A Modified Huber Nonnegative Matrix Factorization Algorithm for Hyperspectral Unmixing

Ziyang Guo 🅐, Anyou Min 🅐, Bing Yang 🅐, Junhong Chen, and Hong Li 🅐, *Member, IEEE*

Abstract—Hypersepctral unmixing (HU) has been one of the most challenging tasks in hyperspectral image research. Recently, nonnegative matrix factorization (NMF) has shown its superiority in hyperspectral unmixing due to its flexible modeling and little prior requirement. But most NMF algorithms tend to use least square function as the objective, which is sensitive to outliers and different kinds of noise. In this article, we propose a modified Huber (mHuber) NMF model to achieve robustness to outliers and different kinds of noise. Under this robust model, we accelerate the half-quadratic optimization algorithm by replacing multiplicative updating rule with a projected nonlinear conjugated gradient rule, which achieves much faster convergence rate. Moreover, a new tuning parameter, rather than a fixed one, is given to adapt to mHuber loss function. Finally, we perform algorithm analysis and experiments in the synthetic and real-world datasets, which confirms the effectiveness and superiority of the proposed method when compared with several state-of-the-art NMF methods in HU.

Index Terms—Modified Huber (mHuber), nonnegative matrix factorization (NMF), projected conjugated gradient, tuning parameter.

I. INTRODUCTION

HYPERSPECTRAL image is often a mixing of the spectrums from different substances, which is acquired by hyperspectral imaging sensor. A significant task in hyperspectral image analysis is hyperspectral unmixing (HU). It decomposes any pixel into the combination of an endmember matrix and an abundance matrix. Due to its high-spectral resolution, low-spatial resolution, and spectral redundancy, HU has also been a challenging work in recent hyperspectral image research [1]. Recent research has generally divided the HU task into two models, i.e., linear mixing model (LMM) and nonlinear mixing model (NLMM). Since nonlinear mixing model incorporates the mutual impact of two or more substances, which leads to variety and complexity of the model, most algorithms for HU in these years are based on LMM, which ignores the mutual impact of endmembers. Existing HU algorithms on LMM can be classified into four categories: Geometry-based, sparsity regression-based, statistics-based, and deep learning- based.

Geometry-based methods often assume the existence of "pure pixel" in the HU, whose typical algorithms include vertex component analysis (VCA) [2], N-Findr [3], pixel purity index [4], and minimum volume-based methods like simplex identification via split augmented Lagrangian (SISAL) [5] and minimum volume simplex analysis (MVSA) [6]. But in real cases, this assumption is not always true. Sparse-regression-based methods assume that each pixel can be represented by the pure spectrums from the known spectral libraries. So the HU problem seeks to find an optimal linear combination of endmembers. Zheng *et al.* [7] proposed a weighted sparse regression model and the corresponding two iterative reweighted algorithm, which combines the advantages of alternating direction method of multipliers and iterative reweighting procedure. Fu *et al.* [8] proposed a sparse regression HU algorithm based on self-dictionary multiple measurement vector (SD-MMV). However, this method relies much on a prior dictionary of endmembers. Deep learning-based methods have received increasing attention with the improvement of deep learning, Savas [9] developed a two-staged autoencoder network by introducing additional layers and a projection metric to achieve an optimum solution. Statistics-based methods have already been a widely used method for HU problem. Popular algorithms including independent component analysis (ICA) [10], nonnegative matrix factorization (NMF) [11], [12] have been proved both theoretically and practically to be powerful in dealing with HU problem. Among these algorithms, NMF achieves public attention due to its advantages on effectiveness, little prior requirement and flexible modeling. Lee *et al.* [13] proposed a brief multiplicative rule to decompose a nonnegative matrix into two nonnegative ones simultaneously. In this HU problem, the two achieved nonnegative matrices can be seen as the spectral endmember matrix and the abundance matrix. In [13], it shows that the two matrices can also be regarded as a sparse representation due to their nonnegativity.

Now we introduce several common NMF-based methods for HU. L_1-NMF [14] added L_1 sparsity constraint to the loss function to ensure the sparsity of the solution. $L_{1/2}$-NMF [11] selected $L_{1/2}$ as the best L_q regularizer, but with more computational cost than L_1-NMF. Li [15] used a $L_{2,1}$ norm as the loss function, not only to determine the optimal endmember number but also get robust results. Since hyperspectral image is originally an image tube, naturally the tube version of NMF generates. Nonnegative tensor factorization (NTF) [16] has been developed by considering spatial information of hyperspectral image. Li *et al.* combined the advantages of NTF and NMF to preserve the local spatial structure of hyperspectral image [17].

Moreover, several existing NMF algorithms tend to use manifold learning regularization for the abundance matrix to keep its intrinsic structure. Lu *et al.* [12] incorporated manifolding regularization into $L_{1/2}$-NMF to preserve the similarity between the original image and the abundance maps. Shu *et al.* proposed a multilayer manifold sparse NMF on both spectral signature matrix and the abundance matrix [18]. Peng *et al.* proposed a general loss model for different kinds of noise [19]. Even with so many pros and achievements, many NMF models and their corresponding updating rules still suffer from several cons like instability, outliers, local minima, and low convergence rate.

