



Article

Maximum Likelihood Estimation Based Nonnegative Matrix Factorization for Hyperspectral Unmixing

Qin Jiang ¹, Yifei Dong ¹, Jiangtao Peng ¹, Mei Yan ¹ and Yi Sun ^{2,*}

¹ Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China; 201822110411080@stu.hubu.edu.cn (Q.J.); 201822110411004@stu.hubu.edu.cn (Y.D.); pengjt1982@hubu.edu.cn (J.P.); 201822110411089@stu.hubu.edu.cn (M.Y.)

² School of Finance, Anhui University of Finance & Economics, Bengbu 233030, China

* Correspondence: sunyi@aufe.edu.cn; Tel.: +86-27-88662127

Abstract: Hyperspectral unmixing (HU) is a research hotspot of hyperspectral remote sensing technology. As a classical HU method, the nonnegative matrix factorization (NMF) unmixing method can decompose an observed hyperspectral data matrix into the product of two nonnegative matrices, i.e., endmember and abundance matrices. Because the objective function of NMF is the traditional least-squares function, NMF is sensitive to noise. In order to improve the robustness of NMF, this paper proposes a maximum likelihood estimation (MLE) based NMF model (MLENMF) for unmixing of hyperspectral images (HSIs), which substitutes the least-squares objective function in traditional NMF by a robust MLE-based loss function. Experimental results on a simulated and two widely used real hyperspectral data sets demonstrate the superiority of our MLENMF over existing NMF methods.

Keywords: hyperspectral unmixing; maximum likelihood estimation; nonnegative matrix factorization



Citation: Jiang, Q.; Dong, Y.; Peng, J.; Yan, M.; Sun, Y. Maximum Likelihood Estimation Based Nonnegative Matrix Factorization for Hyperspectral Unmixing. *Remote Sens.* **2021**, *13*, 2637. <https://doi.org/10.3390/rs13132637>

Academic Editors: Danfeng Hong and Gwanggil Jeon

Received: 21 May 2021

Accepted: 3 July 2021

Published: 5 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A hyperspectral image (HSI) can be represented as a three-dimensional data cube, containing both spectral and spatial information to characterize radiation properties, spatial distribution and geometric characteristics of ground objects [1,2]. Compared with panchromatic, RGB and multispectral pictures that have only several broad bands, HSI usually has hundreds of spectral bands. The rich spectral information of HSI can be used to discriminate subtle differences between similar ground objects, which makes HSI suitable for different applications, such as target recognition, mineral detection, precision agriculture [1–3]. Due to the scattering of ground surface and low spatial resolution of the hyperspectral sensor, an observed HSI pixel is often a mixture of multiple ground materials [4–6]. This is the so called “mixed pixel”. The presence of “mixed pixels” seriously affects the application of HSIs. To address the problem of mixed pixels, hyperspectral unmixing (HU) techniques have been developed [4–8]. HU aims to decompose a mixed spectral into a collection of pure spectra (endmembers) while also providing the corresponding fractions (abundances). In terms of the spectral mixture mechanism, HU algorithms can be roughly categorized into linear and non-linear ones [4,5]. Although, in general, the nonlinear mixing assumption represents the most-real cases better, the linear mixing assumption (although more simplified) has been proved to work very satisfactory in many cases in practice. Taking into account its mathematical tractability, it has attracted significant attention from the scientific community. For these reasons, the linear mixture model is adopted in the present paper, in which a measured spectral can be represented as a linear combination of several endmembers.

Nonnegative matrix factorization (NMF) is a widely used linear HU method [9–20]. In this framework, HU is regarded as a blind source separable problem, and decomposes an observed HSI matrix into the product of the pure pixel matrix (endmember matrix) and

corresponding proportion matrix (abundance matrix). Respecting the physical constraints, nonnegative constraints on the endmembers and abundances, and abundance sum-to-one constraint (ASC) are imposed. The NMF algorithm has the characteristics of intuition and interpretability. However, due to the existence of large number of unknown dependent variables, the solution space of NMF model is too large. To restrict its solution space, many NMF variants are proposed by adding constraints on the abundance or endmember [10–16]. Miao et al. incorporated a volume constraint of endmember into the NMF formulation and proposed a minimum volume constrained NMF (MVC-NMF) model [10], which can perform unsupervised endmember extraction from highly mixed image data without the pure-pixel assumption. Jia et al. introduced two constraints to the NMF [11], i.e., piecewise smoothness of spectral data and sparseness of abundance fraction. Similarly, two constraints on abundance (i.e., abundance separation constraint and abundance smoothness constraint) were added into the NMF [12]. Qian et al. imposed an $l_{1/2}$ -norm-based sparse constraints on the abundance and proposed an $l_{1/2}$ -NMF unmixing model [13]. Lu et al. considered the manifold structure of HSI and incorporated manifold regularization into the $l_{1/2}$ -NMF [14]. Wang et al. added endmember dissimilarity constraint into the NMF [15].

Although the aforementioned NMF methods improved the classical NMF unmixing model at a certain extent, they ignored the effect of noise. As the objective function of NMF is the least squares loss, NMF is sensitive to noise and corresponding unmixing results are usually inaccurate and unstable. To suppress the effect of noise and improve the robustness of the model, many robust NMF methods were proposed [17–20]. He et al. proposed a sparsity-regularized robust NMF by adding a sparse matrix into the linear mixture model to model the sparse noise [17]. Du et al. introduced a robust entropy-induced metric (CIM) and proposed a CIM-based NMF (CIM-NMF) model, which can effectively deal with non-Gaussian noise [18]. Wang et al. proposed a robust correntropy-based NMF model (CENMF) [19], which contained a correntropy-based loss function and an l_1 -norm sparse constraint on the abundance. Based on the Huber's M-estimator, Huang et al. constructed $l_{2,1}$ -norm and $l_{1,2}$ -norm based loss functions to obtain a new robust NMF model [20,21]. Defining the $l_{2,1}$ -norm ($l_{1,2}$ -norm) based loss function actually assumes that the column-wise (row-wise) approximation residual follows Laplacian (Gaussian) distribution from the viewpoint of maximum likelihood estimation (MLE). However, in practice this assumption may not hold well, especially when HSI contains complex mixture noise, such as impulse noise, stripes, deadlines, and other types of noise [22,23].

Inspired by the robust regression theory [23,24], we design the approximation residual as an MLE-like estimator and propose a robust MLE-based $l_{1/2}$ -NMF model (MLENMF) for HU. It replaces the least-squares loss in the original NMF by a robust MLE-based loss, which is a function (associated with the distribution of the approximation residuals) of the approximation residuals [24]. The proposed MLENMF can be converted to a weighted $l_{1/2}$ -NMF model and can be solved by a re-weighted multiplication update iteration algorithm [9,13]. By choosing an appropriate weight function, MLENMF can automatically assign small weights to bands with large residuals, which can effectively reduce the effect of noisy bands and improve the unmixing accuracy. Experimental results on simulated and real hyperspectral data sets show the superiority of MLENMF over existing NMF methods.

The rest of the paper is organized as follows. Section 2 introduces the NMF and $l_{1/2}$ -NMF. Section 3 describes our proposed MLENMF method. The experimental results and analysis are provided in Section 4. Section 5 discusses the effect of parameters in the algorithm. Finally, Section 6 concludes the paper.

2. NMF Unmixing Model

Under the linear spectral mixing mechanism, an observed spectral $\mathbf{h} \in R^{M \times 1}$ can be represented linearly by the endmember $\mathbf{z}_1, \dots, \mathbf{z}_P$ [4,10–13]:

$$\mathbf{h} = \mathbf{Z}\mathbf{s} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{Z} = [z_1, \dots, z_p] \in R^{M \times P}$ represents the endmember matrix, $\mathbf{s} \in R^{P \times 1}$ is the coefficient (abundance) vector, and ϵ is the residual. Applying the above linear mixing model (1) for all hyperspectral pixels $\mathbf{h}_1, \dots, \mathbf{h}_N$, the following matrix representation can be obtained:

$$\mathbf{H} = \mathbf{ZS} + \mathbf{E}, \quad (2)$$

where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_N] \in R^{M \times N}$, $\mathbf{S} = [s_1, \dots, s_N] \in R^{P \times N}$ are nonnegative hyperspectral data matrix and abundance matrix, respectively. $\mathbf{E} \in R^{M \times N}$ is the residual matrix.

In Equation (2), to make the decomposition result as accurate as possible, the residual should be minimized. Then, an NMF unmixing model can be obtained by considering the nonnegative property of endmember and abundance matrices:

$$\min_{\mathbf{Z}, \mathbf{S}} \|\mathbf{H} - \mathbf{ZS}\|_F^2, \quad \text{s.t.}, \mathbf{Z} \geq 0, \mathbf{S} \geq 0, \quad (3)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, and $\mathbf{Z} \geq 0$ means that each element of \mathbf{Z} is nonnegative. As each column of abundance matrix \mathbf{S} records the proportion of endmembers in representing a pixel, the columns of \mathbf{S} (each one corresponding to a pixel) should satisfy the sum-to-one constraint, i.e., $\sum_{p=1}^P S_{pn} = 1$, $n = 1, \dots, N$.

The above NMF Model (3) can be easily solved by the multiplication update algorithm [9,13]. However, its solution space is very large [13]. To restrict the solution space, an $l_{1/2}$ -constraint can be added to the abundance matrix \mathbf{S} , and an $l_{1/2}$ -NMF model can be obtained as [13]:

$$\min_{\mathbf{Z}, \mathbf{S}} \|\mathbf{H} - \mathbf{ZS}\|_F^2 + \lambda \|\mathbf{S}\|_{1/2}, \quad \text{s.t.}, \mathbf{Z} \geq 0, \mathbf{S} \geq 0, \quad (4)$$

where λ is a regularization parameter and $\|\mathbf{S}\|_{1/2}$ is the $l_{1/2}$ -regularizer [13]. As proved in Refs. [13,25], $l_{1/2}$ -regularizer is a good choice in enforcing the sparsity of hyperspectral unmixing because the sparsity of the l_q ($1/2 \leq q < 1$) solution increases as q decreases, whereas the sparsity of the solution for l_q ($0 < q \leq 1/2$) shows little change with respect to q . Meanwhile, the sparsity represented by $l_{1/2}$ also enforces the volume of the simplex to be minimized [13].

