

Article

Kernel-Based Nonlinear Spectral Unmixing with Dictionary Pruning

Zeng Li, Jie Chen *  and Susanto Rahardja 

School of Marine Science and Technology, Northwestern Polytechnical University; Key Laboratory of Ocean Acoustics and Sensing, Ministry of Industry and Information Technology, Xi'an 710072, China; lizeng@mail.nwpu.edu.cn (Z.L.); susantorahardja@ieee.org (S.R.)

* Correspondence: dr.jie.chen@ieee.org

Received: 21 January 2019; Accepted: 1 March 2019; Published: 5 March 2019



Abstract: Spectral unmixing extracts subpixel information by decomposing observed pixel spectra into a collection of constituent spectra signatures and their associated fractions. Considering the restriction of linear unmixing model, nonlinear unmixing algorithms find their applications in complex scenes. Kernel-based algorithms serve as important candidates for nonlinear unmixing as they do not require specific model assumption and have moderate computational complexity. In this paper we focus on the linear mixture and nonlinear fluctuation model. We propose a two-step kernel-based unmixing algorithm to address the case where a large spectral library is used as the candidate endmembers or the sparse mixture case. The sparsity-inducing regularization is introduced to perform the endmember selection and the candidate library is then pruned to provide more accurate results. Experimental results with synthetic and real data, particularly those laboratory-created labeled, show the effectiveness of the proposed algorithm compared with state-of-art methods.

Keywords: hyperspectral images; spectral unmixing; nonlinear unmixing; kernel model; sparse model

1. Introduction

Hyperspectral images integrate the spectrum representing radiant attributes of an observed scene with homological images standing for spatial and geometric relations. Thanks to the high spectrum resolution and rich spectral information, hyperspectral imaging has been widely used in applications ranging from forest mapping and monitoring, land cover change detection, mineral exploration, object identifying, etc.

Hyperspectral image analysis plays a key role in revealing information from high-dimensional data. In the last decade, spectral unmixing has received considerable attention. It aims to decompose the observed pixel spectra into a collection of constituent spectra signatures (called endmembers) and estimate fractions associated with each component (called abundances) [1]. Spectral unmixing is particularly useful to analyze pixels with low spatial or with photon interactions among materials. Unmixing techniques have been extensively studied and most of them perform the endmember extraction and abundance estimation either in an independent or simultaneous manner. Provided that the pure pixels are present in the observed scene, methods such as pixel purity index algorithm (PPI) [2], vertex component analysis (VCA) [3] and the N-FINDR algorithm [4] are proposed for extracting endmembers. Meanwhile some algorithms consider to generate virtual endmembers and abundances due to the absence of pure pixels, for example, the minimum volume simplex analysis (MVSA) [5], the minimum volume constrained nonnegative matrix factorization [6]. To pursue interpretable and tractable solutions, most algorithms are based on the linear mixture model (LMM). The LMM assumes that an observed pixel is a combination of signature spectra weighted by the abundances and to be physically meaningful, the problem is usually subject to two constraints: the abundance nonnegativity

constraint (ANC) and the abundance sum-to-one constraint (ASC). The linear unmixing algorithms can be classified as the least squares principle [7], statistical framework [8,9], sparse regression [10,11], and independence component analysis (ICA) based algorithms [9,12,13].

Since the LMM may not be appropriate in some practical situations where the light is scattered by multiple reflective or interacted materials [14], the nonlinear mixture model (NLMM) then provides an alternative to overcome the limitation of the LMM. For instance, a specific class of nonlinear models referred to as bilinear models has been studied in [15–19] for modeling the second-order reflectance by introducing extra additional bilinear terms to LMM. The postnonlinear mixture model (PNMM) has also been introduced in [20,21] that considers an appropriate nonlinear function mapping from $[0, 1]^L$ into $[0, 1]^L$ [18]. In [22], based on nonlinear models, the authors proposed a parameter-free unmixing algorithm where abundance fractions and a set of nonlinear parameters are under the specific constraints which can be satisfied by minimizing the penalty function. The above algorithms are presented based on certain models, and thus lack the flexibility. The neural networks based methods can be regarded as model-free methods. In [23], the authors designed a multi-layer perceptron neural network combined with a Hopfield neural network to deal with nonlinear mixture. In [24] the auto-associative neural network for abundance estimation was introduced for nonlinear unmixing purposes consisting of dimension reduction stage and the mapping stage from feature to the abundance percentages. Furthermore, in [25] the authors proposed an end-to-end unmixing method based on the convolutional neural network to improve the accuracy where the spatial information is taken into account. The kernel-based methods can also be regarded as the model-free methods.

The kernel-based methods serve as one of the most popular tools in addressing nonlinear learning problems. These methods map the data to a feature space of higher dimension where the mapped data can be represented with a linear model in this space [26]. It is important to note that finding the explicit mapping is however bypassed via the kernel trick [27–31]. In [32,33], the authors proposed a model consisting of a linear mixture and a nonlinear fluctuation. The nonlinear fluctuation function characterizes the high-order interactions between endmembers and is restricted within a reproducing kernel Hilbert space (RKHS). The so called K-Hype and SK-Hype algorithms are proposed therein. In contrast to several other classes of algorithms, the kernel-based methods is independent of a specific observation model and has the generalization ability in multiple scenes. Thus, subsequent models have been further studied in the literature. An ℓ_1 -spatial regularization term is added to the K-Hype problem in [34] to promote the piece-wise spatial continuity of the abundance maps. In [35,36], the nonlinear fluctuation term is also called the residual term and the associated algorithm is called the residual component analysis. Post-nonlinear models as well as some robust unmixing models that consider such a fluctuation are presented in [37]. In [38], the authors extended this method by accounting for band-dependent and neighboring nonlinear contributions using separable kernels. Further, kernel-based nonnegative matrix factorization (NMF) techniques are studied to simultaneously capture nonlinear dependence features and estimate the abundance.

The above kernel-based algorithms has its inherent limitations. On one hand, these algorithms focus on the abundance estimation, and do not consider the extraction of the endmembers. On the other hand, the structures of the K-hype type algorithms and its variations impose a regularization with unclear physical interpretation on the abundance vector (as elaborated in Section 2) and the fluctuation is independent of the abundance fractions. In this work, we propose a kernel-based sparse nonlinear spectral unmixing. The nonlinear unmixing algorithm is designed to run with a large number of candidate spectral signatures. A sparse regularization step and a dictionary pruning step are conducted in a sequential manner: the former select endmembers with significant contributions; and the latter then performs the abundance estimation with a more proper optimization problem using the pruned endmember dictionary. The contributions of this work are summarized as follows:

1. A kernel based sparse nonlinear unmixing problem is formulated and a two-step solving strategy is proposed. This strategy allows to use a spectral library for selecting the endmembers and bypass the problem of endmember extraction in the nonlinear unmixing.

2. A more reasonable formulation of the optimization problem is proposed for solving the linear mixture/nonlinear fluctuation model. This formulation improves the K-Hype formulation in several aspects and serves as a key component in the proposed sparse unmixing scheme.
3. The algorithm is tested using real data with ground-truth created in our laboratory. Lack of publicly available data sets with ground truth imposes difficulties to compare unmixing algorithms. Most of the existing works rely on the use of numerically produced synthetic data and real data without ground truth. Using a labeled real data provides a more meaningful comparison results.

The remainder of this paper is organized as follows. Section 2 introduces the related work based on kernel methods. The proposed basic kernel-based algorithm is described in Section 3. Section 4 describes experimental results with simulated hyperspectral data and real hyperspectral data sets. Finally the conclusion is given in Section 5.

2. Kernel-Based Nonlinear Abundance Estimation

Notation. Scalars are denoted by italic letters. Vectors and matrices are denoted by boldface small and capital letters respectively. Specifically, if each pixel of the hyperspectral image consists of a reflectance vector in L contiguous spectral bands, then $\mathbf{r} = [r_1, r_2, \dots, r_L]^T \in \mathbb{R}^L$ is an observed pixel, $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_L]^T \in \mathbb{R}^{L \times R}$ is the endmember matrix which is a spectral library consisting of spectral signatures \mathbf{m}_i with R numbers of endmembers, and $\mathbf{m}_{\lambda_\ell}$ is the vector of R endmember signatures at the ℓ -th wavelength bands of \mathbf{M} , $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_R]^T \in \mathbb{R}^R$ is the abundance vector.