A well-known knowledge is that most NMF models utilize the least square function (L_2 function) as their loss function, but it suffers from a drawback that when the reconstruction error is larger than 1, the square error is larger than the original error; while the reconstruction error is smaller than 1, it is smaller than the original error. That is to say, L_2 function is sensitive to large values such as outliers and different kinds of noise in real scenarios. The lack of robustness in existing NMF models naturally arouses extra exploration for more robust NMF models. As we know, $\boldsymbol{M}$-estimators is a significant robust technique. The common $\boldsymbol{M}$-estimators include: L_2 estimator, L_1 estimator (i.e., absolute value), $L_{2,1}$ estimator, L_p estimator (i.e., least-powers), Fair estimator, Huber estimator, cauchy estimator, and so on. Huber function is recommended for almost all situations in multivariate linear regression, since it not only incorporates the superiority of L_2 and L_1 but also shows broad applicability.

But the research of Huber function and its modified models in NMF for unmixing is far from enough. Du *et al.* [20] gave an iterative algorithm for both correntropy NMF and Huber NMF under the scheme of half-quadratic optimization, but with low convergence rate. Wang studied the truncated cauchy estimator for NMF for HU in [21], and downplayed the effect of extreme outliers, but suffered from more parameters to be learned. Peng *et al.* [22] proposed a self-paced NMF model and made comparison with several robust NMF methods by adding noise in different ways.

In this article, we propose to replace the L_2 loss of residuals by a modified Huber (mHuber) function for NMF to achieve robust matrix factorization and avoid the instability of standard Huber NMF. After adding some common regularization, we solve the new loss function by half-quadratic optimization. The existing half-quadratic optimization to solve Huber NMF adopts multiplicative updating rule (MU). Since the MU in NMF are often shown to have low convergence rate, we develop a projected nonlinear conjugated gradient algorithm under the half-quadratic scheme as an improved and fast version of the algorithm given in [20]. Then we introduce an new tuning parameter to adapt mHuber. Moreover, we give the corresponding convergence analysis and robustness analysis for our proposed model. In order to show the performance of our proposed method, we design several experiments and introduce a new indicator SFF in order to help evaluate the performance of unmixing.

The contributions of this article are listed as follows.

1) We introduce a modified version of Huber loss function into NMF model, not only to provide robustness against outliers and different kinds of noises but also to ensure stability and better performance.

2) A projected nonlinear conjugated gradient algorithm under half-quadratic scheme for updating $\boldsymbol{S}$ is proposed to make up the low convergence rate of MUs; convergence rate, computational complexity, and robustness analysis are also included.

3) A new tuning parameter is proposed to adapt the new loss function, instead of the normal fixed median method.

The rest of this article is organized as follows: Section II illustrates the basic knowledge of the LMM, Huber NMF model, and the mHuber NMF. Section III presents a mHuber NMF model, its updating algorithm and algorithm analysis. Section IV shows the experiments on synthetic data and real-world hyperspectral dataset. Section V concludes this article.

II. PRELIMINARY

A. LMM and Multiplicative Update Rule

In the hyperspectral images, each pixel is regarded as a combination of one or several pure endmembers of different substances. Regardless of the mutual impact of endmembers to the spectral reflectance, the linear mixed model can be applied to decompose a pixel of M bands from the observation into K pure endmembers and their corresponding abundance matrix. For a given pixel vector $\boldsymbol{x} = [x_1, \ldots, x_M]^T$, M denotes the number of the spectral bands. The LMM can be given as

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{\varepsilon}$$

in which $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_K] \in \mathbb{R}^{M \times K}$ is the endmember matrix, $\boldsymbol{s} = [s_1, \ldots, s_K]^T \in \mathbb{R}^K$ represents the abundance vector of pixel, k is the number of endmembers, and $\boldsymbol{\varepsilon}$ denotes an $M \times 1$ vector of the additive noise as the measurement errors. Given two constraints of $\boldsymbol{x}$ in LMM: The abundance nonnegativity constraint (i.e., $s_i \geq 0$), and the abundance sum-to-one constraint (i.e., $\sum_{i=1}^k s_i = 1$), respectively. By matrix notation, the LMM for N pixels in the image can rewritten as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \mathcal{E} \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{M \times N}$, $\boldsymbol{A} \in \mathbb{R}^{M \times K}$, $\boldsymbol{S} \in \mathbb{R}^{K \times N}$, and $\mathcal{E} \in \mathbb{R}^{M \times N}$ represent the observed data, the endmember reflectance, the endmember abundances, and noise, respectively. In most cases, the noise $\mathcal{E}$ is assumed to be independent and identically distributed Gaussian noise. But there will also be some other kinds of noise in it, such as heavy-tailed noise, outliers.