3. MLENMF Unmixing Model

In the NMF model (3) or (4), the objective function $\|\mathbf{H} - \mathbf{ZS}\|_F^2$ is the least-squares (LS) function which is sensitive to noise. Here, we employ a new robust MLE-based loss to replace the LS objective function and propose an MLE-based NMF (MLENMF) model for HU.

Firstly, the matrix norm form is transformed into vector norm form:

$$\|\mathbf{H} - \mathbf{ZS}\|_F^2 = \sum_{i=1}^M \|\mathbf{H}^i - (\mathbf{ZS})^i\|_2^2, \quad (5)$$

where \mathbf{H}^i is the i -th row of matrix \mathbf{H} .

We can regard the least squares objective function as the sum of approximation residuals, and then construct an MLE-like robust estimator to approximate the minimum of objective function. Denote the approximation residual of the i -th band as $e_i = \|\mathbf{H}^i - (\mathbf{ZS})^i\|_2$ and define residual vector $\mathbf{e} = [e_1, \dots, e_M]^T$, the above Formula (5) can be rewritten as:

$$J(\mathbf{e}) = \|\mathbf{e}\|_2^2 = \sum_{i=1}^M e_i^2. \quad (6)$$

Assume that e_1, \dots, e_M are independent and identically distributed (i.i.d) random variables, which follow the same probability distribution function $g_\theta(e_i)$, where θ is the distribution parameter. The likelihood function can be expressed as:

$$J_\theta(e_1, \dots, e_n) = \prod_{i=1}^M g_\theta(e_i). \quad (7)$$

According to the principle of MLE, the following objective function should be minimized:

$$-\ln J_\theta = \sum_{i=1}^M \varphi_\theta(e_i), \quad (8)$$

where $\varphi_\theta(e_i) = -\ln g_\theta(e_i)$. If we replace the objective function $\|\mathbf{H} - \mathbf{ZS}\|_F^2$ in Equation (4) by the loss in Equation (8), we can get the following optimization problem:

$$\min_{\mathbf{Z}, \mathbf{S}} \sum_{i=1}^M \varphi_\theta(e_i) + \lambda \|\mathbf{S}\|_{1/2}, \quad \text{s.t., } \mathbf{Z} \geq 0, \mathbf{S} \geq 0, \quad (9)$$

In fact, the aim is to construct a loss function to replace the least squares function to reduce the impact of noise. To construct the loss function, we analyze its Taylor expansion. Assume that g_θ is symmetric, and $g_\theta(e_i) < g_\theta(e_j)$ if $|e_i| > |e_j|$. We can infer that: (1) $g_\theta(0)$ is global maximum of g_θ and $\varphi_\theta(0)$ is the global minimum of φ_θ ; (2) $\varphi_\theta(e_i) = \varphi_\theta(-e_i)$; (3) $\varphi_\theta(e_i) > \varphi_\theta(e_j)$ if $|e_i| > |e_j|$. For simplicity, we assume $\varphi_\theta(0) = 0$. Define $D_\theta(\mathbf{e}) = \sum_{i=1}^M \varphi_\theta(e_i)$. According to the first-order Taylor expansion around \mathbf{e}_0 , $D_\theta(\mathbf{e})$ can be approximated as [24]:

$$\tilde{D}_\theta(\mathbf{e}) = D_\theta(\mathbf{e}_0) + (\mathbf{e} - \mathbf{e}_0)^T D'_\theta(\mathbf{e}_0) + \frac{1}{2} (\mathbf{e} - \mathbf{e}_0)^T \mathbf{W} (\mathbf{e} - \mathbf{e}_0), \quad (10)$$

where $D'_\theta(\mathbf{e}_0)$ is the first order derivative of $D_\theta(\mathbf{e})$ at \mathbf{e}_0 , and \mathbf{W} is the Hessian matrix. We can get the mixed partial derivatives $\frac{\partial^2 D_\theta}{\partial e_i \partial e_j} = 0$ ($e_i \neq e_j$) as the error residuals e_i and e_j are assumed i.i.d., and hence \mathbf{W} is a diagonal matrix. Taking the derivative of $\tilde{D}_\theta(\mathbf{e})$ with respect to \mathbf{e} , it gets

$$\tilde{D}'_\theta(\mathbf{e}) = D'_\theta(\mathbf{e}_0) + \mathbf{W}(\mathbf{e} - \mathbf{e}_0). \quad (11)$$

As $\varphi_\theta(0) = 0$ is the global minimum of φ_θ , the minimum of $D_\theta(\mathbf{e})$ is $D_\theta(\mathbf{0})$. $\tilde{D}_\theta(\mathbf{e})$ should also reach its minimum at $\mathbf{e} = \mathbf{0}$ for it is an approximation of $D_\theta(\mathbf{e})$, so $\tilde{D}'_\theta(\mathbf{0}) = 0$ and then we can derive the following formulas from Equation (11):

$$D'_\theta(\mathbf{e}_0) - \mathbf{W}\mathbf{e}_0 = 0 \quad (12)$$

$$W_{i,i} = \frac{\varphi'_\theta(e_{0,i})}{e_{0,i}} \quad (13)$$

where $W_{i,i}$ is the i -th diagonal element of \mathbf{W} . Denote $w_i = W_{i,i}$, Equation (13) can be written as

$$\varphi'_\theta(e_i) = w_i e_i \quad (14)$$

As $\varphi_\theta(x)$ is a nonlinear and nonconvex function, it is difficult to solve the model (9) directly. Inspired by the above Formula (14), we can get:

$$\varphi_\theta(e_i) = w_i e_i^2, \quad (15)$$

and then the Model (9) can be expressed as a weighted NMF model:

$$\min_{\mathbf{Z}, \mathbf{S}} \sum_{i=1}^M w_i e_i^2 + \lambda \|\mathbf{S}\|_{1/2}, \quad \text{s.t.}, \quad \mathbf{Z} \geq 0, \mathbf{S} \geq 0, \quad (16)$$

The objective function of Model (16) can be rewritten as:

$$\begin{aligned} \sum_{i=1}^M w_i e_i^2 + \lambda \|\mathbf{S}\|_{1/2} &= \sum_{i=1}^M w_i \|\mathbf{H}^i - (\mathbf{Z}\mathbf{S})^i\|_2^2 + \lambda \|\mathbf{S}\|_{1/2} \\ &= \sum_{i=1}^M \|\sqrt{w_i} \mathbf{H}^i - (\sqrt{w_i} \mathbf{Z}\mathbf{S})^i\|_2^2 + \lambda \|\mathbf{S}\|_{1/2} \\ &= \|\tilde{\mathbf{H}} - \tilde{\mathbf{Z}}\mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_{1/2}, \end{aligned} \quad (17)$$

where $\tilde{\mathbf{H}} = \sqrt{\mathbf{W}}\mathbf{H}$, $\tilde{\mathbf{Z}} = \sqrt{\mathbf{W}}\mathbf{Z}$. Then, the Model (16) can be expressed as:

$$\min_{\mathbf{Z}, \mathbf{S}} \sum_{i=1}^M \|\tilde{\mathbf{H}} - \tilde{\mathbf{Z}}\mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_{1/2}, \quad \text{s.t.}, \quad \mathbf{Z} \geq 0, \mathbf{S} \geq 0. \quad (18)$$

It is easy to see that model (18) is also an $l_{1/2}$ -NMF algorithm, and can be solved by the multiplication update iteration rule as follows [9,13]:

$$\tilde{\mathbf{Z}} \leftarrow \tilde{\mathbf{Z}} \cdot * (\tilde{\mathbf{H}}\mathbf{S}^T) ./ (\tilde{\mathbf{Z}}\mathbf{S}\mathbf{S}^T) \quad (19)$$

$$\mathbf{S} \leftarrow \mathbf{S} \cdot * (\tilde{\mathbf{Z}}^T \tilde{\mathbf{H}}) ./ \left(\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}}\mathbf{S} + \frac{\lambda}{2} \mathbf{S}^{-\frac{1}{2}} \right) \quad (20)$$

The final endmember matrix is $\mathbf{Z} = \mathbf{W}^{-\frac{1}{2}} \tilde{\mathbf{Z}}$.

In the model (18), a key factor is the weight. In this paper, the weight function is set as the logistic function [23,24,26]:

$$w_i \triangleq w(e_i) = \frac{1}{1 + \exp(-\gamma(\tau - e_i^2))} = \frac{\exp(\gamma(\tau - e_i^2))}{1 + \exp(\gamma(\tau - e_i^2))}, \quad (21)$$

where γ, τ are positive scalars. Parameter γ controls the decreasing rate from 1 to 0, and τ controls the location of demarcation point [24]. It is clear that the value of weight function decreases rapidly with the increase of residual e_i .

MLE weight function in Equation (21) can approximate the weight of commonly used loss functions, such as $l_{2,1}$, maximum correntropy and Huber weights.

When $\gamma = 2$ and $\tau \rightarrow 0$, the MLE weight function is:

$$w(e_i) = \frac{1}{1 + \exp(2e_i^2)} \xrightarrow{\text{for small } e_i} \frac{1}{2(1 + e_i^2)} \quad (22)$$

which is close to $l_{2,1}$ weight: $\frac{1}{1+e_i^2}$. The corresponding weights are shown as red and blue lines in Figure 1a.