In this section, we first review the general unmixing model and the kernel-based linear mixture/nonlinear fluctuation model proposed in [33]. A general mixing mechanism can be formulated as

$$\mathbf{r} = \psi(\mathbf{M}, \boldsymbol{\alpha}) + \mathbf{n} \quad (1)$$

where ψ is an unknown function that defines the interactions between the endmembers in matrix \mathbf{M} parameterized by their associated abundance fractions $\boldsymbol{\alpha}$, and \mathbf{n} is the modeling noise. Though general enough, this strategy may fail if the function ψ cannot be adequately and finitely parameterized. A semi-parametric model proposed in [33] is described by:

$$\psi(\mathbf{m}_{\lambda_\ell}) = \mathbf{m}_{\lambda_\ell}^T \boldsymbol{\alpha} + \psi_{\text{nl}}(\mathbf{m}_{\lambda_\ell}). \quad (2)$$

This model is composed by a linear mixture term and a nonlinear fluctuation term defined by ψ_{nl} . Several useful nonlinear mixing model can be considered as a specific case of (2). For instance, (2) reduces to a bilinear model if the second-order polynomial is used for ψ_{nl} . Models and algorithms of the residual component analysis follows the same principle [35–37]. Assume that the endmember matrix \mathbf{M} is known, and $\psi_{\text{nl}}(\mathbf{m}_{\lambda_\ell})$ is a real-valued function of a reproducing kernel Hilbert space \mathcal{H} , endowed with the reproducing kernel κ , i.e.,

$$\psi_{\text{nl}}(\mathbf{m}_{\lambda_\ell}) = \langle \psi_{\text{nl}}, \kappa(\cdot, \mathbf{m}_{\lambda_\ell}) \rangle_{\mathcal{H}}. \quad (3)$$

Selecting a proper kernel is essential for describing the mixture. For instance, the second-order homogeneous polynomial kernel

$$\kappa(\mathbf{m}_{\lambda_k}, \mathbf{m}_{\lambda_\ell}) = (\mathbf{m}_{\lambda_k}^T \mathbf{m}_{\lambda_\ell})^2 \quad (4)$$

is able to represent the second-order interactions between endmembers; and the Gaussian kernel

$$\begin{aligned} \kappa(\mathbf{m}_{\lambda_k}, \mathbf{m}_{\lambda_\ell}) &= \exp\left(-\frac{\|\mathbf{m}_{\lambda_k} - \mathbf{m}_{\lambda_\ell}\|^2}{2}\right)^2 \\ &= \exp\left(-\frac{\|\mathbf{m}_{\lambda_k}\|^2 + \|\mathbf{m}_{\lambda_\ell}\|^2}{2\sigma^2}\right) \sum_{i=0}^{\infty} \frac{(\mathbf{m}_{\lambda_k}^T \mathbf{m}_{\lambda_\ell})^i}{i! \sigma^{2i}} \end{aligned} \quad (5)$$

involves an infinity order of interactions since it can be expanded as the sum of polynomials of all orders. Then [33] proposes to evaluate the abundances and the nonlinear function by solving the following optimization problem:

$$\begin{aligned} \alpha^*, \psi_{\text{nl}}^* &= \arg \min_{\psi_{\text{nl}} \in \mathcal{H}} \frac{1}{2} (\|\alpha\|^2 + \|\psi_{\text{nl}}\|_{\mathcal{H}}^2) + \frac{1}{2\mu} \sum_{\ell=1}^L e_{\ell}^2 \\ \text{s.t. } e_{\ell} &= r_{\ell} - \mathbf{m}_{\lambda_{\ell}}^{\top} \alpha - \psi_{\text{nl}}(\mathbf{m}_{\lambda_{\ell}}) \\ \alpha &\succeq \mathbf{0} \text{ and } \mathbf{1}^{\top} \alpha = 1, \end{aligned} \tag{6}$$

where μ is a positive parameter. The above problem minimizes the reconstruction error, as well as the regularity of ψ_{nl} and α is characterized by the squared ℓ_2 -norm with the ANC and ASC. This convex problem can be solved via its duality theory and leads to the so-called K-Hype algorithm. The ℓ -th element of this pixel is then reconstructed by

$$r_{\ell}^* = \mathbf{m}_{\lambda_{\ell}}^{\top} \alpha^* + \psi_{\text{nl}}^*(\mathbf{m}_{\lambda_{\ell}}). \tag{7}$$

where $*$ denotes the optimal estimates. K-Hype has been shown efficient in addressing several nonlinear mixture scenarios. However, it has the following restrictions:

- **R1:** While the ℓ_2 -regularization on ψ_{nl} controls its regularity, the ℓ_2 -regularization on α does not possess a clear interpretation, as there is no reason to consider that an abundance vector is having a small ℓ_2 -norm.
- **R2:** Problem (6) aims to estimate the abundances with the known endmember matrix. Considering to use a large spectral library as candidates, specific variants for extracting active endmembers are needed.

Facts **R1** and **R2** motivate us to derive an improved model and the associated unmixing algorithms by modifying the ℓ_2 -regularization and considering an endmember selection strategy.

3. Kernel-Based Nonlinear Unmixing Problem

In this section, we derive a new sparse kernel-based nonlinear unmixing algorithm with the data model (2). We shall first consider **R1** by removing the ℓ_2 -regularization of the abundances and devise the associated algorithm. Thereafter, we consider **R2** by proposing a sparse unmixing technique consisting of an ℓ_1 -regularization step and a dictionary pruning step.

3.1. Nonlinear Unmixing with the Regularity Constraint on ψ_{nl} .

Under the assumption that the endmember matrix M is known, we estimate the abundance vector α and to infer the nonlinear function $\psi_{\text{nl}} \in \mathcal{H}$ by solving the following functional optimization problem, that is,

$$\begin{aligned} \alpha^*, \psi_{\text{nl}}^* &= \arg \min_{\psi_{\text{nl}} \in \mathcal{H}} \frac{1}{2} \|\psi_{\text{nl}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L e_{\ell}^2 \\ \text{s.t. } e_{\ell} &= r_{\ell} - \mathbf{m}_{\lambda_{\ell}}^{\top} \alpha - \psi_{\text{nl}}(\mathbf{m}_{\lambda_{\ell}}) \\ \alpha &\succeq \mathbf{0} \text{ and } \mathbf{1}^{\top} \alpha = 1. \end{aligned} \tag{8}$$

In contrast to (6), the ℓ_2 -norm regularization of α is discarded, then the optimal α^* can not be explicitly expressed with the dual variables as in ([33] Equation (18)). In this case, resorting to the semi-parametric Representation Theorem [39] leads to the following expression for the nonlinear function ψ_{nl} :

$$\psi_{\text{nl}} = \sum_{\ell=1}^L \beta_{\ell} \kappa(\cdot, \mathbf{m}_{\lambda_{\ell}}). \tag{9}$$

By substituting (9) to (8), we get the following problem in a quadratic form with respect to parameters α and β :

$$\begin{aligned} \min_{\alpha, \beta} & \frac{1}{2} \beta^\top K \beta + \frac{1}{2\mu} \|r - M\alpha - K\beta\|_2^2 \\ & = \frac{1}{2} \begin{bmatrix} \beta \\ \alpha \end{bmatrix}^\top \begin{bmatrix} K + \frac{1}{\mu} K^\top K & \frac{1}{\mu} K^\top M \\ \frac{1}{\mu} M^\top K & \frac{1}{\mu} M^\top M \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \begin{bmatrix} -\frac{1}{\mu} K^\top r \\ -\frac{1}{\mu} M^\top r \end{bmatrix}^\top \begin{bmatrix} \beta \\ \alpha \end{bmatrix} \\ \text{s.t. } & \alpha \succeq \mathbf{0} \text{ and } \mathbf{1}^\top \alpha = 1, \end{aligned} \tag{10}$$

where K is a Gram matrix with (k, ℓ) -th entry being $\kappa(m_{\lambda_k}, m_{\lambda_\ell})$. Then the abundance vector can be evaluated by solving the problem (10) under the ASC and ANC via a quadratic programming solver.

3.2. Kernel-Based Sparse Nonlinear Unmixing with Dictionary Pruning

In this subsection, we consider a kernel-based model with pruning dictionary via a two-step method. It is possible to use a large spectral library as a dictionary, and then select a few elements as the endmembers that contribute to the mixture. This way also provides a solution to the simultaneous endmember extraction and abundance estimation for the nonlinear unmixing.

3.3. Problem Formulation

We first consider to add ℓ_1 -norm regularization to (8) and to promote the sparsity of estimated abundances. The optimization problem is then formulated as

$$\begin{aligned} \alpha^\dagger, \psi_{\text{nl}}^\dagger & = \arg \min_{\psi_{\text{nl}} \in \mathcal{H}} \frac{1}{2} \|\psi_{\text{nl}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L e_\ell^2 + \lambda \|\alpha\|_1 \\ \text{s.t. } & e_\ell = r_\ell - m_{\lambda_\ell}^\top \alpha - \psi_{\text{nl}}(m_{\lambda_\ell}) \\ & \alpha \succeq \mathbf{0}. \end{aligned} \tag{11}$$

Note that the ASC is ignored since it is contradictory to the ℓ_1 -norm regularization. In order to solve (11), we introduce an auxiliary variable ζ and reformulate the problem to an equivalent form:

$$\begin{aligned} \alpha^\dagger, \zeta^\dagger, \psi_{\text{nl}}^\dagger & = \arg \min_{\Psi_{\text{nl}} \in \mathcal{H}} \frac{1}{2} \|\psi_{\text{nl}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L e_\ell^2 + \lambda \|\zeta\|_1 + \mathcal{I}_{\mathbb{R}^+}(\zeta) \\ \text{s.t. } & e_\ell = r_\ell - m_{\lambda_\ell}^\top \alpha - \psi_{\text{nl}}(m_{\lambda_\ell}) \\ & \alpha = \zeta, \end{aligned} \tag{12}$$

where the $\mathcal{I}_{\mathbb{R}^+}(\cdot)$ is the indicator function such that $\mathcal{I}_{\mathbb{R}^+}(\cdot)$ is 0 if its argument belongs to the nonnegative orthant and $+\infty$ otherwise. This new variable ζ allows us to decouple the non-smooth ℓ_1 -norm regularization functional from the constrained problem. As studied in [40], the split-Bregman iteration algorithm is an efficient method to deal with a broad class of ℓ_1 -regularized problems. By applying this framework to (12), the following formulation is obtained:

$$\begin{aligned} & \alpha^{(k+1)}, \psi_{\text{nl}}^{(k+1)}, \zeta^{(k+1)} \\ & = \arg \min_{\alpha, \psi_{\text{nl}} \in \mathcal{H}, \zeta} \frac{1}{2} \|\psi_{\text{nl}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L [r_\ell - m_{\lambda_\ell}^\top \alpha - \psi_{\text{nl}}(m_{\lambda_\ell})]^2 + \lambda \|\zeta\|_1 + \mathcal{I}_{\mathbb{R}^+}(\zeta) + \frac{\rho}{2} \left\| \alpha - \zeta + \zeta^{(k)} \right\|_2^2 \end{aligned} \tag{13}$$

and

$$\zeta^{(k+1)} = \zeta^{(k)} + (\alpha^{(k+1)} - \zeta^{(k+1)}). \tag{14}$$

Because of the way that we have split the components of the cost function, we can now perform the above minimization efficiently by iteratively minimizing with respect to α and ψ_{nlm} , and ζ separately. The two steps that we have to perform are as follows:

1. Optimization with respect to α and ψ_{nlm} : Discarding irrelevant variables, the optimization problem (13) reduces to

$$\begin{aligned} \min_{\alpha, \psi_{\text{nlm}} \in \mathcal{H}} \quad & \frac{1}{2} \|\psi_{\text{nlm}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L e_{\ell}^2 + \frac{\rho}{2} \|\alpha - \zeta^{(k)} + \xi^{(k)}\|_2^2 \\ \text{s.t.} \quad & e_{\ell} = r_{\ell} - \mathbf{m}_{\lambda_{\ell}}^{\top} \alpha - \psi_{\text{nlm}}(\mathbf{m}_{\lambda_{\ell}}). \end{aligned} \tag{15}$$

By introducing the Lagrange multipliers $\{\beta\}_{\ell=1}^L$, the Lagrange function associated with the problem (15) can be written as

$$\mathcal{L}_1 = \frac{1}{2} \|\psi_{\text{nlm}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L e_{\ell}^2 + \frac{\rho}{2} \|\alpha - \zeta^{(k)} + \xi^{(k)}\|_2^2 + \sum_{\ell=1}^L \beta_{\ell} (r_{\ell} - \mathbf{m}_{\lambda_{\ell}}^{\top} \alpha - \psi_{\text{nlm}}(\mathbf{m}_{\lambda_{\ell}}) - e_{\ell}). \tag{16}$$

The conditions for optimality of \mathcal{L}_1 with respect to the primal variables are given by

$$\begin{cases} \alpha = \frac{1}{\rho} \mathbf{M}^{\top} \beta + \zeta^{(k)} - \xi^{(k)} \\ \psi_n = \sum_{\ell=1}^L \beta_{\ell} \kappa(\cdot, \mathbf{m}_{\lambda_{\ell}}) \\ e_{\ell} = \mu \beta_{\ell}, \text{ for } \ell = 1, \dots, L, \end{cases} \tag{17}$$

where $\beta = [\beta_1, \dots, \beta_L]^{\top}$. By substituting (17) into (16), this results in a quadratic form with respect to the Lagrange multiplier and we get the following dual problem:

$$\max_{\beta} -\frac{1}{2} \beta^{\top} \left(\mathbf{K} + \mu \mathbf{I} + \frac{1}{\rho} \mathbf{M} \mathbf{M}^{\top} \right) \beta + \left(\frac{1}{\rho} \mathbf{M} \zeta^{(k)} - \mathbf{M} \xi^{(k)} + \mathbf{r} \right)^{\top} \beta. \tag{18}$$

This is a quadratic program and the updated value $\beta^{(k+1)}$ is readily obtained by

$$\beta^{(k+1)} = \left(\mathbf{K} + \mu \mathbf{I} + \frac{1}{\rho} \mathbf{M} \mathbf{M}^{\top} \right)^{-1} \times \left(\frac{1}{\rho} \mathbf{M} \zeta^{(k)} - \mathbf{M} \xi^{(k)} + \mathbf{r} \right). \tag{19}$$

Clearly, matrix $\mathbf{K} + \mu \mathbf{I} + \frac{1}{\rho} \mathbf{M} \mathbf{M}^{\top}$ is positive-definite and invertible.

2. Optimization with respect to ζ : Discarding irrelevant variables, the problem reduced to

$$\min_{\zeta} \lambda \|\zeta\|_1 + \mathcal{I}_{\mathbb{R}^+}(\zeta) + \frac{\rho}{2} \|\alpha^{(k+1)} - \zeta + \xi^{(k)}\|_2^2. \tag{20}$$

Its solution can be expressed by the well-known soft threshold function

$$\zeta^{(k+1)} = \mathcal{S}(\alpha^{(k+1)} + \xi^{(k)})^+ \tag{21}$$

where $\mathcal{S}(\cdot, \cdot)$ denotes the component-wise application of the soft threshold function, given by

$$\mathcal{S}(x, y) = \text{sign}(x) \max(|x| - y, 0), \tag{22}$$

and $(\cdot)^+$ projects the argument to the nonnegative orthant by setting a negative value to 0.

These iterations are repeated until convergence. The stopping criteria as in [41] can be used where the primal and dual residuals must be smaller than some tolerance thresholds. We denote the optimal solution obtained from this step by α^{\dagger} , as this quantity will be used in the later part.

3.4. Dictionary Pruning

By solving the above problem, the abundance vector is obtained from a relatively large original dictionary. We then propose to refine the estimation and prune the large spectral library to a smaller one by discarding the endmembers with zero fractions. Therefore, in this second step, we reformulate the optimization problem using this pruned dictionary as follows:

$$\begin{aligned}
 &\alpha^*, \psi_{\text{nlín}}^* = \\
 &\underset{\psi_{\text{nlín}} \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{2} \|\psi_{\text{nlín}}\|_{\mathcal{H}}^2 + \frac{1}{2\mu} \sum_{\ell=1}^L [r_{\ell} - \mathbf{x}^{\top} (\mathbf{m}_{\lambda_{\ell}} \odot \operatorname{supp}(\alpha^{\dagger})) - \psi_{\text{nlín}}(\mathbf{m}_{\lambda_{\ell}} \odot \operatorname{supp}(\alpha^{\dagger}))]^2 + \lambda \|\alpha\|_1 \quad (23) \\
 &\text{s.t. } \alpha \succeq \mathbf{0},
 \end{aligned}$$

where $\operatorname{supp}(\alpha^{\dagger}) \in \{0, 1\}$ is a vector that denotes the support of the abundance estimated by the first step and \odot denotes the Hadamard product. The Hadamard product between the endmember information and the support set discards the inactive ones defined in α^{\dagger} . Note that a large number of endmembers interact in $\psi_{\text{nlín}}$ with the model in Section 3.3, and only a small active part is involved after pruning the dictionary, which is reasonable for providing a better model. Problem (23) can be approached in a similar way as in Section 3.3.

4. Experiments

In this section, experiments are conducted to validate the proposed unmixing algorithms and compare them with several state-of-the-art methods. The algorithms are tested on synthetic data, real data captured created in our laboratory with ground-truth, and the well-known Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite image over the Cuprite mining region in NV, USA. Note that testing unmixing algorithms using real data with ground-truth would produce more meaningful results.

4.1. Experiments with Synthetic Data

4.1.1. Simulation Settings

Two synthetic datasets are generated for the experiments. In the first scene (denoted by \mathcal{S}_1), a large spectral library is considered, since the sparse unmixing algorithm aims to select proper endmembers from this library. A pruned-version endmember library [42] is used with 342 candidates whose reflectance values are measured for 224 spectral bands distributed uniformly in the interval 0.4–2.5 μm . A total of 1000 pixels are generated for each mixture setting. In each pixel, three endmembers in the library are randomly activated, with their associated abundance vectors being generated in the simplex defined by non-negative and sum-to-one constraints.

In the second scene (denoted by \mathcal{S}_2), nine endmembers are selected from the above library to generate a 50-by-50 image with $N = 2500$ pixels, using the upper-left region of ([11] Figure 2) for the associated abundances. These mixtures in this scene involved only nine endmembers, however, very few endmembers are activated in each pixel as shown in Figure 1.

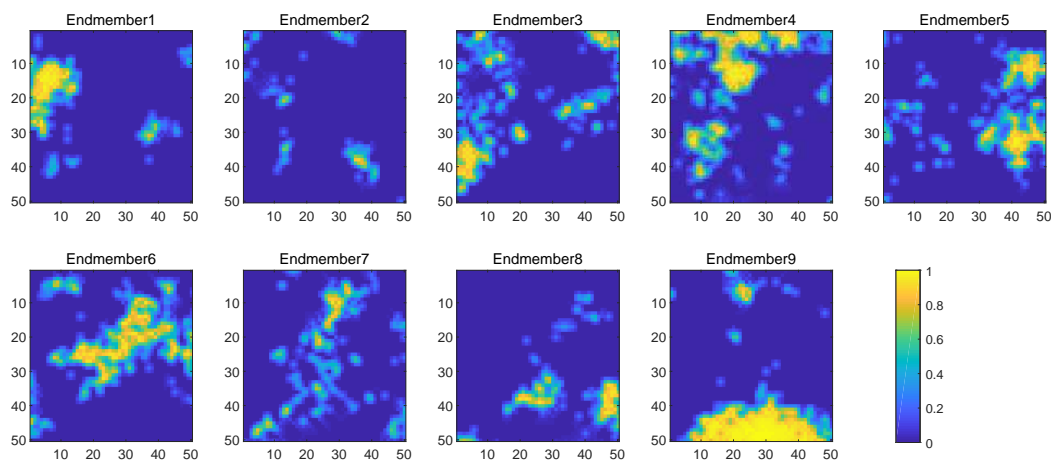


Figure 1. Ground-truth abundance maps associated to nine endmembers used in \mathcal{S}_2 .