The widely used optimization algorithm for this model is multiplicative iterative algorithm which applies a multiplicative update rule in NMF as follows:

$$\boldsymbol{A} \leftarrow \boldsymbol{A}. * \boldsymbol{X}\boldsymbol{S}^T./\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^T \tag{2}$$

$$\boldsymbol{S} \leftarrow \boldsymbol{S}. * \boldsymbol{A}^T\boldsymbol{X}./(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{S}. \tag{3}$$

B. Huber Criterion

To be robust to the outliers or noise corruptions, the loss function based on Huber's criterion is adopted, which is a critical theory as a $\boldsymbol{M}$-estimator in robust analysis. It has lowered the sensitivity of the least square function by replacing the square

term into L_1 term in suitable conditions. As Huber function is so satisfactory that it has been recommended for almost all situations [23]. Then, the Huber loss is given by

$$\mathbb{H}\left(\boldsymbol{E}_{ij}\right) = \begin{cases} \frac{1}{2}\boldsymbol{E}_{ij}^2, & \|\boldsymbol{E}_{ij}\|_2 \le c \\ c\|\boldsymbol{E}_{ij}\|_2 - \frac{1}{2}c^2, & \text{else.} \end{cases} \quad (4)$$

The corresponding derivative ψ_{ij} of the Huber loss are given by

$$\psi_{ij} = \begin{cases} \boldsymbol{E}_{ij}, & |\boldsymbol{E}_{ij}| \le c \\ c \cdot \frac{\boldsymbol{E}_{ij}}{|\boldsymbol{E}_{ij}|}, & \text{else} \end{cases} \quad (5)$$

in which $\boldsymbol{E}_{ij} = \boldsymbol{X}_{ij} - \sum_{k=1}^{K} \boldsymbol{A}_{ik}\boldsymbol{S}_{kj}$. When $|\boldsymbol{E}_{ij}|$ is small, the Huber Loss acts as a quadratic increase; while it acts as a linear increase when $|\boldsymbol{E}_{ij}|$ is large. c is a cutoff parameter, which is set to control the level of robustness. It is easy to see that the loss function reduces to least-squares pattern as $c \to \infty$, while it reduces to least absolute deviation when $c \to 0$. In [23], Huber proposed a fixed c value of 1.345 to achieve 95% efficiency at $N(0,1)$, $N(0,1)$ denotes the standard normal distribution. But there are other adaptive methods to get c automatically. Moreover, from (5), we can get that $|\psi_{ij}| \le c$.

Even Huber function gives remarkable performance, it encounters difficulties because of a lack of stability in the gradient values. To be specific, it may have two discontinuous second derivative 1 and 0 when $\boldsymbol{E}_{ij}$ varies during different iterations. The modified form is less sensitive than standard Huber loss and benefits from a smooth second derivative. Then a mHuber model is proposed in [24], which guarantees a unique solution of the new Huber Loss in, i.e.,

$$\mathbb{H}_{\mathcal{M}}\left(\boldsymbol{E}_{ij}\right) = \begin{cases} c^2\left[1 - \cos\left(\boldsymbol{E}_{ij}/c\right)\right], & |\boldsymbol{E}_{ij}|/c \le \pi/2, \\ c\left|\boldsymbol{E}_{ij}\right| + c^2\left(1 - \pi/2\right), & \text{else} \end{cases}. \quad (6)$$

The corresponding first derivative is

$$\psi_{\mathcal{M}}\left(\boldsymbol{E}_{ij}\right) = \begin{cases} c\sin\left(\boldsymbol{E}_{ij}/c\right), & |\boldsymbol{E}_{ij}|/c \le \pi/2, \\ c \cdot \text{sgn}\left(\boldsymbol{E}_{ij}/c\right), & \text{else} \end{cases}. \quad (7)$$

The corresponding second derivative is

$$\psi_{\mathcal{M}}'\left(\boldsymbol{E}_{ij}\right) = \begin{cases} \cos\left(\boldsymbol{E}_{ij}/c\right), & |\boldsymbol{E}_{ij}|/c \le \pi/2, \\ 0, & \text{else} \end{cases}. \quad (8)$$

III. ALGORITHM FOR MODIFIED HUBER-BASED NMF UNMIXING

A. Modified Huber-Based NMF

Now, we introduce the mHuber loss into the NMF model in the HU, which is defined as follows:

$$f(A, S) = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \mathbb{H}_{\mathcal{M}}\left(\boldsymbol{E}_{ij}\right) \quad (9)$$

in which $\boldsymbol{E}_{ij} = \boldsymbol{X}_{ij} - (\boldsymbol{AS})_{ij}$.