When $\gamma = \frac{1}{\sigma^2}$ and $\tau \rightarrow 0$, the MLE weight function is:

$$w(e_i) = \frac{1}{\left(1 + \exp\left(\frac{e_i^2}{\sigma^2}\right)\right)} \quad (23)$$

which is close to the weight of maximum correntropy criterion: $\exp\left(-\frac{e_i^2}{\sigma^2}\right)$ (σ is a parameter). The corresponding weights are shown in Figure 1b.

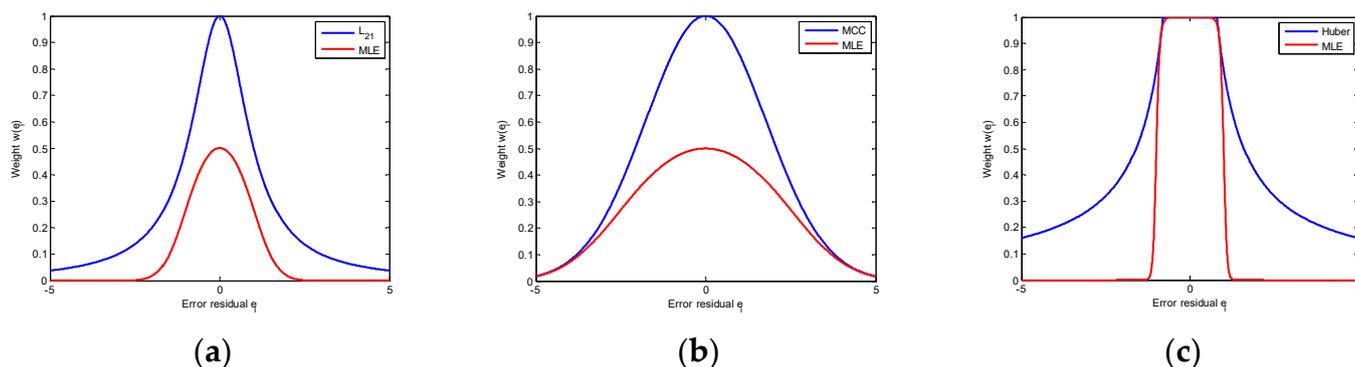


Figure 1. MLE weights fit $l_{2,1}$ weight (a), maximum correntropy weight (b), Huber weight (c).

By choosing appropriate parameters, the MLE weight can also approximate the Huber weight:

$$w_{\text{Huber}}(e_i) = \begin{cases} 1, & |e_i| \leq c \\ \frac{c}{|e_i|}, & |e_i| > c \end{cases} \tag{24}$$

as shown in Figure 1c.

Based on Equations (14) and (21), the objective function of MLE can be obtained as:

$$\begin{aligned} \varphi_{\theta}(e_i) &= \int_0^{e_i} \varphi'_{\theta}(e_i) de_i = \int_0^{e_i} e_i w(e_i) de_i \\ &= \int_0^{e_i} e_i \frac{\exp(\gamma(\tau - e_i^2))}{1 + \exp(\gamma(\tau - e_i^2))} de_i \\ &= \frac{-1}{2\gamma} \ln(1 + \exp(\gamma(\tau - e_i^2))) \Big|_0^{e_i} \\ &= \frac{-1}{2\gamma} \ln \frac{1 + \exp(\gamma(\tau - e_i^2))}{1 + \exp(\gamma\tau)} \end{aligned} \tag{25}$$

From Equations (8) and (25), we can see that the probability distribution function $g_{\theta}(e_i)$ has the form:

$$g_{\theta}(e_i) = \left(\frac{1 + \exp(\gamma(\tau - e_i^2))}{1 + \exp(\gamma\tau)} \right)^{\frac{1}{2\gamma}} \tag{26}$$

If $\tau = 0, \gamma \rightarrow 0$, the probability distribution function $g_{\theta}(e_i)$ is actually a Gaussian distribution:

$$\ln g_{\theta}(e_i) = \frac{\ln \frac{1 + \exp(-\gamma e_i^2)}{2}}{2\gamma} \xrightarrow{\gamma \rightarrow 0} \frac{\exp(-\gamma e_i^2) \cdot (-e_i^2)}{1 + \exp(-\gamma e_i^2)} \xrightarrow{\gamma \rightarrow 0} -\frac{e_i^2}{2} \tag{27}$$

$$g_{\theta}(e_i) \xrightarrow{\gamma \rightarrow 0} \exp\left(-\frac{e_i^2}{2}\right) \tag{28}$$

In this case, the weight defined in Equation (21) is: $\omega_i = 1/2$, which is the LS case.

In Figure 2a, we compare the MLE objective function with the LS loss function. MLE objective function is controlled by the parameters γ, τ , and is truncated to a constant for large residuals (e.g., $|e_i| > 2$). As the constant has no effect on the optimization model, the negative effect of noise (points with large residuals) can be automatically diminished. Compared with the MLE function, LS loss function is global and increases quadratically as the increase of residual. When there has heavy noise, the objective function of LS model will be dominated by the points with heavy noise. Figure 2b shows the influence function [22,27] of MLE and LS. The influence function of a loss $\varphi(e)$ is defined as: $\psi(e) = \partial\varphi(e)/\partial e$, which measures the robustness of loss function as the increase of error residual. For residual $e_i > 0$, the influence function of MLE increases first, then decreases and finally reaches the zero value. It means that larger errors finally have no effect on the MLE-based model.

However, the influence function of LS continues to grow linearly. So, the LS loss function is seriously affected by noise. In the presence of noise, MLE is obviously more robust than LS.

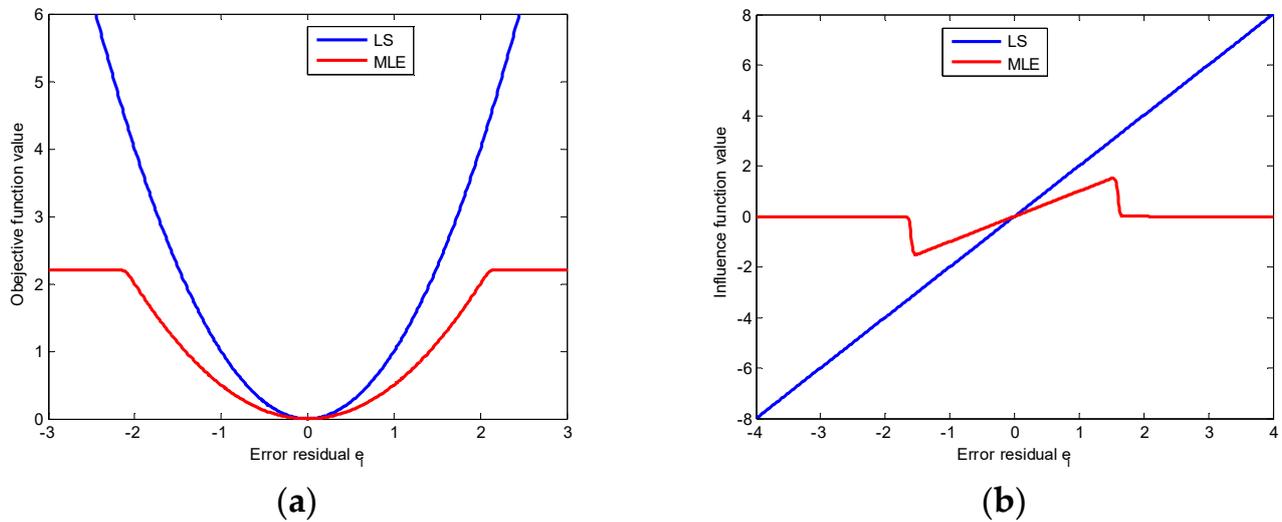


Figure 2. Comparison of objective function (a) and influence function (b) between MLE and LS.

The procedure of the proposed MLENMF is shown in Algorithm 1.

Algorithm 1 MLENMF.

Input: hyperspectral matrix H , the parameter γ, τ

Initialization: endmember Z_0 and abundance S_0 ,

Output: estimated endmember and abundance matrices.

1. Initialize $Z^{(0)} = Z_0, S^{(0)} = S_0, v = 1, W = I$
2. Run the following steps until convergence: (a) Compute the errors:

$$(e_i^2)^{(v)} = \|H^i - (Z^{(v-1)}S^{(v-1)})^i\|_2^2$$

- (b) Calculate the weight of each entry:

$$w^{(v)}(e_i) = \frac{\exp(\gamma\tau - \gamma(e_i^2)^{(v)})}{1 + \exp(\gamma\tau - \gamma(e_i^2)^{(v)})}$$

- (c) Compute the weighted matrices:

$$\begin{aligned} \tilde{H} &= (W^{(v)})^{\frac{1}{2}} H \\ \tilde{Z}^{(v-1)} &= (W^{(v)})^{\frac{1}{2}} Z^{(v-1)} \end{aligned}$$

- (d) Updating endmember matrix and weighted abundance matrix:

$$\begin{aligned} (\tilde{Z}^{(v)}, S^{(v)}) &= L_{1/2} \text{NMF}(\tilde{H}, \tilde{Z}^{(v-1)}, S^{(v-1)}) \\ Z^{(v)} &= (W^{(v)})^{-\frac{1}{2}} \tilde{Z}^{(v)} \end{aligned}$$

- (e) $v = v + 1$
-

Remark. In the current method, it assumes that different bands are independent and then an MLE solution can be deduced. The band independence assumption is only used in the derivation of MLE estimator. By means of this assumption, it can finally generate a weighted NMF model where the weight function can be used to reduce the effect of noisy bands. Although hyperspectral bands are not independent from each other in practice, the final weighted NMF model (i.e., MLENMF) can still alleviate negative effects of noise.