Both of these two scenes are generated with the general bilinear model (GBM) defined by:

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \sum_{i=1}^{R-1} \sum_{j=i+1}^R \gamma_{i,j} \alpha_i \alpha_j \mathbf{m}_i \odot \mathbf{m}_j + \mathbf{n},$$

and the PNMM defined by

$$\mathbf{r} = (\mathbf{M}\mathbf{x})^\tau + \mathbf{n}. \quad (24)$$

In these experiments, the parameter $\gamma_{i,j}$ is simply set to 1 and τ is set to 0.7, same as [33]. Finally, all data are corrupted with an additive white Gaussian noise with two levels of SNR, namely, 30 dB and 15 dB respectively.

4.1.2. Comparative Simulations

Several algorithms are conducted in order to compare their unmixing performance. Their tuning parameters are set during preliminary experiments on independent data, via a simple search over the grids defined hereafter.

- The FCLS algorithm [7]: This classical algorithm relies on the linear model and seeks the optimal solution of a least-square problem subject to the ASC and ANC.
- The sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) [10]: This algorithm fits the observed (mixed) hyperspectral vectors with sparse linear mixtures of spectral signatures from a large dictionary available a priori. The parameter λ_{SUnSAL} controls the trade-off between the fitting and sparsity.
- The K-Hype algorithm [33]: This algorithm is based on the scalar RKHS model. The parameter $\mu_{\text{K-Hype}}$ controls the trade-off between the fitting and the functional regularity.
- The nonlinear neighbor and band dependent unmixing (NDU) [38]: This method extends the K-Hype by accounting for band-dependent and neighboring nonlinear contributions with vector-valued kernel function. The separable (Sp.) structure is considered due to the high computational complexity. This algorithm is characterized by the trade-off parameter λ_{NDU} , fitting parameter μ_{NDU} and penalty parameter ρ_{NDU} .
- The improved incremental KNMF (IKNMF) [31]: This method extends KNMF by introducing partition matrix theory and considering the relationships among dividing blocks. The incremental KNMF (IKNMF) is proposed to reduce the computing requirements for large-scale data and IKNMF aims to further improve the abundance results. The parameter k controls the number of blocks. For each experiment of IKNMF, 5 independent runs are carried out and the results are averaged.

The root mean square error between the true values $\boldsymbol{\alpha}_n$ and the estimates $\hat{\boldsymbol{\alpha}}_n$ defined by

$$RMSE = \sqrt{\frac{1}{NR} \sum_{n=1}^N \|\boldsymbol{\alpha}_n - \hat{\boldsymbol{\alpha}}_n\|^2} \quad (25)$$

is used to evaluate the performance of the abundance estimation. All the kernel-based methods are tested with a Gaussian (\mathcal{G}) given by

$$\kappa(\mathbf{m}_{\lambda_\ell}, \mathbf{m}_{\lambda_p}) = \exp\left(-\frac{\|\mathbf{m}_{\lambda_\ell} - \mathbf{m}_{\lambda_p}\|^2}{2\sigma^2}\right), \quad (26)$$

and a second order homogeneous polynomial (\mathcal{P}) kernel given by

$$\kappa(\mathbf{m}_{\lambda_\ell}, \mathbf{m}_{\lambda_p}) = (\mathbf{m}_{\lambda_\ell}^\top \mathbf{m}_{\lambda_p})^2, \quad (27)$$

respectively. The bandwidth σ of the Gaussian kernel in the algorithms K-Hype, NDU, IKNMF and the proposed algorithm is varied within $\{1, 2, 3, \dots, 10\}$. In these algorithms, the tuning parameters λ_{SUnSAL} , λ_{NDC} and λ of the sparse regularization are tested in the range $\{1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005\}$, and parameters μ , $\mu_{\text{K-Hype}}$ and μ_{NDU} for the fitting term are tested in the range $\{1000, 500, 100, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1, 0.05, 0.01, 0.005\}$. The penalty parameter ρ_{NDU} in NDU and ρ in the proposed method are both set to 1. All the parameters used in the experiments of \mathcal{S}_1 and \mathcal{S}_2 are reported in Appendix Section (Tables A1 and A2 respectively). In the experiment with \mathcal{S}_1 , the results of IKNMF are not given, since its computational time is beyond acceptance with a very large endmember library. Besides selecting the Gaussian kernel parameter from the candidate values, we also use an automatic parameter setting strategy, where we set $\sigma = \max_{\ell,k} \{d(\mathbf{m}_{\lambda_\ell}, \mathbf{m}_{\lambda_k})\}$, so that the kernel width covers the variability range of the input. Associated performance of this setting is denoted by “(auto)” in the results.

4.1.3. Results

The obtained RMSEs for \mathcal{S}_1 and \mathcal{S}_2 are reported in Tables 1 and 2. Generally, under the low SNR (15 dB), the compared algorithms have similar performance. Meanwhile all of the algorithms show an increased accuracy as the SNR increases. It is observed that NDU performs slightly better than our proposed algorithm in some settings of \mathcal{S}_1 , however its computational complexity is several times higher. In \mathcal{S}_2 , it is observed that the proposed algorithm outperforms the competing methods in most cases, and even in the low SNR setting, the proposed algorithm significantly outperforms the others. We also observe that the automatic kernel parameter setting strategy leads to results that are comparable to those from the pre-selected setting. The synthetic data can be restrictive to provide a proper performance evaluation due to their inherent limitation. In the next section, we shall validate the advantage of the proposed algorithm with labeled real data.

Table 1. RMSE comparison of \mathcal{S}_1 .

	SNR = 15 dB		SNR = 30 dB	
	GBM	PNMM	GBM	PNMM
FCLS	0.0296 ± 0.00055	0.0319 ± 0.00052	0.0224 ± 0.00046	0.0235 ± 0.00033
SUnSAL	0.0283 ± 0.00054	0.0321 ± 0.00054	0.0171 ± 0.00034	0.0214 ± 0.00034
K-Hype (\mathcal{G})	0.0295 ± 0.00037	0.0319 ± 0.00036	0.0214 ± 0.00040	0.0244 ± 0.00032
K-Hype (\mathcal{P})	0.0292 ± 0.00038	0.0322 ± 0.00036	0.0213 ± 0.00038	0.0254 ± 0.00032
NDU (\mathcal{G})	0.0278 ± 0.00054	0.0311 ± 0.00050	0.0153 ± 0.00030	0.0192 ± 0.00028
NDU (\mathcal{P})	0.0277 ± 0.00054	0.0309 ± 0.00052	0.0150 ± 0.00031	0.0193 ± 0.00028
Proposed (\mathcal{G})	0.0279 ± 0.00051	0.0315 ± 0.00051	0.0152 ± 0.00029	0.0210 ± 0.00032
Proposed (auto) (\mathcal{G})	0.0291 ± 0.00054	0.0315 ± 0.00051	0.0155 ± 0.00030	0.0210 ± 0.00032
Proposed (\mathcal{P})	0.0278 ± 0.00052	0.0315 ± 0.00051	0.0150 ± 0.00032	0.0210 ± 0.00032

Table 2. RMSE comparison of \mathcal{S}_2 .

	SNR = 15 dB		SNR = 30 dB	
	GBM	PNMM	GBM	PNMM
FCLS	0.0564 ± 0.0031	0.1185 ± 0.0082	0.0488 ± 0.0027	0.1139 ± 0.0070
SUnSAL	0.0523 ± 0.0035	0.0839 ± 0.0060	0.0288 ± 0.0008	0.0561 ± 0.0016
K-Hype (\mathcal{G})	0.0482 ± 0.0029	0.0726 ± 0.0051	0.0161 ± 0.0003	0.0338 ± 0.0011
K-Hype (\mathcal{P})	0.0945 ± 0.0070	0.1346 ± 0.0107	0.0533 ± 0.0026	0.1218 ± 0.0078
NDU (\mathcal{G})	0.0551 ± 0.0030	0.1107 ± 0.0090	0.0447 ± 0.0020	0.0888 ± 0.0070
NDU (\mathcal{P})	0.0524 ± 0.0025	0.1127 ± 0.0093	0.0355 ± 0.0011	0.0825 ± 0.0068
IIKNMF (\mathcal{G})	0.2396 ± 0.0488	0.2290 ± 0.0364	0.2418 ± 0.0472	0.2360 ± 0.0439
IIKNMF (\mathcal{P})	0.2155 ± 0.0324	0.2066 ± 0.0253	0.2241 ± 0.0349	0.2107 ± 0.0342
Proposed (\mathcal{G})	0.0529 ± 0.0039	0.0662 ± 0.0052	0.0278 ± 0.0008	0.0269 ± 0.0005
Proposed (auto) (\mathcal{G})	0.0536 ± 0.0037	0.0665 ± 0.0053	0.0233 ± 0.0005	0.0271 ± 0.0006
Proposed (\mathcal{P})	0.0515 ± 0.0037	0.0867 ± 0.0061	0.0159 ± 0.0003	0.0604 ± 0.0011

4.1.4. Sensitivity Analysis

We analyze the sensitivity of the proposed method with respect to the algorithm parameters λ and μ . This test is conducted with \mathcal{S}_1 using both Gaussian and Polynomial kernels under SNR of 30 dB. The results are shown in Figures 2 and 3. As can be seen within a reasonable range around the optimal parameter values, the algorithm exhibits satisfactory RMSE.