In most hyperspectral images, most pixels will not consist of all endmembers, especially when the number of endmembers K is relatively large. So we can expect low rank and sparseness of reconstruction image. Naturally, we propose the nuclear norm of the reconstruction to ensure its low rank and sparseness. Moreover, graph regularization [12] is often added into unmixing

model to build the link between the inner manifold structures of the abundance and the observation. This is derived from the assumption that pixels that similar to each other tend to be close in the new abundance space. Thus, the $f(A, S)$ can be transformed into

$$f(A, S) = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \mathbb{H}_{\mathcal{M}}\left((\boldsymbol{X} - \boldsymbol{AS})_{ij}\right)$$
$$+ \lambda_1\|\boldsymbol{AS}\|_* + \lambda_2 Tr\left(\text{SLS}^T\right) \quad (10)$$

where $Tr(\cdot)$ is the trace. $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{G}$ is an $N \times N$ laplacian matrix, in which $\boldsymbol{D}$ is a diagonal matrix whose entries are column sums of $\boldsymbol{G}$, $\boldsymbol{G}$ is the weight matrix of graph calculated by common methods, such as heat kernel.

The nuclear norm has the following relaxation as $\|\boldsymbol{AS}\|_* \to 1/2(\|\boldsymbol{A}\|_F^2 + \|\boldsymbol{S}\|_F^2)$, the second term of our loss can be transformed into $\lambda_1(\|\boldsymbol{A}\|_F^2 + \|\boldsymbol{S}\|_F^2)$, then the new $f(\boldsymbol{A}, \boldsymbol{S})$ will be

$$f(A, S) = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \mathbb{H}_{\mathcal{M}}\left((\boldsymbol{X} - \boldsymbol{AS})_{ij}\right)+$$
$$\lambda_1\left(\|\boldsymbol{A}\|_F^2 + \|\boldsymbol{S}\|_F^2\right) + \lambda_2 Tr\left(\text{SLS}^T\right). \quad (11)$$

Then the abundance nonnegativity constraint (ANC) and abundance sum-to-one constraint (ASC) constraints of matrix form are denoted by

$$\boldsymbol{A} \ge 0, \boldsymbol{S} \ge 0, \ 1_K^T\boldsymbol{S} = 1_N^T$$

where 1_K and 1_N denote all one column vectors of size K and N.

Till now, our robust mHuber NMF model is built with the first term robust to outliers, the second term guaranteeing the low rank and sparseness, the third term keeps the intrinsic manifold structure of abundance.

B. Updating Rules

Since the cost function f of our mHuber NMF model is not convex to $\boldsymbol{A}$ and $\boldsymbol{S}$ together, the classical MUs for NMF may confront difficulties. Due to the higher efficiency of half-quadratic minimization than gradient-based methods for this kind of problem [25], we adopt it here to optimize our model. To realize it, we introduce additional auxiliary function $/varPhi(\boldsymbol{W}_{ij})$ and the conjugate variable $\boldsymbol{W}_{ij}$ of $\boldsymbol{E}_{ij}$, which reformulates our loss function into an augmented loss function with a quadratic term in an enlarged parameter space. Then the following equation holds:

$$\mathcal{L} = \sum_{i,j=1}^{M,N} \left(\frac{1}{2}\boldsymbol{W}_{ij}\boldsymbol{E}_{ij}^2 + /varPhi\left(\boldsymbol{W}_{ij}\right)\right)$$
$$+ \lambda_1\left(\|\boldsymbol{A}\|_F^2 + \|\boldsymbol{S}\|_F^2\right) + \lambda_2 Tr\left(\text{SLS}^T\right). \quad (12)$$

Via this, we have the augmented objective function

$$\underset{\boldsymbol{A},\boldsymbol{S},\boldsymbol{W},\boldsymbol{c}}{\arg\min}\mathcal{L} = \underset{\boldsymbol{A},\boldsymbol{S}}{\arg\min}f\left(\boldsymbol{A}, \boldsymbol{S}\right). \quad (13)$$

The half-quadratic scheme will be a three-step scheme.

TABLE I
HYBRID STRATEGY FOR β_{k+1}

	β_{k+1}
Hybrid Strategy	$\beta_{k+1}^{HS} = \dfrac{(vec(\partial\mathcal{L}/\partial\boldsymbol{S}_{k+1}))^T y_k}{vec(\Xi_k^T)y_k}$ $\beta_{k+1}^{DY} = \dfrac{\|vec(\partial\mathcal{L}/\partial\boldsymbol{S}_{k+1})\|^2}{vec(\Xi)_k^T y_k}$ $\beta_{k+1} = \max\left\{0, \min\left(\beta_{k+1}^{HS}, \beta_{k+1}^{DY}\right)\right\}.$

1) When $\boldsymbol{A}$, $\boldsymbol{S}$, and c are fixed, the minimization of the (12) becomes a convex problem to $\boldsymbol{W}$. The optimal $\boldsymbol{W}$ is given by

$$W_{ij} = \frac{\mathbb{H}_{\mathcal{M}}^{(\boldsymbol{E}_{ij})}}{\boldsymbol{E}_{ij}}. \tag{14}$$

It is easy to see that when a large error exists, $\boldsymbol{W}_{ij}$ tends to be small, with larger $\boldsymbol{W}_{ij}$ when small error exists.