4. Results

In this section, we perform experiments on a simulated and two real hyperspectral data sets to test the performance of MLENMF model and compare the results with $l_{1/2}$ -NMF [13], l_{21} -NMF [21], CENMF [19], CIMNMF [18], and HuberNMF (HubNMF for short) [18].

4.1. Evaluation Metrics

Spectral angular distance (SAD) and root mean square error (RMSE) are used to quantitatively evaluate the accuracy of estimated endmembers and abundances.

The formula of SAD is:

$$\text{SAD}_k = \arccos\left(\frac{\mathbf{z}_k^T \hat{\mathbf{z}}_k}{\|\mathbf{z}_k\| \cdot \|\hat{\mathbf{z}}_k\|}\right), \quad (29)$$

where SAD_k represents the similarity between the k -th real endmember \mathbf{z}_k and estimated endmember $\hat{\mathbf{z}}_k$.

The RMSE is:

$$\text{RMSE}_k = \left(\frac{1}{N} \|\mathbf{s}^k - \hat{\mathbf{s}}^k\|^2\right)^{\frac{1}{2}}, \quad (30)$$

where \mathbf{s}_k and $\hat{\mathbf{s}}_k$ are the k -th real and estimated abundance maps (i.e., the k -th row vector in \mathbf{S} and $\hat{\mathbf{S}}$), respectively. N is the number of pixels in HSI.

4.2. Implementation Details

The vertex component analysis (VCA) and fully constrained least squares (FCLS) methods are used to generate the initial endmember \mathbf{Z}_0 and abundance \mathbf{S}_0 for different unmixing methods [11–19]. The regularization parameter λ in $l_{1/2}$ -NMF and CENMF is dependent on the sparsity of the material abundances and is estimated based on the sparseness criterion in Ref. [13]. The parameter of CIMNMF and Huber-NMF are set to be the recommended values in Ref. [18]. The proposed MLENMF contains two parameters γ and τ as shown in Equation (21). It is clear that τ is related to the amplitude of residual e_i^2 . For different data sets, the amplitude of residuals may be different. So, it is difficult to determine a specific value of τ . Here, we set τ in a data-dependent way: τ is the (100ξ) -th percentile of residual vector $\tilde{\mathbf{e}} = [e_1^2, \dots, e_M^2]^T$, where $\xi \in (0, 1]$ controls the ratio of inliers. Following Reference [24], parameter γ is set as $\gamma = c/\tau$, $c \in (0, 10]$. So, in the experiments, we only need to tune the parameters ξ and c .

4.3. Experiments on Simulated Data

Seven spectral signatures (i.e., “Carnallite NMNH98011”, “Actinolite NMNHR16485”, “Andradite WS487”, “Diaspore HS416.3B”, “Erionite+Merlinoit GDS144”, “Halloysite NMNH106236”, “Hypersthene NMNHC2368”) from the USGS spectral library (<https://www.usgs.gov/labs/spec-lab>, accessed on 2 July 2019) are selected to construct the endmember matrix $\mathbf{Z} \in R^{224 \times 7}$. Then, these seven spectra are mixed according to the method described in Ref. [28] to form the corresponding abundance matrix $\mathbf{S} \in R^{7 \times 4096}$. The hyperspectral data matrix is obtained by the product of endmember and abundance matrices, i.e., $\mathbf{H} = \mathbf{ZS}$. To simulate the real situations, Gaussian noise is added into the data matrix \mathbf{H} such that the signal noise ratio (SNR) of different bands follow the Gaussian distribution, i.e., $\text{SNR} \sim \mathcal{N}(\mu, \delta^2)$ with $\delta = 5$. In the experiment, $\mu \in \{5, 10, 15, 20\}$ are

considered. A large μ corresponds to small noise. In the MLENMF, parameter ζ is set as 0.4, and parameter $c = 10$ is used for $\mu \leq 10$ and $c = 1$ is used for $\mu \geq 10$.

Tables 1 and 2 show the average results of 20 random experiments under different degrees of noise. Each SAD (RMSE) value is the mean of SAD (RMSE) over seven endmembers. It is clear that the performance of different methods are improved as the increase of SNR or μ , and MLENMF shows better results in different degrees of noise.

Table 1. The SAD results of different unmixing methods for simulated data.

μ	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
5	0.4972	0.4754	0.4324	0.4088	0.4732	0.4718	0.3895
10	0.3513	0.3146	0.2901	0.2813	0.3093	0.3086	0.2537
15	0.1997	0.1749	0.1764	0.1626	0.1781	0.1772	0.1134
20	0.0988	0.0950	0.0936	0.0865	0.0985	0.0946	0.0689

Table 2. The RMSE results of different unmixing methods for simulated data.

μ	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
5	0.2679	0.2763	0.2737	0.2586	0.2653	0.2650	0.2459
10	0.2491	0.2554	0.2497	0.2345	0.2435	0.2448	0.2114
15	0.1951	0.1956	0.1983	0.1811	0.1954	0.1920	0.1440
20	0.1137	0.1219	0.1233	0.1145	0.1208	0.1203	0.0599

To visualize the results of different methods, the real and estimated spectra for the endmember 1 (i.e., “Carnallite NMNH98011”) at $\mu = 20$ are shown in Figure 3. Here, we only show the results for the endmember 1 due to space limitations. Similar good results are obtained for the other endmembers. It can be seen that the spectral curve estimated by the MLENMF can well approximate the reference one while the curve of other methods exhibit deviations in amplitude from the reference spectral. As the reference spectral and estimated spectral by different methods have similar shape, the SAD of different methods show small differences as shown in Table 1. Notwithstanding, the estimated abundance map of different methods show large differences as shown in Figure 4. Taking into account both SAD and RMSE results, we can see that our MLENMF method is more robust than other NMF methods when the data contains noise.

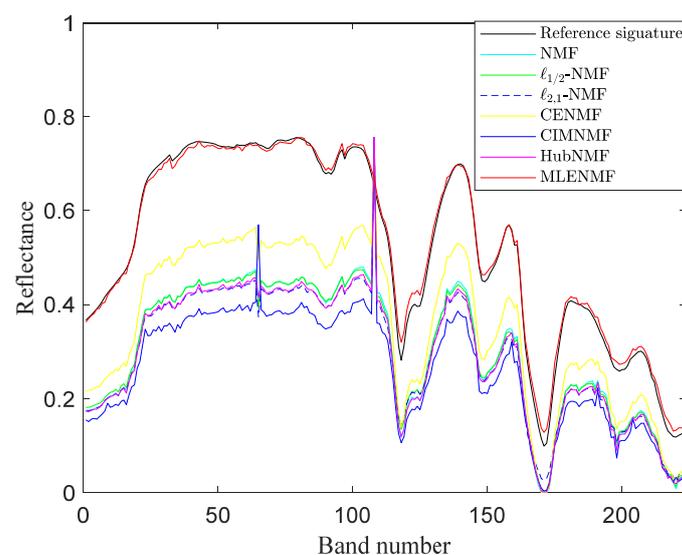


Figure 3. The reference and estimated spectra for the endmember 1 of simulated data.

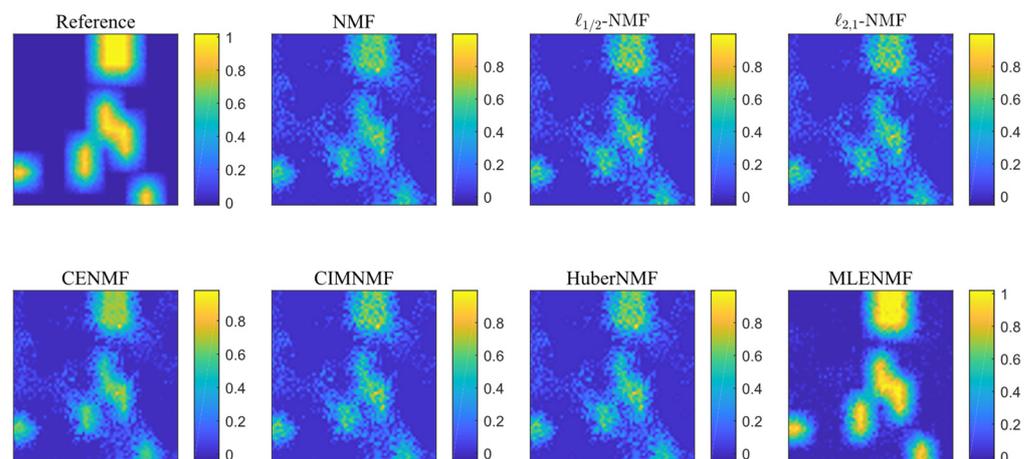


Figure 4. The reference and estimated abundance maps for the endmember 1 of simulated data.

4.4. Experiments on Real Data

Two real hyperspectral unmixing data sets, i.e., Urban and Japser are used to evaluate the performance of different NMF unmixing methods (Available at <https://rslab.ut.ac.ir/data>, <https://sites.google.com/site/feiyunzhuhompage/datasets-ground-truths>, accessed on 2 July 2019). The Urban data was obtained by the HYDICE sensor. This scene has the size of 307×307 pixels and each pixel corresponds to an 2×2 m² area. The original data has 210 bands, where band 1–4, 76, 87, 101–111, 136–153, 198–210 are severely affected by dense water vapor and atmosphere. After removing these noisy bands, 162 bands are kept. This scene contains four reference materials: Asphalt road, Grass, Tree and Roof, which are also available at the <https://rslab.ut.ac.ir/data>, accessed on 2 July 2019.