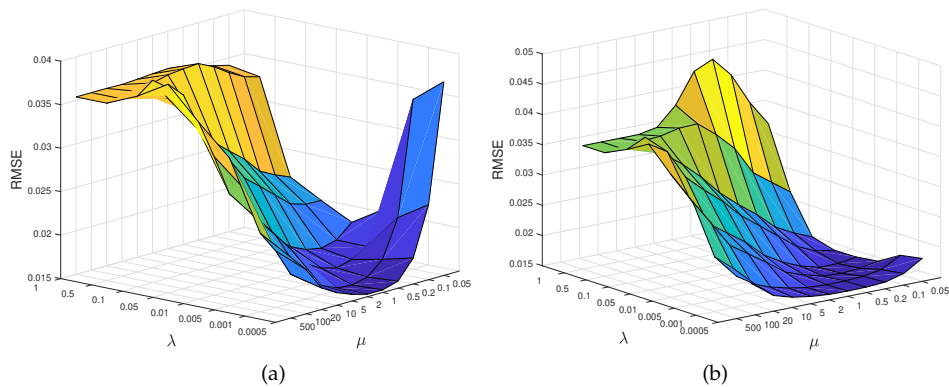


Figure 2. RMSE as a function of the regularization parameters for the proposed method with S_1 using Gaussian kernel.

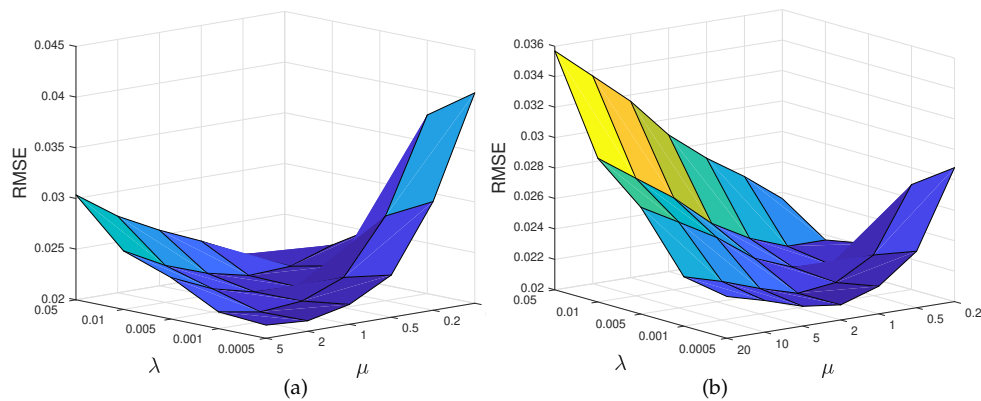


Figure 3. RMSE as a function of the regularization parameters for the proposed method with S_1 using Polynomial kernel.

4.2. Experiments with the Laboratory Created Real Data

4.2.1. Data Presentation

In order to perform quantitative evaluation of unmixing performance, we designed several experimental scenes in our laboratory with the GaiaField (Sichuan Dualix Spectral Image Technology CO. Ltd., GaiaField-V10) and GaiaSorter systems. Our GaiaField is a push-boom imaging spectrometer with a HSIA-OL50 lens, covering the visible and NIR wavelengths ranging from 400 nm to 1000 nm. GaiaSorter sets an environment that isolates external lights, and is endowed with a conveyor to move samples for the push-boom imaging. A dataset is created by imaging these scenes with the hyperspectral camera in our laboratory, providing mixtures with 256 wavelength bands. The experimental settings are strictly controlled so that pure material spectral signatures and material compositions are known. See [43,44] for experiment and data description in detail.

In this subsection, uniform mixture #1, mixture #7 and two non-uniform mixtures cases of Scene-II in [44] are used (In [44], to get the ground truth, the aligned high-resolution RGB images of S_5 and S_6 were captured, and then the percentage of each colored sand in a low-resolution hyperspectral pixel could be analyzed with the help of the associated RGB image). In this scene, quartz sand of different colors are mixed, and the diameter of the granules is controlled to be approximately 0.03 mm (see Figure 4). Seven endmembers (four different pure colored quartz sand and three other materials in the dataset) are included as candidate spectral signatures, shown in Figure 5. The true abundances are provided either by priorly known fractions (uniform mixtures) or analysis of aligned high-resolution images (mixtures with spatial patterns). One uniform mixture is formed by mixing 50% blue sand and 50% white sand so that the true abundance vector $\alpha = [0.5, 0.5, 0, 0, 0, 0, 0]^T$ (denoted by S_3). The other uniform mixture is formed by mixing the blue, red and green sand so that the true abundance vector $\alpha = [1/3, 0, 1/3, 1/3, 0, 0, 0]^T$ (denoted by S_4). S_5 and S_6 respectively denote the two mixtures with spatial patterns formed by mixing two and three types of sand. As this scene mimics the intimate mixture model, the Gaussian kernel is used for this high-order nonlinear case.

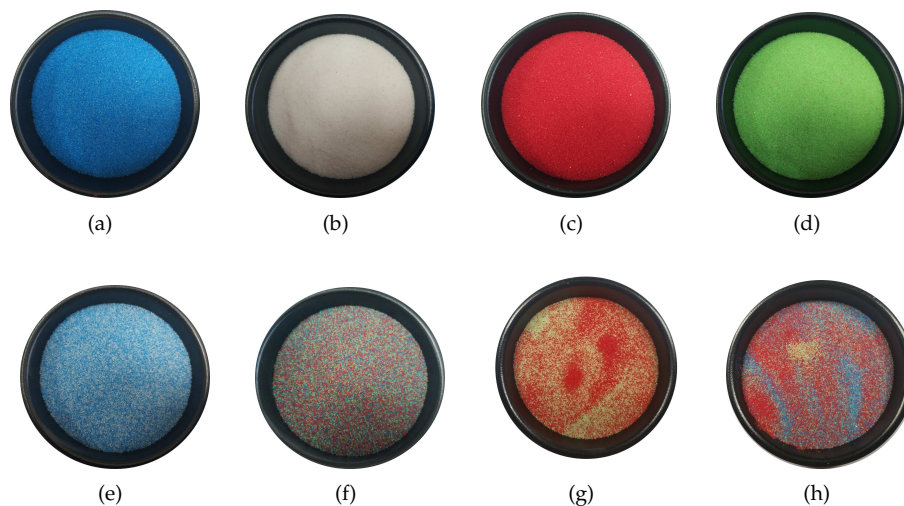


Figure 4. Laboratory created data for unmixing performance evaluation (RGB images). Subfigures (a–d): pure quartz sand of 0.03 mm of four colors, and they serve as pure materials for providing endmembers. Subfigures (e,f): uniform mixtures of sand of two (e) and three colors (f). Subfigures (g,h): mixtures with spatial patterns of sand of two (g) and three colors (h). Square regions of 60-by-60 pixels in the center of each subfigures are clipped out and used in experiments.

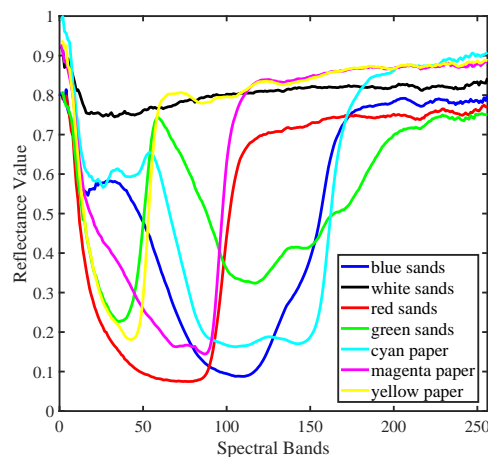


Figure 5. Spectral signatures of seven materials extracted from the dataset. They form the endmember matrix that is used in Section 4.2.

4.2.2. Results

The obtained RMSEs are reported in Table 3, with the algorithm parameters listed in Table A3. We observe that the proposed algorithm achieves the lowest RMSEs, and the proposed algorithm with automatically set bandwidth also leads to satisfactory performance. For data \mathcal{S}_5 and \mathcal{S}_6 , abundance maps estimated by the compared algorithms are shown in Figures 6 and 7. We see that most algorithms recovered the spatial pattern of the abundance maps. However, the K-Hype algorithm produces a low contrast map for the first component in \mathcal{S}_5 . The proposed algorithm results in sharper abundance maps with lowest RMSEs. Experiments with these labeled real data show the benefit of the proposed algorithm.

Table 3. RMSE comparison of $\mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5, \mathcal{S}_6$.