2) When $\boldsymbol{W}$ and c are fixed, our loss becomes a weighted least square NMF problem, then multiplicative rules to alternatively iterate $\boldsymbol{A}$ and $\boldsymbol{S}$ has been used in standard Huber NMF in [20]. But multiplicative rules is often criticized for its slow convergence rate. Thus, we propose a new nonlinear projected conjugated gradient rule here to update $\boldsymbol{A}$ and $\boldsymbol{S}$.

3) When $\boldsymbol{A}$, $\boldsymbol{S}$, and $\boldsymbol{W}$ are fixed, c can be determined by several methods.

The detail of our proposed algorithm is illustrated as follows.

1) Update $\boldsymbol{W}$: When $\boldsymbol{A}$, $\boldsymbol{S}$, and c are fixed, The weight matrix $\boldsymbol{W}$ can be updated by

$$\boldsymbol{W}_{\mathcal{M}}(\boldsymbol{E}_{ij}) = \begin{cases} \frac{c*\sin(\boldsymbol{E}_{ij}/c)}{\boldsymbol{E}_{ij}}, & |\boldsymbol{E}_{ij}|/c \leq \pi/2 \\ \frac{c \cdot sgn(\boldsymbol{E}_{ij}/c)}{\boldsymbol{E}_{ij}}, & \text{else.} \end{cases} \tag{15}$$

2) Updating $\boldsymbol{S}$: When the weight matrix $\boldsymbol{W}$ is fixed, the optimization reduced to a weighted NMF problem, then the common MUs can be applied. But because of its low convergence rate, we develop a nonlinear projected conjugated gradient method in the iteration process of half-quadratic framework for both $\boldsymbol{S}$. First taking the partial derivative with respect to $\boldsymbol{S}$ yields

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{S}} = -\boldsymbol{A}^T(\boldsymbol{W} \otimes \boldsymbol{X}) + \boldsymbol{A}^T(\boldsymbol{W} \otimes (\boldsymbol{AS})) + \lambda_1 \boldsymbol{S} + \lambda_2 \boldsymbol{S} L \tag{16}$$

in which $\otimes$ means the Hadamard product. Then the iteration direction will be

$$\Xi_{k+1} = -\partial\mathcal{L}/\partial\boldsymbol{S}_{k+1} + \beta_{k+1}\Xi_k. \tag{17}$$

Now in terms of the choice of β_{k+1}, the Dai–Yuan (DY) formulas [26] ensures strong convergence properties but may suffer from a tiny step and bad search direction. The Hestenes–Stiefel (HS) formulas [27] often guarantees better search direction and step size, but with poor convergence. Thus we select a hybrid strategy to choose β_{k+1}, which are given in Table I. Before given the strategy, we define y_k as $y_k = vec(\frac{\partial\mathcal{L}}{\partial\boldsymbol{S}_{k+1}} - \frac{\partial\mathcal{L}}{\partial\boldsymbol{S}_k})$.

After getting the iteration direction $/varXi_{k+1}$, we need to decide the step length α_{k+1}. In order to avoid bad-scaled search direction that may happen in conjugated gradient method, an expedient is to integrate current information for an initial guess

of α_{k+1} as

$$\alpha_{\text{init}} = \frac{\alpha_k(\text{vec}(\partial\mathcal{L}/\partial\boldsymbol{S}_{k+1}))^T \text{vec}(/varXi_k)}{(\text{vec}(\partial\mathcal{L}/\partial\boldsymbol{S}_k))^T \text{vec}(/varXi_{k+1})}. \tag{18}$$

When $k = 0$, the initial step size is $1/\|vec(\partial\mathcal{L}/\partial\boldsymbol{S}_0)\|_\infty$. Then by utilizing an inexact line search method satisfying the weak Wolfe condition:

(i) $f(\boldsymbol{S}_{k+1}) - f(\boldsymbol{S}_k) \leq c_1\alpha_k/varXi_k f'(\boldsymbol{S}_k)$;

(ii) $/varXi_k^T f'(\boldsymbol{S}_{k+1}) \geq c_2/varXi_k f'(\boldsymbol{S}_k)$;

where $0 < c_1 < c_2 < 1$; (i) is the Armoji condition for sufficient descent, while (ii) is the curvature condition to avoid very short step; thus we have

$$\boldsymbol{S}_{k+1} = \boldsymbol{S}_k + \alpha_{k+1}/varXi_{k+1}. \tag{19}$$

As the above updating cannot ensure nonnegativity of $\boldsymbol{S}$, we need a project operation by

$$\boldsymbol{S}_{k+1} = \max\{0, \boldsymbol{S}_{k+1}\}. \tag{20}$$

The stopping criterion of the iteration is

$$\frac{\|vec(\partial\mathcal{L}/\partial S_{k+1})\|_\infty}{(1 + \|vec(\partial\mathcal{L}/\partial S_0)\|_\infty)} < \epsilon \tag{21}$$

where $\|\cdot\|_\infty$ denotes the infinity norm, and $\epsilon > 0$ denotes a specified threshold (default value is 10^{-6}).