We first perform experiments on the Urban data with 162 bands. The parameters of MLENMF are set as: $\zeta = 0.8$ and $c = 1$. The estimated endmembers and abundances by different unmixing methods are compared with the groundtruth references and then the SAD and RMSE results are computed, as shown in Tables 3 and 4, respectively. Compared with other NMF methods, the proposed MLENMF shows better overall results. Figure 5 shows the estimated endmembers by different methods. It can be seen that the other methods cannot well estimate the endmember ‘Roof’, while our MLENMF generates spectral curve that is similar to the reference signature. From the abundance maps in Figure 6, we can see that the maps of MLENMF are more consistent with the reference maps than comparison methods.

Table 3. The SAD results of different methods for Urban data with 162 bands.

	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
Asphalt	0.1587	0.1079	0.1576	0.1575	0.1634	0.1601	0.1127
Grass	0.2954	0.2477	0.2983	0.2857	0.2922	0.2826	0.0543
Tree	0.1919	0.1058	0.1917	0.1943	0.1938	0.1923	0.1011
Roof	0.6709	0.5748	0.6841	0.6501	0.6727	0.6534	0.0914
Mean	0.3292	0.2591	0.3329	0.3219	0.3305	0.3221	0.0899

Table 4. The RMSE results of different methods for Urban data with 162 bands.

	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
Asphalt	0.2126	0.2843	0.2123	0.2114	0.2126	0.2127	0.1383
Grass	0.2386	0.2683	0.2395	0.2391	0.2394	0.2385	0.1318
Tree	0.1692	0.1387	0.1693	0.1699	0.1693	0.1691	0.0605
Roof	0.1574	0.1718	0.1568	0.1574	0.1569	0.1574	0.0538
Mean	0.1945	0.2158	0.1945	0.1945	0.1945	0.1944	0.0961

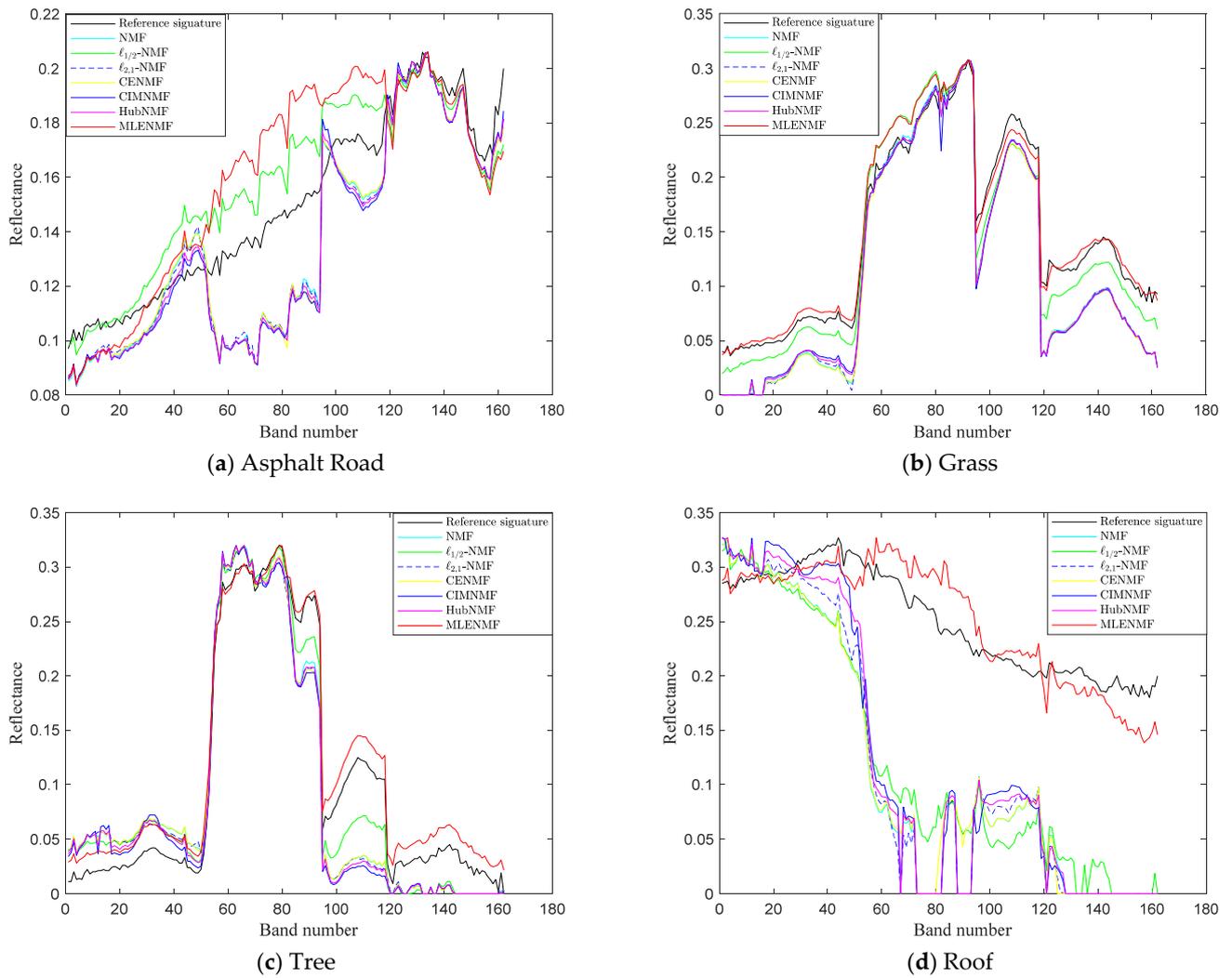
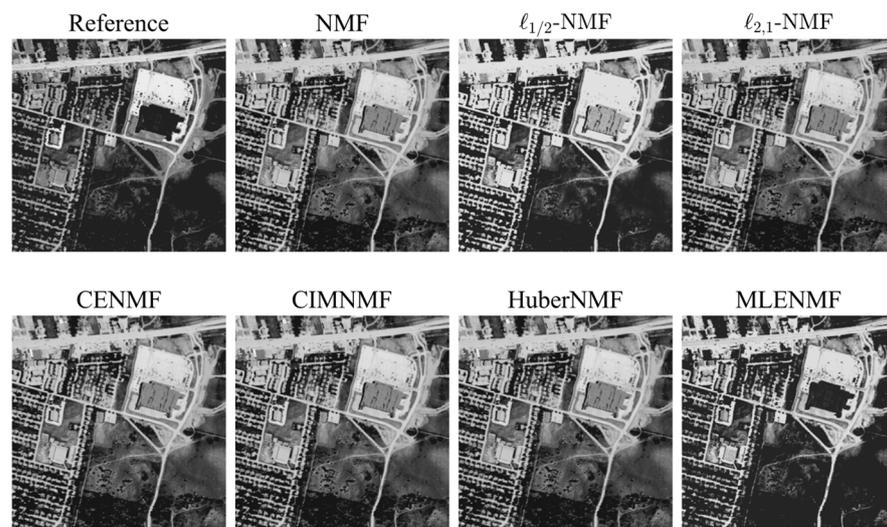


Figure 5. Comparison of endmembers estimated by different methods for Urban data with 162 bands for different materials: (a) Asphalt Road, (b) Grass, (c) Tree, (d) Roof.



(a) Asphalt Road

Figure 6. Cont.

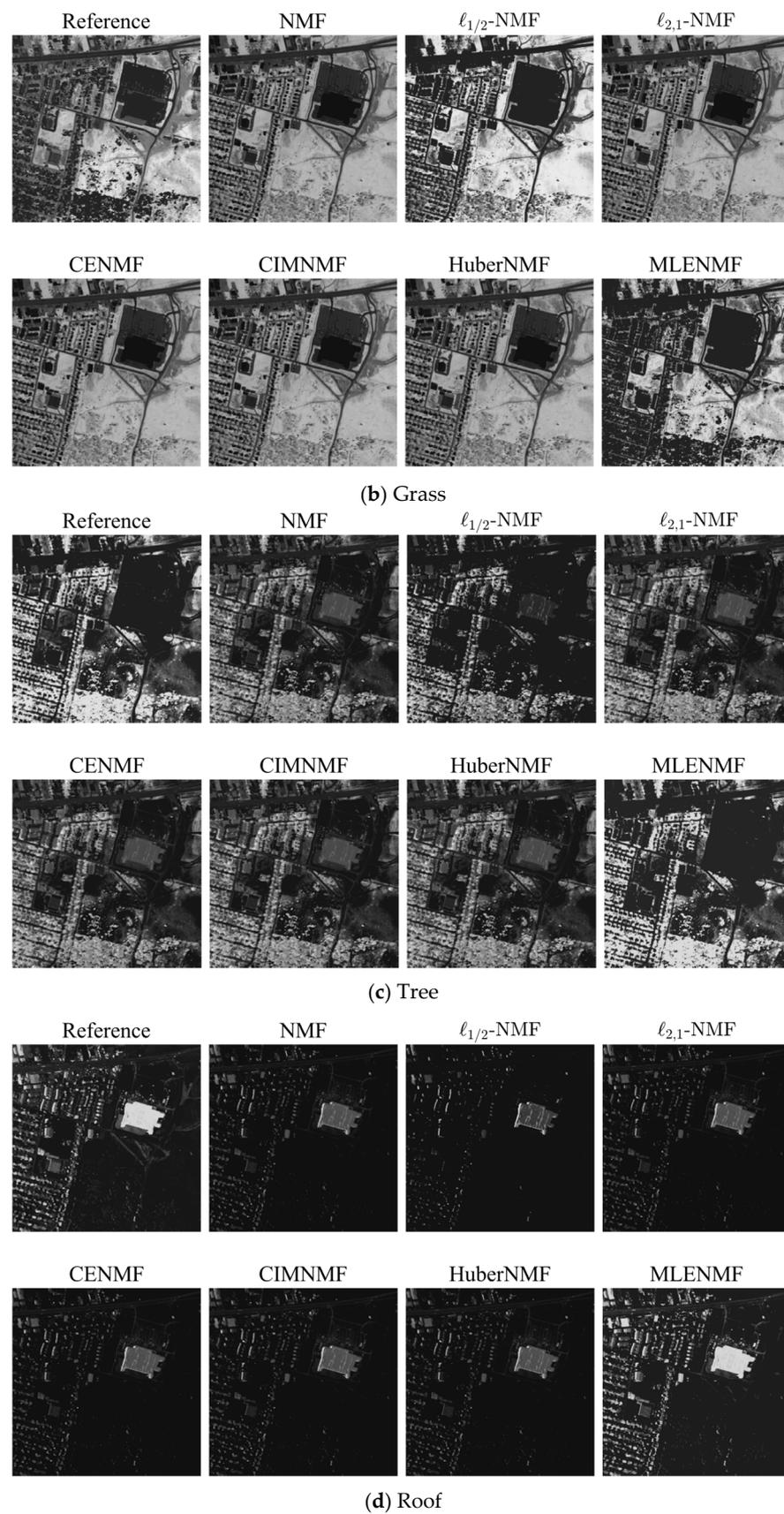


Figure 6. Comparison of abundance maps estimated by different algorithms for Urban data with 162 bands for different materials: (a) Asphalt Road, (b) Grass, (c) Tree, (d) Roof.

