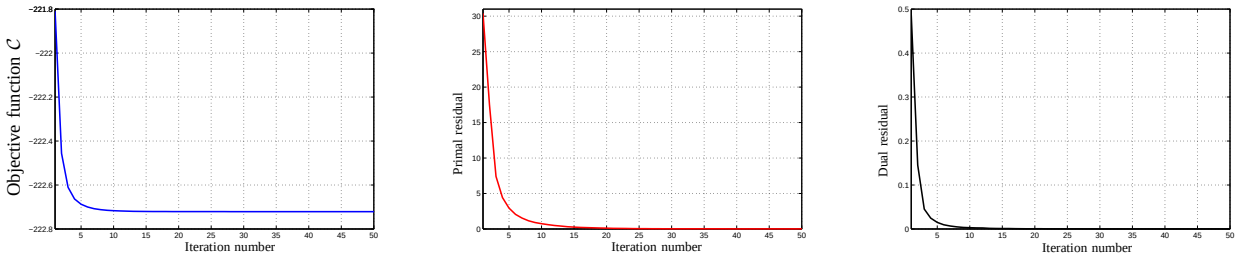
accuracy. We examine the robustness of CUSAL-FC with respect to the initialization, and compare it with the half-quadratic approach proposed for maximum correntropy criterion, namely NMF-MCC [19]. The elements in $\boldsymbol{X}$ are identically initialized for each method, using (i) uniform distribution $x_{rt} \sim \mathcal{U}(0,1)$; (ii) normal distribution $x_{rt} \sim \mathcal{N}(0.5, 0.1)$ combined with rounding all the negative values up to zero; and (iii) 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]

[1]It 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.

TABLE VI

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

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



(a) Objective function value (b) Primal residual (c) Dual residual

Fig. 3. Illustration of the convergence property of CUSAL-FC, by plotting the objective function value, the primal and dual residuals over the first 50 iterations, using a fixed value of μ=0.01.

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-sparse and ℓ_1-CENMF, on a series of data with sparse abundance matrices. We study the influence of (i) the noise level over bands, namely $\overline{\text{SNR}}$, and (ii) the sparsity level of the abundances. Each image, of 15×15 pixels, is generated by the linear mixture model. The endmember matrix is composed by $R = 62$ USGS signatures, where the angle between any two different endmembers is larger than 10° [25]. The K nonzero entries in each abundance vector $\boldsymbol{x}_t$ 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

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

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

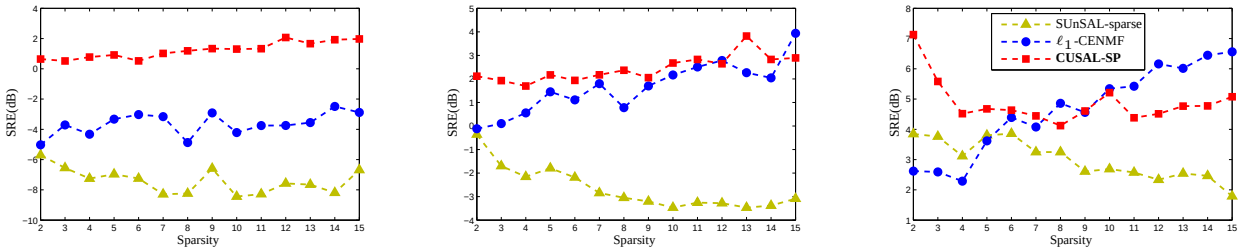
The unmixing performance with the sparsity-promoting algorithms is evaluated using the signal-to-reconstruction error, measured in decibels, according to [12], [25]. It is defined by

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

The results, averaged over ten 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 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 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 [7], [25]. The raw data contains $L = 224$ bands, covering a wavelength



(a) $\overline{\text{SNR}} = 10$ (b) $\overline{\text{SNR}} = 20$ (c) $\overline{\text{SNR}} = 30$

Fig. 4. The averaged signal-to-reconstruction error (SRE) with respect to the sparsity level K, averaged over ten Monte-Carlo realizations. Comparison for various average noise level $\overline{\text{SNR}}$.