Model	LMM				NLMM			
	Algorithm	FCLS	SUnSAL	K-Hype (\mathcal{G})	NDU (\mathcal{G})	IKNMF (\mathcal{G})	Proposed (\mathcal{G})	Proposed (auto) (\mathcal{G})
RMSE	\mathcal{S}_3	0.1192 ± 0.0054	0.1410 ± 0.0072	0.1445 ± 0.0053	0.1235 ± 0.0093	0.2213 ± 0.0646	0.1124 ± 0.0044	0.1138 ± 0.0045
	\mathcal{S}_4	0.0891 ± 0.0041	0.1020 ± 0.0049	0.0662 ± 0.0042	0.1032 ± 0.0112	0.1709 ± 0.0184	0.0573 ± 0.0025	0.0585 ± 0.0025
	\mathcal{S}_5	0.1619 ± 0.0279	0.1733 ± 0.0318	0.1802 ± 0.0229	0.1617 ± 0.0280	0.2502 ± 0.0753	0.1576 ± 0.0275	0.1585 ± 0.0283
	\mathcal{S}_6	0.1799 ± 0.0250	0.1817 ± 0.0252	0.1895 ± 0.0249	0.1789 ± 0.0249	0.2398 ± 0.0598	0.1699 ± 0.0242	0.1703 ± 0.0244

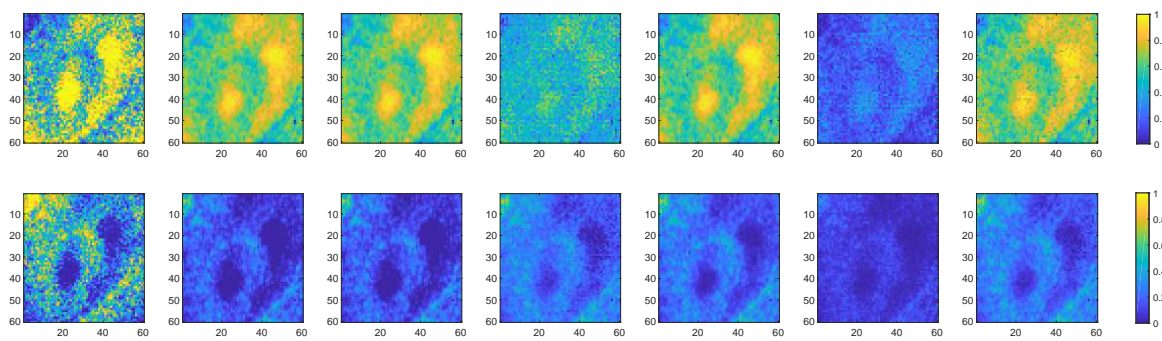


Figure 6. Abundances maps of S_5 . From left to right columns: ground-truth, estimated abundances of FCLS, SUnSAL, K-Hype (\mathcal{G}), NDU (\mathcal{G}), IKNMF (\mathcal{G}), proposed (\mathcal{G}). From top to bottom rows: distributions of red quartz sand and green quartz sand.

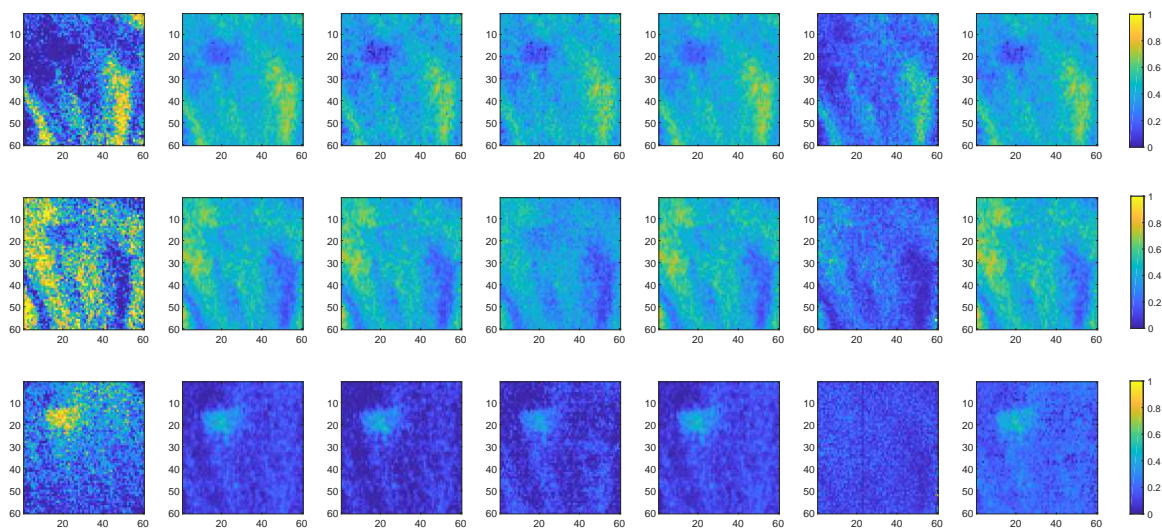


Figure 7. Abundances maps of S_6 . From left to columns: ground-truth, estimated abundances of FCLS, SUnSAL, K-Hype (\mathcal{G}), NDU (\mathcal{G}), IKNMF (\mathcal{G}), proposed (\mathcal{G}). From top to bottom rows: distributions of blue quartz sand, red quartz sand and green quartz sand.

4.3. Experiment with AVIRIS Cuprite Data

This section illustrates the performance of the proposed algorithms and other algorithms when applied to real remote sensing hyperspectral data. The scene that is used for our experiment is the well-known image captured on the Cuprite mining district (NV, USA) by AVIRIS. A sub-image of 250×191 pixels is chosen to evaluate the algorithms. This area of interest has 188 spectral bands. This scene is denoted by S_7 . The same twelve spectra signatures as in [33] are used for the endmember matrix, as shown in Figure 8. Results from most existing works illustrate that only few materials are active in each pixel, thus sparse unmixing is potentially beneficial for analyzing this scene.

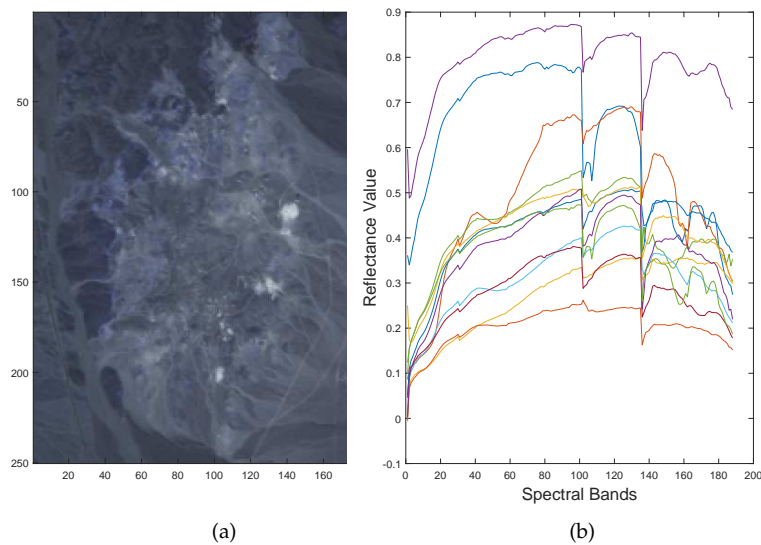


Figure 8. AVIRIS Cuprite data (denoted by S_7) used in the experiment. (a) The false color image illustrated with bands $\{30, 40, 80\}$ of the sub-image of the AVIRIS Cuprite data set. (b) Twelve endmembers used for unmixing.

Note that this real data is extensively used in the literature, however no ground-truth information is available for an objective performance evaluation. The quality of these algorithms can be defined by the Reconstruction Error (RE) which is defined as

$$RE = \sqrt{\frac{1}{NL} \sum_{n=1}^N \|r_n - \hat{r}_n\|^2}. \quad (28)$$

In general, RE can be used to provide a quantitative information of the unmixing performance. However, as pointed in [33], without the ground-truth information on the abundances for the real data, the reconstruction quality measured by RE is not necessarily proportional to the quality of the abundance estimation, and therefore it can only be considered as complementary information. In this experiment, the selected parameters are given in Table A4 and the RE comparison by the linear and nonlinear models is reported in Table 4. Figure 9 illustrates the maps of the reconstruction error. These tables and figures clearly indicate that nonlinear unmixing algorithms lead to significantly lower REs compared to linear algorithms, especially at several particular spatial regions. The proposed algorithm exhibits a comparable RE compared to K-Hype with Gaussian kernel, recalling that it is common that a regularization term increases the data fitting cost. Figure 10 shows the abundance maps of selected materials estimated by these algorithms. The proposed algorithm provides a sharper map with several particular locations emphasized.

Table 4. RE comparison of S_7 .

Model	LMM			NLMM						
	FCLS	SUnSAL	K-Hype (\mathcal{G})	NDU(\mathcal{G})	IKNMF (\mathcal{G})	Proposed (\mathcal{G})	K-Hype (\mathcal{P})	NDU(\mathcal{P})	IKNMF (\mathcal{P})	Proposed (\mathcal{P})
Algorithm										
RE	0.0056	0.0050	0.0027	0.0282	0.0164	0.0028	0.0021	0.0282	0.0519	0.0034

5. Conclusions

In this work a new nonlinear sparse unmixing method is proposed. Under the assumption that the model consists of a linear mixture and a nonlinear fluctuation defined in an RHKS, the unmixing is conducted by the kernel-based method with a sparsity promoting term and the dictionary pruning technique. Experiments with both synthetic and real data were conducted to illustrate the performance of the proposed algorithm and compared algorithms. The proposed method outperformed the competing ones in most cases. Particularly, quantitative abundance estimation errors were reported with laboratory created real data, and highlighted the advantage of the proposed algorithm. Although the NDU performed slightly better in one synthetic data setting, it had higher computation complexity. In future work the automatic kernel parameter learning can be considered to increase the flexibility of the algorithm.

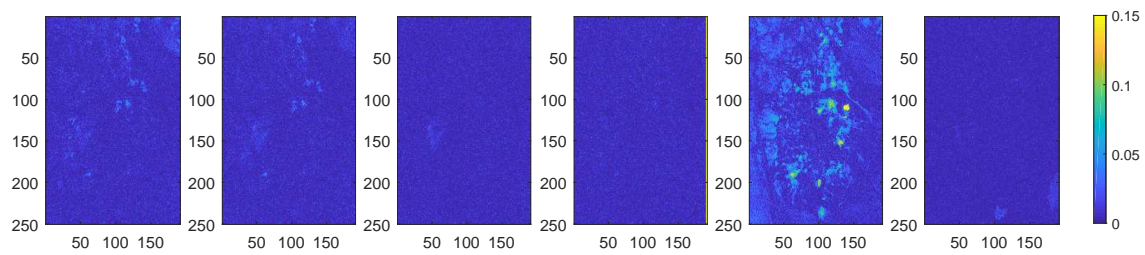


Figure 9. Maps of reconstruction error. From left to right: FCLS, SUnSAL, K-Hype (\mathcal{G}), NDU (\mathcal{G}), IIKNMF (\mathcal{G}), Proposed (\mathcal{G}).