To test the unmixing performance of different methods in the case of noisy bands, we also calculate the SAD and RMSE for the Urban data with the whole 210 bands and show the results in Tables 5 and 6, respectively. The parameters of MLENMF in this case are set as: $\zeta = 0.4$ and $c = 10$. Even with some known bad bands, our MLENMF also provides the best results.

Table 5. The SAD results of different methods for Urban data with 210 bands.

	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
Asphalt	0.1229	0.1234	0.1295	0.1271	0.1345	0.1312	0.1481
Grass	0.4703	0.3544	0.4607	0.4854	0.4637	0.4618	0.1436
Tree	0.3197	0.2057	0.3139	0.3401	0.3035	0.3048	0.1979
Roof	0.4446	0.3048	0.4083	0.4705	0.4466	0.4348	0.3039
Mean	0.3394	0.2471	0.3281	0.3558	0.3371	0.3332	0.1984

Table 6. The RMSE results of different methods for Urban data with 210 bands.

	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
Asphalt	0.2443	0.3204	0.2453	0.2391	0.2428	0.2441	0.2659
Grass	0.3763	0.4655	0.3770	0.3761	0.3770	0.3769	0.2985
Tree	0.3374	0.3711	0.3381	0.3373	0.3380	0.3380	0.2679
Roof	0.2213	0.2377	0.2217	0.2155	0.2197	0.2208	0.2616
Mean	0.2948	0.3487	0.2955	0.2920	0.2944	0.2949	0.2735

The Jasper data is collected by the AVIRIS sensor, covering a spectral range of 380 to 2500 nm, with a total of 224 spectral bands, including 26 noisy bands. The spectral resolution is 9.46 nm, and the image size is 100×100 . The image mainly contains four materials: Tree, Water, Soil, and Road. The parameters of MLENMF are set as: $\zeta = 0.4$ and $c = 1$. The SAD and RMSE results of different unmixing methods on this data set are shown in Tables 7 and 8, respectively. Figures 7 and 8 show the estimated endmembers and abundance maps of different methods. It can be seen from these results that the proposed MLENMF can provide more accurate estimation on the endmembers and abundances.

Table 7. The SAD results of different methods for Jasper data.

	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
Tree	0.3040	0.2787	0.2981	0.2503	0.2189	0.3029	0.0851
Water	0.3156	0.1163	0.2799	0.2974	0.2968	0.2969	0.1841
Soil	0.2834	0.1660	0.3186	0.3871	0.2490	0.1872	0.0908
Road	0.6469	0.5341	0.6627	0.6953	0.7052	0.6961	0.2269
Mean	0.3875	0.2738	0.3898	0.4075	0.3675	0.3708	0.1468

Table 8. The RMSE results of different methods for Jasper data.

	NMF	$l_{1/2}$ -NMF	l_{21} -NMF	CENMF	CIMNMF	HubNMF	MLENMF
Tree	0.2310	0.3032	0.2711	0.2251	0.2287	0.2350	0.1055
Water	0.1563	0.1726	0.1738	0.1623	0.1502	0.1508	0.1029
Soil	0.3382	0.3789	0.3342	0.3209	0.3045	0.3207	0.2510
Road	0.2385	0.2634	0.2534	0.2639	0.2351	0.2337	0.2350
Mean	0.2410	0.2796	0.2581	0.2430	0.2296	0.2351	0.1736

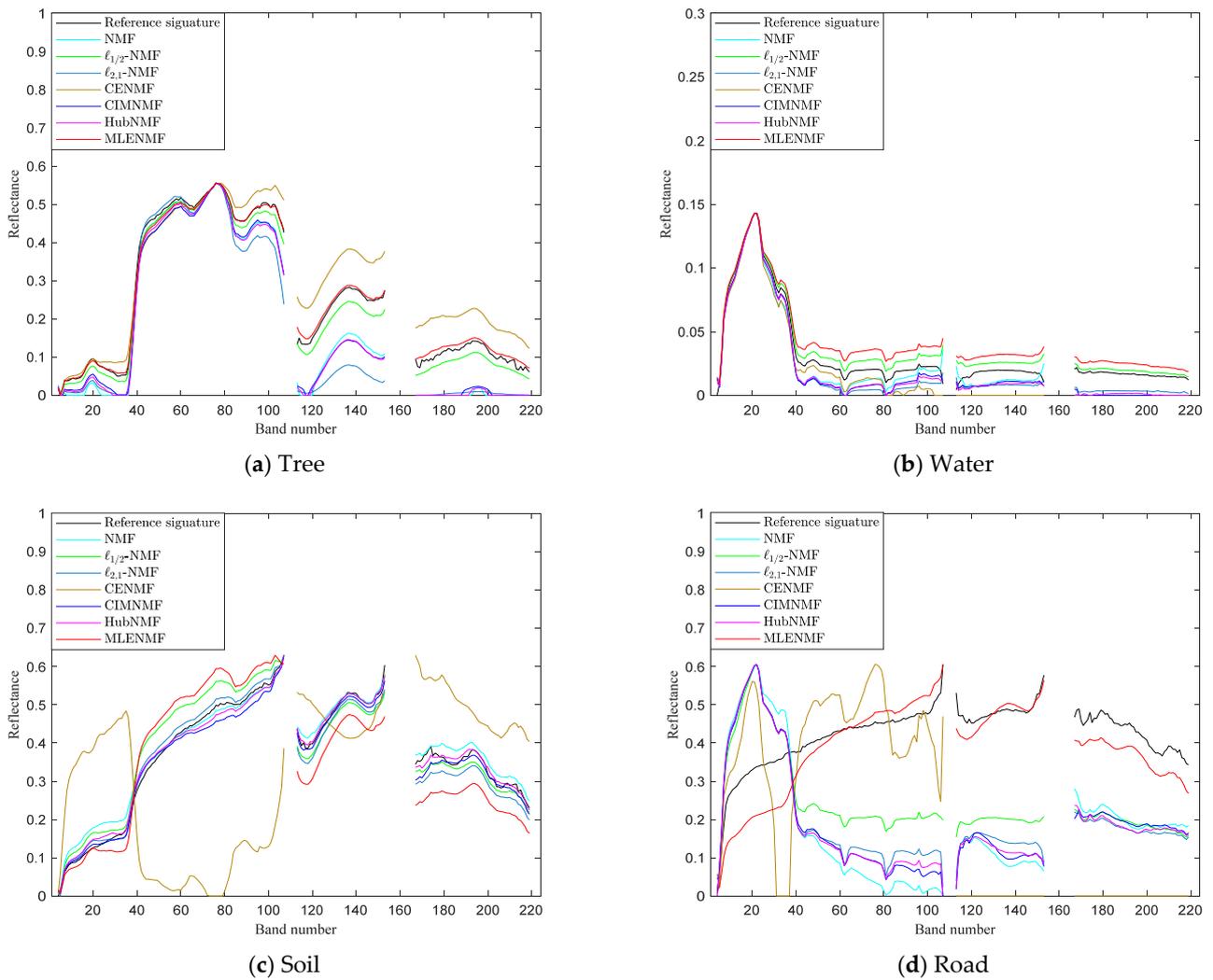
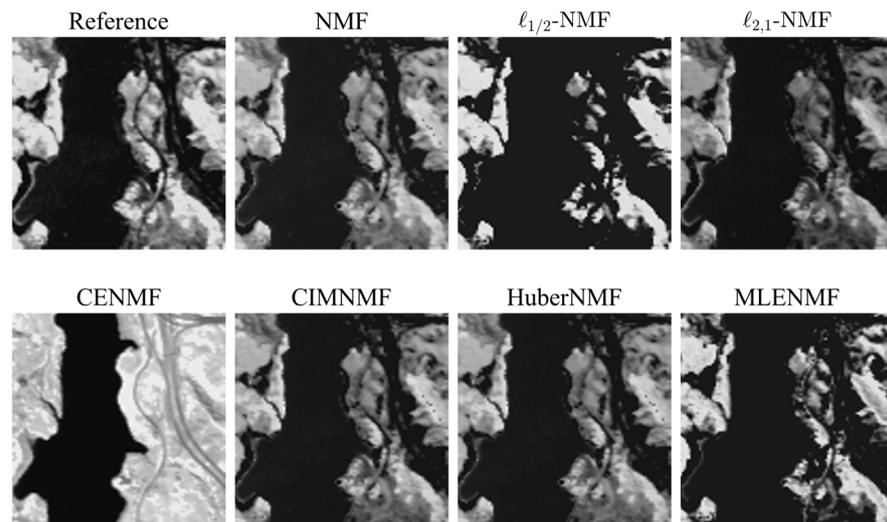


Figure 7. Comparison of endmembers estimated by different methods on Japser data for different materials: (a) Tree, (b) Water, (c) Soil, (d) Road.