3) Updating $\boldsymbol{A}$: Terming to updating $\boldsymbol{A}$, we have two updating rules. Obviously, when fixing $\boldsymbol{S}$, the partial derivative of $\boldsymbol{A}$ is given as

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{A}} = -(\boldsymbol{W} \otimes \boldsymbol{X})\boldsymbol{S}^T + (\boldsymbol{W} \otimes (\boldsymbol{AS}))\boldsymbol{S}^T + \lambda_1\boldsymbol{A}. \tag{22}$$

The original MU of classical NMF is proposed as follows:

$$\boldsymbol{A} \leftarrow \boldsymbol{A}.*(\boldsymbol{W} \otimes \boldsymbol{X})\boldsymbol{S}^T./\left[(\boldsymbol{W} \otimes \boldsymbol{AS})\boldsymbol{S}^T + \lambda_1\boldsymbol{A}\right]. \tag{23}$$

4) Updating c: c is critical in determining the threshold of inliers and outliers. Huber showed that when $c = 1.345$, one gets approximately 95% efficiency in the Gaussian cases compared to the optimal estimators. c used in [20] is the median of reconstruction errors, which is defined as $c = \text{med}(|\boldsymbol{E}_{ij}|)$. But they are not appropriate in all cases. c is the tuning constant that should be used to adjust the efficiency of the resulting estimators for specific distributions.

Therefore, we adopt Holland and Welsch's rule [28], which suggests that c is chosen as the product of a constant and the median of the absolute value of reconstruction error. Such as for in standard Huber function, c can be

$$c = 1.345 \cdot \text{median}(|\boldsymbol{E}_{ij}|). \tag{24}$$

Since our new loss function is a mHuber, we change the 1.345 by 1.2107 to get 95% asymptotic efficiency on the standard normal distribution, i.e.,

$$c = 1.2107 \cdot \text{median}(|\boldsymbol{E}_{ij}|). \tag{25}$$

5) Stopping Criterion: At last, the whole algorithm will terminate when reaching the maximum iteration number and the reconstruction error is less than a threshold (defaulted value is 10^{-4}).

Algorithm 1: Projected Nonlinear CG algorithm of S.

Input: X, S, S, λ_1, λ_2, c, γ, iternum

Output: Estimated endmember matrix S;

1 Initialize S_0, S_0, λ_1 and λ_2 with default or random values, $k = 0$, $\gamma = 0.5$;

2 Evaluate $f(S_0)$, $\partial\mathcal{L}/\partial S_0$, and Ξ_0;

3 Determine a_0 by $1/\|vec(\partial\mathcal{L}/\partial S_0)\|_\infty$;

4 Determine $\partial\mathcal{L}/\partial S_{k+1}$ by (16);

5 Determine β_{k+1} by a hybrid strategy as Table I;

6 Determine Ξ_{k+1} by (17);

7 $\alpha_k = \alpha_{init}$ by (18), compute S_{k+1} by (19) and (20);

8 **repeat**

9   **else** choose $a_k = \gamma * a_{kguess}$;

10   Obtain S_{k+1} by (19) and (20);

11 **until** a_k *satisfies the weak Wolfe conditions*;

12 Let $S \leftarrow S_{k+1}$;

    return S.

Algorithm 2: Improved Half-quadratic Algorithm for mHuber NMF.

Input: X, A_0, S_0, λ_1, λ_2, *stopping criterion*

Output: Estimated endmember matrix A and abundance matrix S;

1 Initialize A_0, S_0, λ_1, λ_2 with default or random values;

2 Evaluate initial E_0;

3 **repeat**

4   Update W by applying Eq.(15);

5   Update S by applying the derivative of S via Eq.(16) and projected conjugated gradient algorithm 1;

6   Update A by applying Eq.(23);

7   Update c by utilizing median sorting as Eq.(24);

8 **until** *The stopping criterion reaches*;

    return A, S.

C. Algorithm Analysis

1) Implementation Issues: There are four issues to be discussed in implementing our proposed method. The first issue is the initialization of A and S. We use common NMF algorithms for HU, such as a) random initialization, b) Nfindr-FCLS initialization, and c) VCA-FCLS initialization. As for random initialization, it randomly generates initial A by choosing columns from the observation, and randomly chooses any values between 0 and 1 to be the entries of initial S. VCA and Nfindr are used to initialize A. FCLS can be adopted in initialize S. The second issue is the coefficient λ_1 and λ_2 of low rank regularization and manifold regularization, which in general are decided by experience. Third, the ASC constraint is often incorporated by a simple simplex method proposed in [29], the matrix X and A are replaced by $\hat{X}$ and $\hat{A}$ in our inputs, respectively, i.e.,