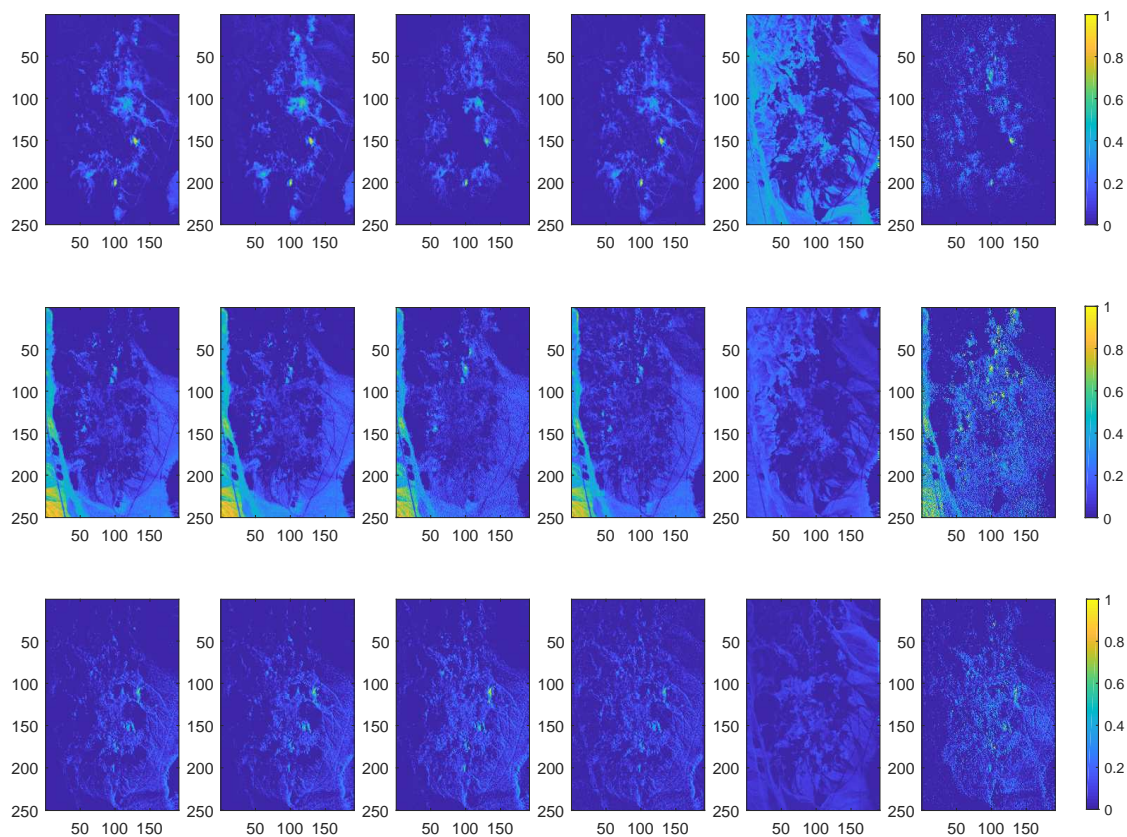


Figure 10. Abundances maps of selected materials of S_7 . From left to right: FCLS, SUnSAL, K-Hype (\mathcal{G}), NDU (\mathcal{G}), IIKNMF (\mathcal{G}), Proposed (\mathcal{G}). From top to bottom: the 1st endmember, the 6th endmember, the 10th endmember.

Author Contributions: Conceptualization, J.C.; methodology, Z.L.; software, Z.L.; formal analysis, Z.L.; investigation, J.C. and Z.L.; supervision, J.C. and S.R.; writing—original draft preparation, Z.L.; writing—review and editing, J.C. and S.R.

Funding: The work of J. Chen was supported in part by the NSFC under grant 61671382, Natural Science Foundation of Shenzhen under grant JCYJ2017030155315873, and 111 project (B18041).

Acknowledgments: The authors would like to thank Risheng Huang for providing MATLAB codes of IIKNMF used in our experiments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The optimal parameters used in the experiments are provided in Tables A1 to A4.

Table A1. The selected algorithm parameters for \mathcal{S}_1 .

	SNR = 15 dB		SNR = 30 dB	
	GBM	PNMM	GBM	PNMM
SUnSAL	$\lambda = 0.1$	$\lambda = 0.05$	$\lambda = 0.005$	$\lambda = 0.005$
K-Hype (\mathcal{G})	$\mu = 0.2, \sigma = 10$	$\mu = 0.2, \sigma = 10$	$\mu = 0.005, \sigma = 10$	$\mu = 0.01, \sigma = 10$
K-Hype (\mathcal{P})	$\mu = 0.2$	$\mu = 0.2$	$\mu = 0.01$	$\mu = 0.01$
NDU (\mathcal{G})	$\lambda = 0.1, \mu = 100, \sigma = 10$	$\lambda = 0.1, \mu = 1, \sigma = 10$	$\lambda = 0.01, \mu = 0.1, \sigma = 10$	$\lambda = 0.01, \mu = 0.2, \sigma = 10$
NDU (\mathcal{P})	$\lambda = 0.1, \mu = 100$	$\lambda = 0.1, \mu = 1$	$\lambda = 0.01, \mu = 0.05$	$\lambda = 0.01, \mu = 0.1$
Proposed (\mathcal{G})	$\lambda = 0.005, \mu = 20, \sigma = 2$	$\lambda = 0.001, \mu = 100, \sigma = 10$	$\lambda = 0.001, \mu = 1, \sigma = 2$	$\lambda = 0.001, \mu = 2, \sigma = 10$
Proposed (auto) (\mathcal{G})	$\lambda = 0.5, \mu = 5, \sigma = 6.62$	$\lambda = 0.001, \mu = 100, \sigma = 6.62$	$\lambda = 0.001, \mu = 1, \sigma = 6.62$	$\lambda = 0.001, \mu = 2, \sigma = 6.62$
Proposed (\mathcal{P})	$\lambda = 0.005, \mu = 10$	$\lambda = 0.001, \mu = 100$	$\lambda = 0.01, \mu = 0.2$	$\lambda = 0.001, \mu = 2$

Table A2. The selected algorithm parameters for \mathcal{S}_2 .

	SNR = 15 dB		SNR = 30 dB	
	GBM	PNMM	GBM	PNMM
SUnSAL	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.0005$	$\lambda = 0.05$
K-Hype (\mathcal{G})	$\mu = 0.01, \sigma = 10$	$\mu = 0.01, \sigma = 10$	$\mu = 0.005, \sigma = 10$	$\mu = 0.005, \sigma = 10$
K-Hype (\mathcal{P})	$\mu = 0.1$	$\mu = 0.1$	$\mu = 0.005$	$\mu = 0.2$
NDU (\mathcal{G})	$\lambda = 0.1, \mu = 100, \sigma = 1$	$\lambda = 0.001, \mu = 0.001, \sigma = 10$	$\lambda = 0.005, \mu = 0.005, \sigma = 2$	$\lambda = 0.005, \mu = 0.005, \sigma = 10$
NDU (\mathcal{P})	$\lambda = 0.1, \mu = 10$	$\lambda = 0.05, \mu = 0.2$	$\lambda = 0.1, \mu = 1$	$\lambda = 0.01, \mu = 0.01$
IKNMF (\mathcal{G})	$k = 100, \sigma = 10$	$k = 100, \sigma = 10$	$k = 100, \sigma = 10$	$k = 100, \sigma = 10$
IKNMF (\mathcal{P})	$k = 100$	$k = 100$	$k = 100$	$k = 100$
Proposed (\mathcal{G})	$\lambda = 0.05, \mu = 1, \sigma = 1$	$\lambda = 0.05, \mu = 2, \sigma = 2$	$\lambda = 0.1, \mu = 0.1, \sigma = 2$	$\lambda = 0.05, \mu = 0.1, \sigma = 1$
Proposed (auto) (\mathcal{G})	$\lambda = 0.05, \mu = 1, \sigma = 1.27$	$\lambda = 0.05, \mu = 2, \sigma = 1.27$	$\lambda = 0.1, \mu = 0.1, \sigma = 1.27$	$\lambda = 0.05, \mu = 0.2, \sigma = 1.27$
Proposed (\mathcal{P})	$\lambda = 0.05, \mu = 1$	$\lambda = 0.1, \mu = 1$	$\lambda = 0.05, \mu = 0.2$	$\lambda = 0.05, \mu = 1$

Table A3. The parameters in algorithms of $\mathcal{S}_3, \mathcal{S}_4, \mathcal{S}_5, \mathcal{S}_6$.