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 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 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 $\boldsymbol{y}_t$ and the reconstructed ones $\widehat{\boldsymbol{y}}_t$, as illustrated in Fig. 5, where the SAD is defined by

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

The estimated abundance maps using 187, 205 and 224 bands are given in Fig. 6, Fig. 7, and Fig. 8, respectively. In 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 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.

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

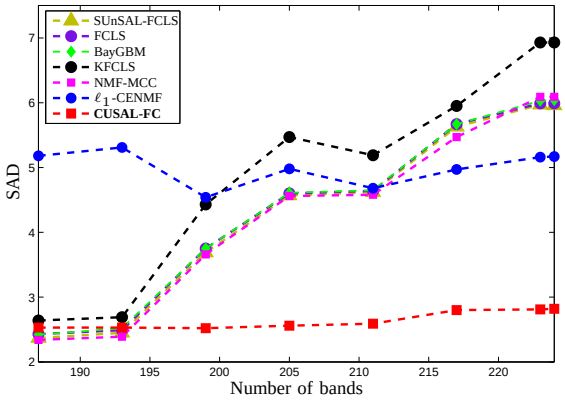


Fig. 5. **Cuprite image:** The averaged spectral angle distance (SAD) using different number of bands, computed without the noisy bands $1-3, 105-115, 150-170$, and $223-224$.

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 non-negativity and sum-to-one constraints, and the second with the non-negativity constraint and a sparsity-promoting term. The alternating direction method of multipliers (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].

Acknowledgment

This work was partially supported by the French ANR under grant HYPANEMA: ANR-12BS03-0033 and by the National Natural Science Fondations of China under grant 11626250.

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] 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.

[4] M. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. of SPIE: Imaging Spectrometry V*, vol. 3753, no. 10, 1999.

[5] J. Li, A. Agathos, D. Zaharie, J. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5067–5082, Sept. 2015.

[6] 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, Mar. 2007.

[7] 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.

[8] ——, "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.