$$\hat{X} = \begin{bmatrix} X \\ \delta 1_N^T \end{bmatrix}, \; \hat{A} = \begin{bmatrix} A \\ \delta 1_K^T \end{bmatrix} \tag{26}$$
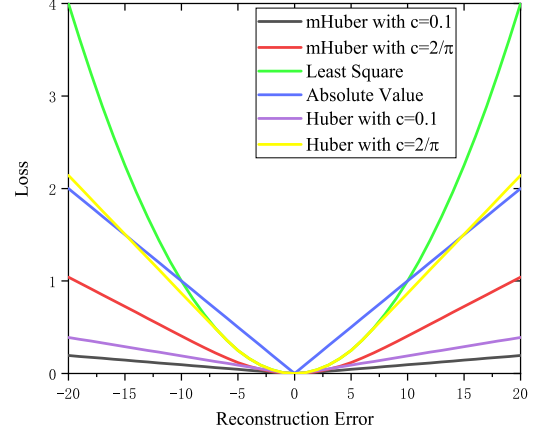

Fig. 1. Comparison of different loss functions.

in which $1_N^T = 1_K^T S$, and δ controls the effect of ASC, most papers choose δ from 10 to 20 to ensure relatively high accuracy and convergence rate. At last, since our algorithm cannot decide the number of endmembers, the number K is given by a prior information by other algorithms, such as VD [30], HySime [31], and R-CoNMF [15].

2) Convergence Analysis: When A, W, c are fixed, there are several requirements to ensure the global convergence of the algorithm 1. c_1 and c_2 are set as $0 < c_1 < 1/2$ and $c_1 < c_2 < 1$ (i.e., $c_1 = 0.0001$, $c_2 = 0.5$). When satisfying the weak Wolfe condition, [32, Th. 3.1] revealed that when the size of β_{k+1} ranges in the interval $r_k \in [\frac{c_2-1}{c_2+1}], \lim_{k\to\infty} inf\|/varXi_k\| = 0$. $\|\cdot\|$ means the Euclidean norm. The global convergence of A is natural when S, W, c are fixed.

3) Computational Complexity Analysis: Now, the computational complexity analysis of our proposed methods will be given in detail. The computational complexity in calculating A can be easily derived the normal NMF method. So we try to get the time of floating-point calculation in calculating S, W, c in each iteration. From [12], the addition and multiplication cost of SL is dKN, respectively, in which d represents the d nonzero entries in each row of G. In our algorithm, t ($t < 8$) is the maximum search number in the inexact line search for searching a_k. The nonlinear conjugate gradient method only has the computational complexity of $\mathcal{O}(\text{MNK})$ for updating S. Like GLNMF, there is only one $\mathcal{O}(MN^2)$ operation in our proposed algorithm to achieve d-nearest neighbor for manifold learning. The calculation times of each iteration are given in the Table II. With γ steps of iteration, the overall cost of our proposed method is $\mathcal{O}(\gamma\text{MNK} + MN^2)$.

4) Robustness Analysis: We propose the robustness analysis of our method. In order to study the property of mHuber, we make a comparison between mHuber and some other M-estimator statistically and geometrically. As shown in Fig. 1, we can see that when the reconstruction error increases, the original Huber and mHuber loss function generally enjoy much lower increasing rate than L_1 and L_2. This indicates the insensitivity of Huber loss. In addition, with the same tuning parameter c of 0.1 and $2/\pi$, the curves of mHuber loss are always under the curves

TABLE II
CALCULATION TIMES IN EACH ITERATION IN MHUBER NMF METHOD

	Update A	Update S	Update W & c	Total
Addition	$(N+1)MK + (M+N)K^2$	$MNK + (M+N)K^2 + (2+3t+dt)NK$	-	$(2N+1)MK + 2(M+N)K^2 + (2+3t+dt)NK$
Multiplication	$(N+1)MK + (M+N)K^2$	$(M+12t+5+dt)NK + (M+N)K^2$	$2MN$	$(12t+dt+5)NK + 2(M+N)K^2 + (2N+2NK+K)M$
Division	MK	$2NK$	$2MN$	$MK + 2NK + 2MN$
Subtraction	$-$	$3NK + t$	MN	$3NK + t + MN$