(a) Tree

Figure 8. Cont.

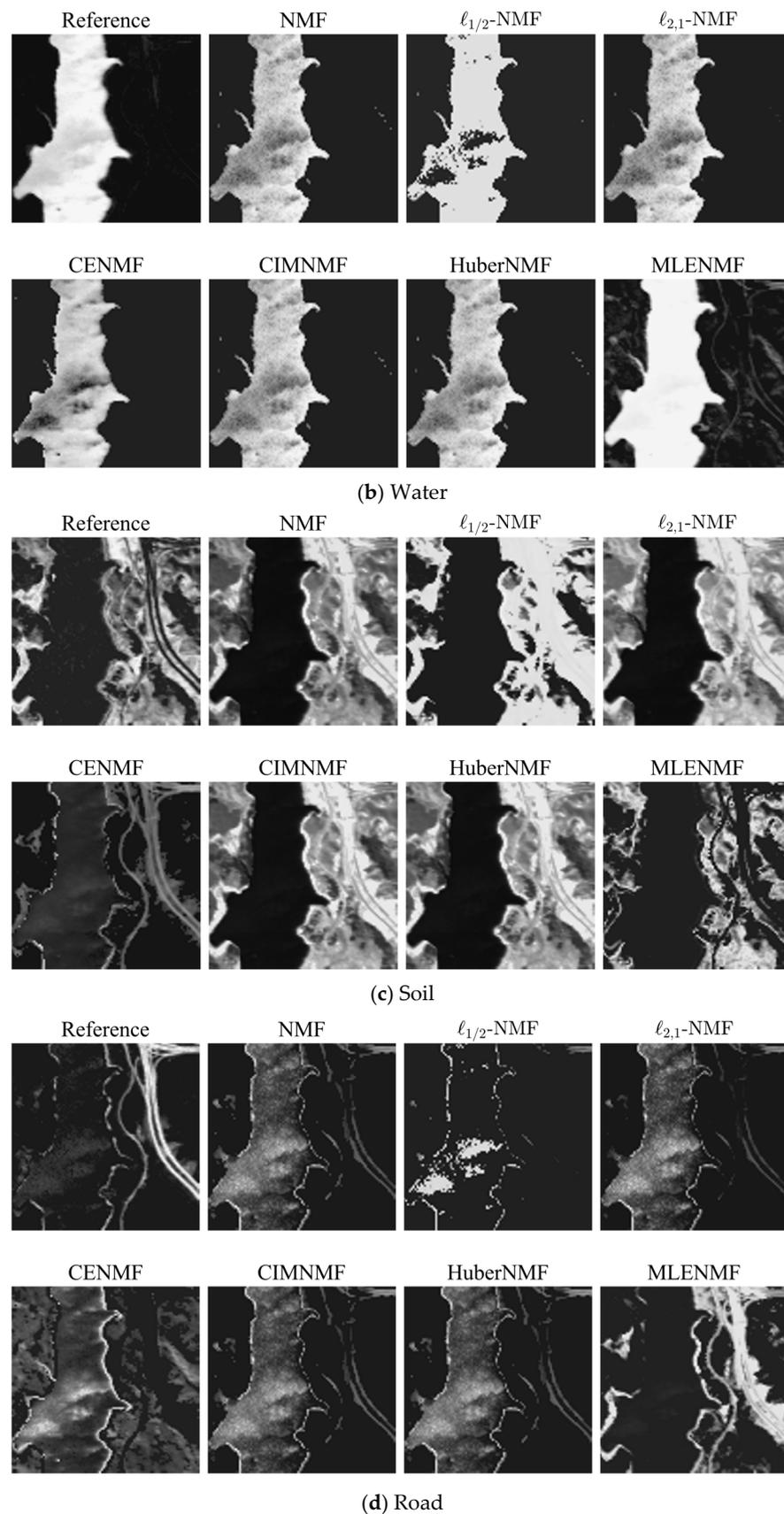


Figure 8. Comparison of abundance estimated by different methods on Japser data for different materials: (a) Tree, (b) Water, (c) Soil, (d) Road.

5. Discussion

As described in Section 4.2, τ is the (100ζ) -th percentile of residual vector $\tilde{e} = [e_1^2, \dots, e_M^2]^T$, and $\gamma = c/\tau$, $c \in (0, 10]$, $\zeta \in (0, 1]$. By tuning the parameters c and ζ , the MLE objective function in Equation (26) can be truncated, as shown in Figure 9. Parameter c and ζ control the decreasing rate and the location of truncation point, respectively. The larger the value of c , the greater the degree of truncation. The smaller the value of ζ , the more forward the position of the truncation point. As shown in Figure 9, when the noise or residual is large, it is better to choose a larger c and a smaller ζ that truncates the weight of larger residuals to a constant (seeing the red dotted line).

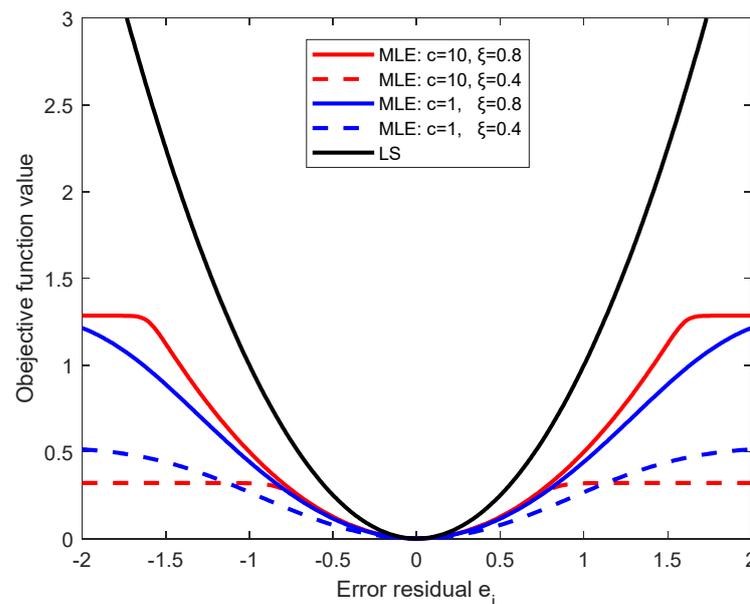


Figure 9. Comparison of MLE objective function and LS under different parameters.

We take the Urban data set as an example to show the effect of parameters c and ζ . Figure 10 shows the SAD results of MLENMF on Urban data with 210 bands. The results in Figure 10a are obtained by fixing $\zeta = 0.4$ and changing c in the set $\{0.1, 0.5, 1, 2, 5, 10\}$. When ζ is fixed, larger c values correspond to better unmixing results. As shown in Figure 9, c affects the degree of truncation. If choosing a large c , the weight of large errors can be truncated to a constant (e.g., the objective function values are constant for errors larger than 1.5, showing as the red solid line in Figure 9). As their objective function values are constant, they have no influence on the model. For Urban data with all 210 bands, MLENMF with a larger c can effectively alleviate the effect of noisy bands. By fixing $c = 10$ and changing ζ in the set $\{0.1, 0.2, 0.4, 0.6, 0.8, 1\}$, Figure 10b shows the SAD of MLENMF versus parameter ζ . It is better to set the parameter ζ in the interval $[0.4, 0.8]$ when c is fixed. Parameter ζ determines the ratio of inliers. As the data contains noisy bands, the value of ζ should be less than 1.

When the known noisy bands on the Urban data are removed, the experimental results on Urban data with 162 bands are obtained at fixing $\zeta = 0.8$ and $c = 1$, respectively. The results are shown in Figure 11. From Figure 11a, we can see that the proposed MLENMF is not sensitive to parameter c because different c values generate similar results for small errors in the case of low noise or no noise data as shown in Figure 9. From Figure 11b, the best result is achieved at $\zeta = 1$, which means that the data points are almost inliers.

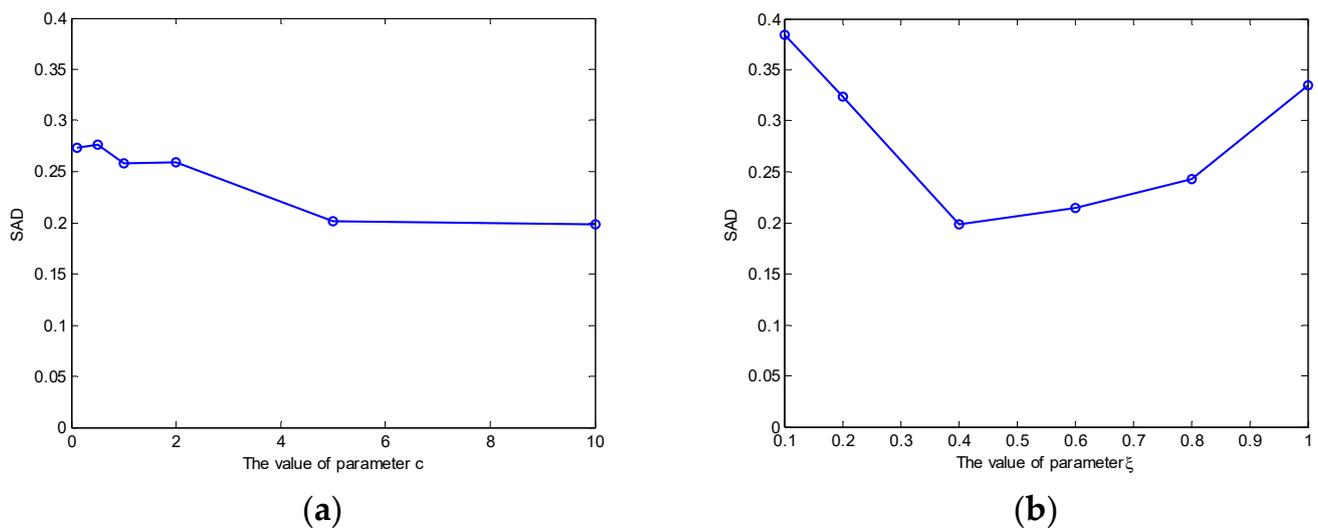


Figure 10. The SAD results under different parameter settings on Urban data with 210 bands. (a) SAD versus c at $\zeta = 0.4$. (b) SAD versus ζ at $c = 10$.