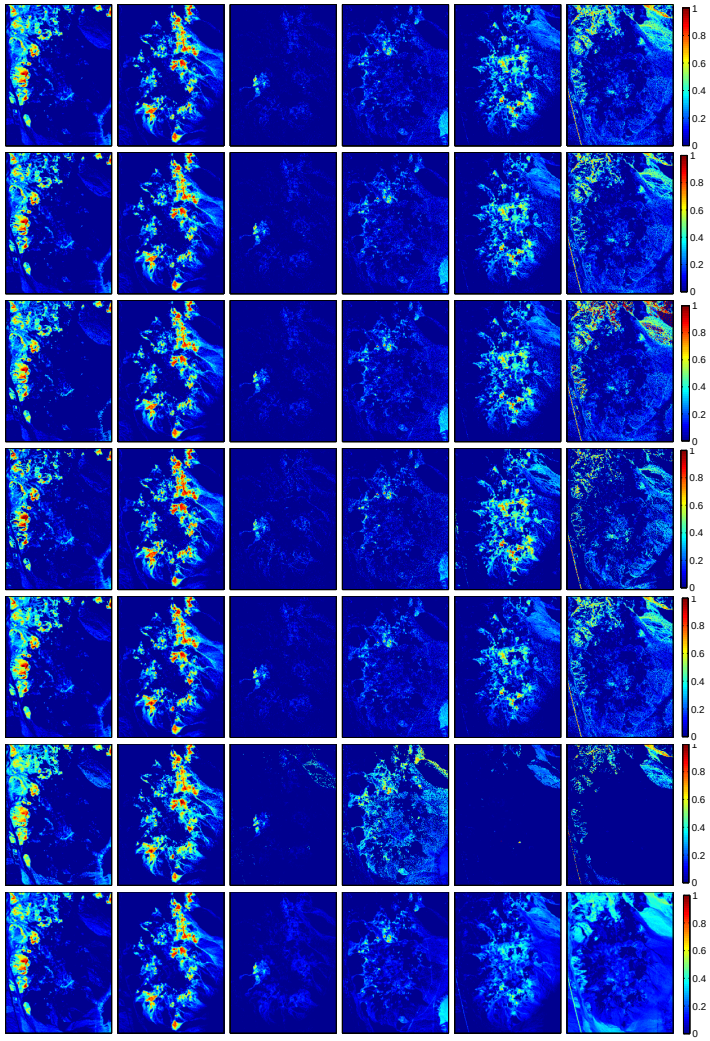


Fig. 6. **Cuprite image:** Estimated abundance maps using 187 clean bands. **Left to right:** sphene, alunite, buddingtonite, kaolinite, chalcedony, highway. **Top to bottom:** SUnSAL-FCLS, FCLS, BayGBM, KFCLS, NMF-MCC, ℓ_1-CENMF, CUSAL-FC.

[9] 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.

[10] 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.

[11] 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, Apr. 2012.

[12] J. M. Bioucas-Dias and M. A. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (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," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[14] 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 (IGARSS)*, Jul. 2006, pp. 387–390.

[15] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina, "Kernel fully constrained least squares abundance estimates," in *IEEE International Geoscience and Remote Sensing Symposium (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,"

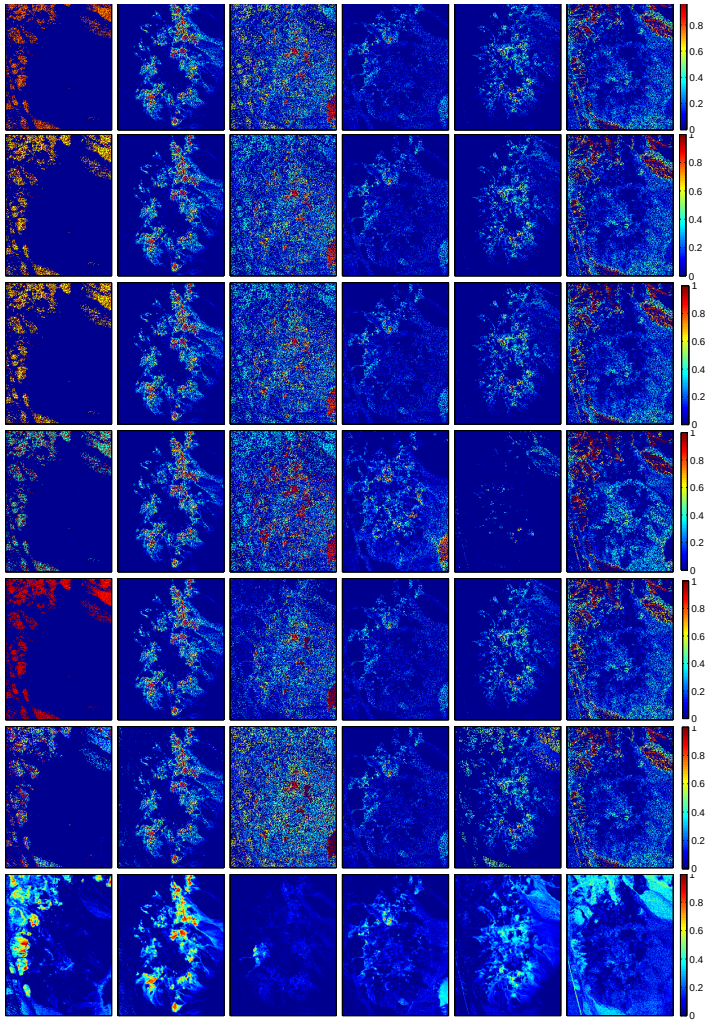


Fig. 7. **Cuprite image:** Estimated abundance maps using 205 bands, with 187 clean bands. Same legend as Fig. 6.

in *IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS)*, Jun. 2013, pp. 1–4.