	SUnSAL	K-Hype	NDU	IKNMF	Proposed	Proposed (auto)
\mathcal{S}_3	$\lambda = 1$	$\mu = 1, \delta = 10$	$\mu = 200, \lambda = 0.001, \delta = 10$	$k = 200, \delta = 10$	$\mu = 500, \lambda = 0.01, \delta = 1$	$\mu = 500, \lambda = 0.01, \delta = 1.54$
\mathcal{S}_4	$\lambda = 0.001$	$\mu = 0.05, \delta = 10$	$\mu = 1000, \lambda = 0.001, \delta = 1$	$k = 200, \delta = 10$	$\mu = 10, \lambda = 0.001, \delta = 1$	$\mu = 10, \lambda = 0.001, \delta = 1.54$
\mathcal{S}_5	$\lambda = 0.001$	$\mu = 0.05, \delta = 10$	$\mu = 200, \lambda = 0.001, \delta = 10$	$k = 200, \delta = 10$	$\mu = 2, \lambda = 0.001, \delta = 10$	$\mu = 5, \lambda = 0.001, \delta = 1.54$
\mathcal{S}_6	$\lambda = 0.01$	$\mu = 0.2, \delta = 10$	$\mu = 1000, \lambda = 0.001, \delta = 10$	$k = 200, \delta = 10$	$\mu = 10, \lambda = 0.001, \delta = 10$	$\mu = 10, \lambda = 0.001, \delta = 1.54$

Table A4. Parameters used in algorithms of \mathcal{S}_7 .

	SUnSAL	K-Hype	NDU	IKNMF	Proposed
LMM	$\lambda = 0.001$	—	—	—	—
\mathcal{G}	—	$\mu = 0.0001, \sigma = 10$	$\mu = 0.01, \sigma = 2, \lambda = 0.001$	$k = 400, \sigma = 2$	$\mu = 0.0001, \sigma = 1, \lambda = 0.001$
\mathcal{P}	—	$\mu = 0.0001$	$\mu = 0.01, \lambda = 0.001$	$k = 400$	$\mu = 0.001, \lambda = 0.001$

References

1. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57. [CrossRef]
2. Boardman, J.W. Automating spectral unmixing of aviris data using convex geometry concepts. In Proceedings of the Fourth Annual JPL Airborne Geoscience Workshop, Arlington, VA, USA, 25–29 October 1993.
3. Nascimento, J.M.P.; Bioucas-Dias, J.M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]
4. Winter, M.E. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. *Proc. SPIE Image Spectrom.* **1999**, *3753*, 266–275.
5. Li, J.; Bioucas-Dias, J.M. Minimum Volume Simplex Analysis: A Fast Algorithm to Unmix Hyperspectral Data. In Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 8–11 July 2008; Volume 3753, pp. 250–253.
6. Miao, L.; Qi, H. Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 765–777. [CrossRef]
7. Heinz, D.C.; Chang, C. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2002**, *39*, 529–545. [CrossRef]
8. Dobigeon, N.; Tournet, J.Y.; Chang, C.I. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. *IEEE Trans. Signal Process.* **2008**, *56*, 2684–2695. [CrossRef]

9. Moussaoui, S.; Schmidt, F.; Jutten, C.; Chanussot, J.; Brie, D.; Benediktsson, J.A. On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation. *Neurocomputing* **2008**, *71*, 2194–2208. [[CrossRef](#)]
10. Bioucas-Dias, J.M.; Figueiredo, M.A.T. Alternating Direction Algorithms for Constrained Sparse Regression: Application to Hyperspectral Unmixing. In Proceedings of the 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Reykjavik, Iceland, 14–16 June 2010; pp. 1–4.
11. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4484–4502. [[CrossRef](#)]
12. Xia, W.; Liu, X.; Wang, B.; Zhang, L. Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2165–2179. [[CrossRef](#)]
13. Jia, S.; Qian, Y. Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 161–173. [[CrossRef](#)]
14. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [[CrossRef](#)]
15. Halimi, A.; Altmann, Y.; Dobigeon, N.; Tourneret, J.Y. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4153–4162. [[CrossRef](#)]
16. Altmann, Y.; Dobigeon, N.; Tourneret, J.Y. Bilinear models for nonlinear unmixing of hyperspectral images. In Proceedings of the IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS), Shanghai, China, 6–9 June 2011; pp. 1–4.
17. Fan, W.; Hu, B.; Miller, J.; Li, M. Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data. *Int. J. Remote Sens.* **2009**, *30*, 2951–2962. [[CrossRef](#)]
18. Qu, Q.; Nasrabadi, N.M.; Tran, T.D. Abundance Estimation for Bilinear Mixture Models via Joint Sparse and Low-Rank Representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4404–4423.
19. Sigurdsson, J.; Ulfarsson, M.O.; Sveinsson, J.R. Blind Sparse Nonlinear Hyperspectral Unmixing Using an ℓ_q Penalty. *IEEE Geosci. Remote Sens. Lett.* **2018**, *99*, 1–5.
20. Altmann, Y.; Halimi, A.; Dobigeon, N.; Tourneret, J.Y. Supervised Nonlinear Spectral Unmixing Using a Postnonlinear Mixing Model for Hyperspectral Imagery. *IEEE Trans. Image Process.* **2012**, *21*, 3017–3025. [[CrossRef](#)] [[PubMed](#)]
21. Zhong, L.; Luo, W.; Gao, L. A particle swarm optimization algorithm for unmixing the polynomial post-nonlinear mixing model. In Proceedings of the Congress on Image and Signal Process., BioMedical Engineering and Informatics (CISP-BMEI), Datong, China, 15–17 October 2016; pp. 596–600.
22. Pu, H.; Chen, Z.; Wang, B.; Xia, W. Constrained Least Squares Algorithms for Nonlinear Unmixing of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 1287–1303. [[CrossRef](#)]
23. Plaza, J.; Martínez, P.; Pérez, R.; Plaza, A. Nonlinear neural network mixture models for fractional abundance estimation in AVIRIS hyperspectral images. In Proceedings of the XIII NASA/Jet Propulsion Lab. Airborne Sci. Workshop, Pasadena, CA, USA, 31 March–2 April 2004.
24. Licciardi, G.A.; Del Frate, F. Pixel unmixing in hyperspectral data by means of neural networks. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4163–4172. [[CrossRef](#)]
25. Zhang, X.; Sun, Y.; Zhang, J.; Wu, P.; Jiao, L. Hyperspectral Unmixing via Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *99*, 1–5. [[CrossRef](#)]
26. Heylen, R.; Parente, M.; Gader, P. A Review of Nonlinear Hyperspectral Unmixing Methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1844–1868. [[CrossRef](#)]
27. Broadwater, J.; Chellappa, R.; Banerjee, A.; Burlina, P. Kernel fully constrained least squares abundance estimates. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 23–28 July 2007; pp. 4041–4044.
28. Zhang, L.; Wu, B.; Huang, B.; Li, P. Nonlinear estimation of subpixel proportion via kernel least square regression. *Int. J. Remote Sens.* **2007**, *28*, 4157–4172. [[CrossRef](#)]
29. Wang, W.; Qian, Y. Kernel based sparse NMF algorithm for hyperspectral unmixing. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 6970–6973.

30. Huang, R.; Li, X.; Zhao, L. Incremental kernel non-negative matrix factorization for hyperspectral unmixing. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 6569–6572.
31. Huang, R.; Li, X.; Zhao, L. Hyperspectral Unmixing Based on Incremental Kernel Nonnegative Matrix Factorization. *IEEE Trans. Geosci. Remote Sens.* **2018**, *99*, 1–18. [[CrossRef](#)]
32. Chen, J.; Richard, C.; Honeine, P. Nonlinear unmixing of hyperspectral images with multi-kernel learning. In Proceedings of the 2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Shanghai, China, 4–7 June 2012; pp. 1–4.
33. Chen, J.; Richard, C.; Honeine, P. Nonlinear Unmixing of Hyperspectral Data Based on a Linear-Mixture/Nonlinear-Fluctuation Model. *IEEE Trans. Signal Process.* **2013**, 480–492. [[CrossRef](#)]
34. Chen, J.; Richard, C.; Honeine, P. Nonlinear estimation of material abundances in hyperspectral images with ℓ_1 -norm spatial regularization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2654–2665. [[CrossRef](#)]
35. Altmann, Y.; Dobigeon, N.; McLaughlin, S.; Tournet, J. Residual component analysis of hyperspectral images—application to joint nonlinear unmixing and nonlinearity detection. *IEEE Trans. Image Process.* **2014**, *23*, 2148–2158. [[CrossRef](#)] [[PubMed](#)]
36. Altmann, Y.; Pereyra, M.; McLaughlin, S. Bayesian Nonlinear Hyperspectral Unmixing with Spatial Residual Component Analysis. *IEEE Trans. Image Process.* **2015**, *1*, 174–185. [[CrossRef](#)]
37. Févotte, C.; Dobigeon, N. Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. *IEEE Trans. Image Process.* **2015**, *24*, 4810–4819. [[CrossRef](#)] [[PubMed](#)]
38. Ammanouil, R.; Ferrari, A.; Richard, C.; Mathieu, S. Nonlinear unmixing of hyperspectral data with vector-valued kernel functions. *IEEE Trans. Image Process.* **2017**, *26*, 340–354. [[CrossRef](#)] [[PubMed](#)]
39. Scholkopf, B.; Herbrich, R.; Smola, A.J. A generalized representer theorem. In Proceedings of the International Conference on Computational Learning Theory, Sydney, NSW, Australia, 8–10 July 2001; pp. 416–426.
40. Goldstein, T.; Osher, S. The Split Bregman Method for L1-Regularized Problems. *SIAM J. Imaging Sci.* **2009**, *2*, 323–343. [[CrossRef](#)]
41. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]
42. Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Sparse Unmixing of Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2014–2039. [[CrossRef](#)]
43. Zhao, M.; Chen, J. A dataset with ground-truth for hyperspectral unmixing. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 5081–5084.
44. Zhao, M.; Chen, J.; He, Z. A laboratory-created dataset with ground-truth for hyperspectral unmixing evaluation. *arXiv* **2019**, arXiv:1902.08347.