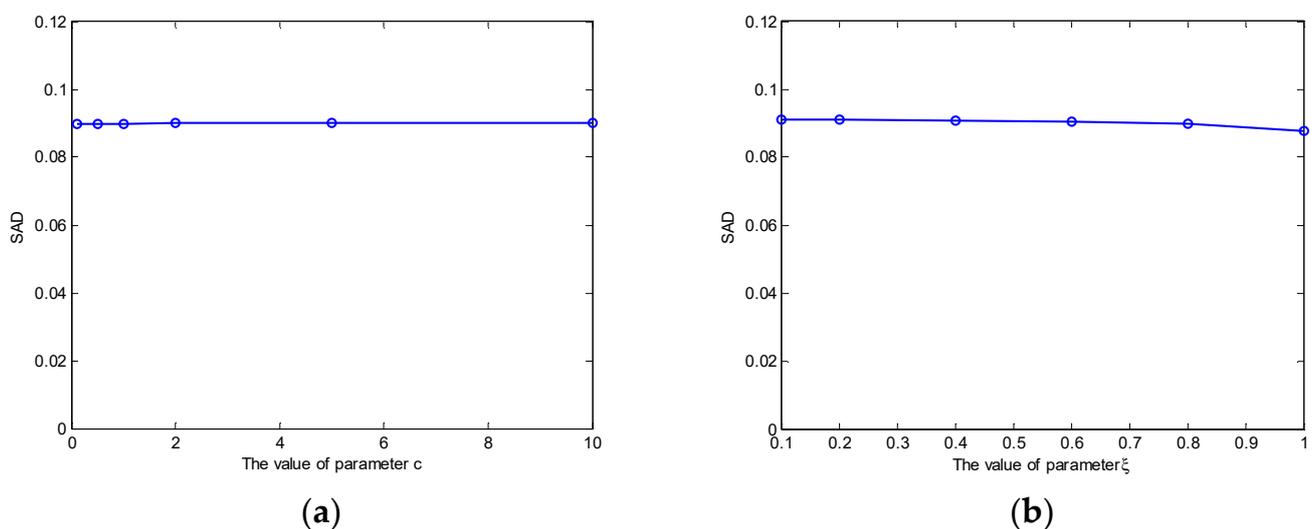


Figure 11. The SAD results under different parameter settings on Urban data with 162 bands. (a) SAD versus c at $\zeta = 0.8$. (b) SAD versus ζ at $c = 1$.

The above analysis recommends setting the parameter ζ in the interval $[0.4, 0.8]$. For data with heavy noise, ζ can be set to be a small value, such as $\zeta = 0.4$. Parameter c is chosen in the interval $[1, 10]$. For data with heavy noise, it can set $c = 10$. Otherwise, a moderate value $c = 1$ is recommended.

6. Conclusions

This paper proposes a maximum likelihood estimation-based nonnegative matrix factorization (MLENMF) model for hyperspectral unmixing. The proposed MLENMF employs an MLE-like loss function that replaces the least-squares loss function in the NMF model. The MLE-like loss is a robust loss, which can truncate the objective function value of noise and can reduce their negative effects on the unmixing model. Experimental results on a simulated data and two real data sets (Urban and Jasper) show that the proposed MLENMF model has obvious noise suppression effect and can obtain more accurate unmixing results. In the current model, it assumes that different bands are independent and then an MLE solution can be deduced. Notwithstanding, in practice the assumption of band independence is not generally valid. Taking also into account the dependence

between different bands, improved the unmixing performance may result. However, this issue deserves further research. In addition, parameter selection is a key problem for the unmixing model. The cross-validation strategy can be considered for parameter selecting, such as dividing the whole hyperspectral image into two disjoint subimages, one for training and another for testing, and then performing cross-validation to automatically select the parameters. This also deserves research in the future.

Author Contributions: Conceptualization, Q.J., J.P., Y.D., Y.S. and M.Y.; Methodology, Q.J., J.P., Y.D. and M.Y.; Software, Q.J., J.P. and Y.D.; Validation, Q.J., Y.D., Y.S. and M.Y.; Formal Analysis, Q.J., Y.D., Y.S. and M.Y.; Investigation, Q.J., Y.D. and M.Y.; Resources, J.P. and Y.S.; Data Curation, Q.J., J.P. and Y.D.; Writing—Original Draft Preparation, Q.J., Y.D. and M.Y.; Writing—Review and Editing, Q.J., J.P., Y.D. and Y.S.; Visualization, Q.J., Y.D. and Y.S.; Supervision, J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant Nos.61871177, 11771130, and by the National Key Research and Development Program of China (No. 2020YFA0714200).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.M.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [[CrossRef](#)]
2. Zhou, Y.; Peng, J.; Chen, C.L. Philip. Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095. [[CrossRef](#)]
3. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78. [[CrossRef](#)]
4. 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 Observ. Remote Sens.* **2012**, *5*, 354–379. [[CrossRef](#)]
5. Heylen, R.; Parente, M.; Gader, P. A review of nonlinear hyperspectral unmixing methods. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 1844–1868. [[CrossRef](#)]
6. Peng, J.; Zhou, Y.; Sun, W.; Du, Q.; Xia, L. Self-paced nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1501–1515. [[CrossRef](#)]
7. 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](#)]
8. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Trans. Image Process.* **2018**, *28*, 1923–1938. [[CrossRef](#)]
9. Danie Lee, D.; Sebastian, S.H. Algorithms for non-negative matrix factorization. In Proceedings of the 13th International Conference on Neural Information Processing Systems, Denver, CO, USA, 27 November–2 December 2020; pp. 556–562.
10. 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](#)]
11. Jia, S.; Qian, Y. Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 161–173. [[CrossRef](#)]
12. Liu, X.; Xia, W.; Wang, B.; Zhang, L. An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 757–772. [[CrossRef](#)]
13. Qian, Y.; Jia, S.; Zhou, J.; Robles-Kelly, A. Hyperspectral unmixing via L1/2 sparsity-constrained nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4282–4297. [[CrossRef](#)]
14. Lu, X.; Wu, H.; Yuan, Y.; Yan, P.; Li, X. Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2815–2826. [[CrossRef](#)]
15. Wang, N.; Du, B.; Zhang, L. An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2013**, *6*, 554–569. [[CrossRef](#)]
16. Peng, J.; Sun, W.; Li, W.; Li, H.; Meng, X.; Ge, C.; Du, Q. Low-rank and sparse representation for hyperspectral image processing: A review. *IEEE Geosci. Remote Sens. Mag.* **2021**. [[CrossRef](#)]

17. He, W.; Zhang, H.; Zhang, L. Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 4267–4279. [[CrossRef](#)]
18. Du, L.; Li, X.; Shen, Y. Robust nonnegative matrix factorization via half-quadratic minimization. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 201–210.
19. Wang, Y.; Pan, C.; Xiang, S.; Zhu, F. Robust hyperspectral unmixing with correntropy-based metric. *IEEE Trans. Image Process.* **2015**, *24*, 4027–4039. [[CrossRef](#)]
20. Huang, R.; Li, X.; Zhao, L. Spectral-spatial robust nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8235–8254. [[CrossRef](#)]
21. Kong, D.; Ding, C.; Huang, H. Robust nonnegative matrix factorization using L21-norm. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow Scotland, UK, 24–28 October 2011; pp. 673–682.
22. Peng, J.; Sun, W.; Jiang, F.; Chen, H.; Zhou, Y.; Du, Q. A general loss based nonnegative matrix factorization for hyperspectral unmixing. *IEEE Geosci. Remote Sens. Lett.* **2020**. [[CrossRef](#)]
23. Peng, J.; Li, L.; Tang, Y.Y. Maximum likelihood estimation based joint sparse representation for the classification of hyperspectral remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 1790–1802. [[CrossRef](#)]
24. Yang, M.; Zhang, L.; Yang, J.; Zhang, D. Robust sparse coding for face recognition. In Proceedings of the CVPR, Colorado Springs, CO, USA, 20–25 June 2011; pp. 625–632.
25. Xu, Z.; Guo, H.; Wang, Y.; Zhang, H. Representative of L1/2 regularization among L_q ($0 < q < 1$) regularizations: An experimental study based on phase diagram. *Acta Autom. Sin.* **2012**, *38*, 1225–1228.
26. Zhang, J.; Jin, R.; Yang, Y.M.; Hauptmann, A.G. Modified logistic regression: An approximation to SVM and its applications in large-scale text categorization. In Proceedings of the 20th International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 888–895.
27. Li, X.; Lu, Q.; Dong, Y.; Tao, D. Robust subspace clustering by Cauchy loss function. *IEEE Trans. Neur. Netw. Learn. Syst.* **2019**, *30*, 2067–2078. [[CrossRef](#)] [[PubMed](#)]
28. Feng, X.; Li, H.; Li, J.; Du, Q.; Plaza, A.; Emery, W. Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6245–6257. [[CrossRef](#)]