[17] 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, Nov. 2007.

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

[19] 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.

[20] 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, Aug. 2011.

[21] 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.

[22] 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.

[23] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems (Classics in Applied Mathematics)*. Society for Industrial Mathematics, 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 Processing*, vol. 81, pp. 945–974, May 2001.

[25] 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, Jun. 2011.

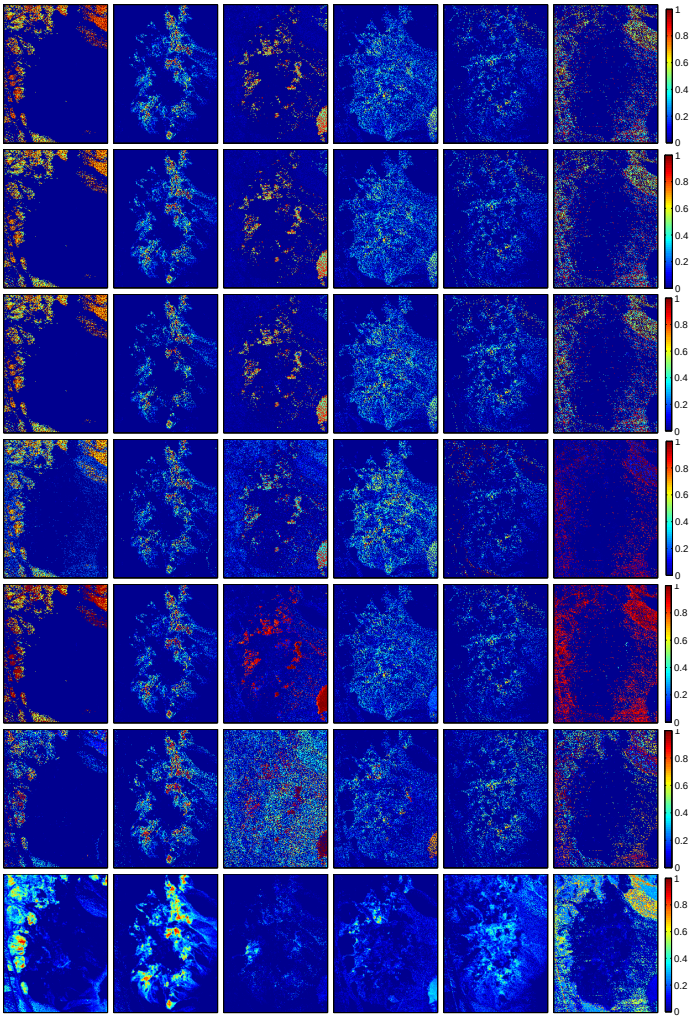


Fig. 8. **Cuprite image:** Estimated abundance maps using all the 224 bands, with 187 clean bands. Same legend as Fig. 6.

[26] M. D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 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 Processing Letters*, 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, p. 7149, 2015.

[30] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 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 Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[32] ——, "Unmixing hyperspectral images using the generalized bilinear model," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Jul. 2011, pp. 1886–1889.

[33] 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.

[34] 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, Dec. 2015.

[35] 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.

[36] P. O. Hoyer and P. Dayan, “Non-negative matrix factorization with sparseness constraints,” *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[37] 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, 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,” *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2951–2962, 2009.

[39] A. Halimi, J. Bioucas-Dias, N. Dobigeon, G. S. Buller, and S. McLaughlin, “Fast hyperspectral unmixing in presence of nonlinearity or mismodelling effects,” *IEEE Transactions on Computational Imaging*, 2017, To appear.

[40] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, “Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects,” *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4565–4579, Oct. 2016.