TABLE III
COMPARISON OF DIFFERENT M-ESTIMATORS

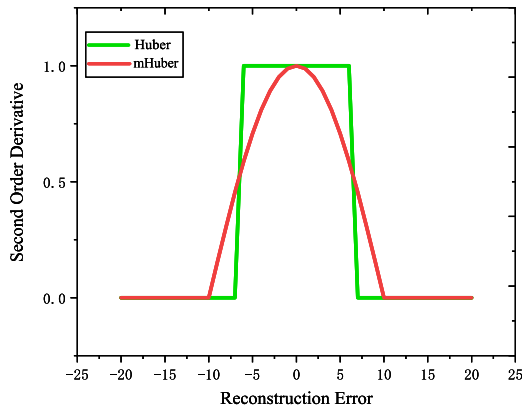| M-estimator	Loss function	Weight	Derivative
L_1	$|x|$	$1/|x|$	$sgn(|x|)$
L_2	$x^2/2$	1	x
$L_{2,1}$	$2\left(\sqrt{1+x^2/2}\right)-1$	$1/\left(\sqrt{1+x^2/2}\right)$	$x/\left(\sqrt{1+x^2/2}\right)$
Cauchy	$\frac{c^2}{2}\log\left(1+(x/c)^2\right)$	$\frac{1}{1+(x/c)^2}$	$\frac{x}{1+(x/c)^2}$
Huber	$\begin{cases} x^2/2, & |x| \le c \\ c|x| - c^2/2, & |x| > c \end{cases}$	$\begin{cases} 1, & |x| \le c \\ c/|x|, & |x| > c \end{cases}$	$\begin{cases} x, & |x| \le c \\ c \cdot x/|x|, & |x| > c \end{cases}$
mHuber	$\begin{cases} c^2[1-\cos(x/c)], |x|/c \le \pi/2, \\ c|x| + c^2(1-\pi/2), else. \end{cases}$	$\begin{cases} c \cdot \sin(x/c)/x \\ c \cdot sgn(x/c)/x \end{cases}$	$\begin{cases} c \cdot \sin(x/c) \\ c \cdot sgn(x/c) \end{cases}$

Fig. 2. Comparison of different second-order derivative

of original Huber. Therefore, the mHuber has better robustness than original Huber loss to reduce the negative effect of large outliers. Third, As can be seen in Fig. 2, mHuber loss enjoys a continuous second-order moment than original Huber loss, which would add more stability in the computation. In Table III, six M-estimators with their corresponding weight matrix and first-order derivative are listed. These estimators are all that can be used in NMF algorithm. Cauchy function has a descending first derivative, which tends to yield erroneous solutions in a way which cannot be observed [33]. L_2 is sensitive to noise, L_1 reduces the influence of both large errors and small errors. $L_{2,1}$ has no tuning parameter to determine the inliers and outliers.

IV. EXPERIMENT

This section will evaluate the performance of our proposed methods. Both synthetic and real-world data are tested. We utilize spectral angle distance (SAD) and root-mean-square error (RMSE) to assess the accuracy of the extracted endmembers and the corresponding abundances. SAD intends to evaluate the similarity of the ground-truth kth endmember signature $\boldsymbol{A}_k$ and the estimation $\hat{\boldsymbol{A}}_k$, it is defined as

$$\text{SAD}_k = \arccos\left(\frac{\boldsymbol{A}_k^T \hat{\boldsymbol{A}}_k}{\|\boldsymbol{A}_k\|\|\hat{\boldsymbol{A}}_k\|}\right). \tag{27}$$

The RMSE evaluates the difference of abundance estimate $\boldsymbol{S}_k$, the ground-truth abundance matrix for kth endmember is $\hat{\boldsymbol{S}}_k$, which is defined as

$$\text{RMSE}_k = \left(1/N\left|\boldsymbol{S}_k - \hat{\boldsymbol{S}}_k\right|^2\right)^{1/2}. \tag{28}$$

Despite of the two common indicators, we introduce spectral feature fitting (SFF) [34],[35] as another indicator to examine the performance of unmixing for mineral dataset: Cuprite dataset. SFF has been widely used in real practice in remote sensing due to its sensitivity in recording precise and subtle mineral absorption features in the spectral diagram of the mineral under consideration. And not all the band ranges are necessarily useful in identifying the substances. Thus we adopt spectra between the band range from 2.0 to 2.5 μm in SFF method for further comparison between the estimated endmembers and reference endmembers. Here, after continuum removal, the estimated $\hat{\boldsymbol{A}}_{mk}$ and ground-truth kth endmember signature $\boldsymbol{A}_{mk}$, m is the mth band. In order to achieve best fitting, we use the minimum of $\sum_m (\hat{\boldsymbol{A}}_{mk} - (a\boldsymbol{A}_{mk} + b))^2$, in which a, b are the least square fitting coefficients. Then the related coefficient f_{cor} of $\boldsymbol{A}_k$ and $\hat{\boldsymbol{A}}_k$ can be calculated by $f_{\text{cor}} = \frac{\text{Cov}(\hat{\boldsymbol{A}}_k, \boldsymbol{A}_k)}{\sqrt{D(\hat{\boldsymbol{A}}_k)}\sqrt{D(\boldsymbol{A}_k)}}$. Then the root mean square R_{sff} is obtained by $R_{\text{sff}} = \sqrt{(1-f_{\text{cor}}^2)D(\boldsymbol{A}_k)}$. Finally the ratio of f_{cor} and R_{sff} are often used to represent the fitting performance, i.e.,

$$\text{SFF} = \frac{f_{\text{cor}}}{R_{\text{sff}}}. \tag{29}$$

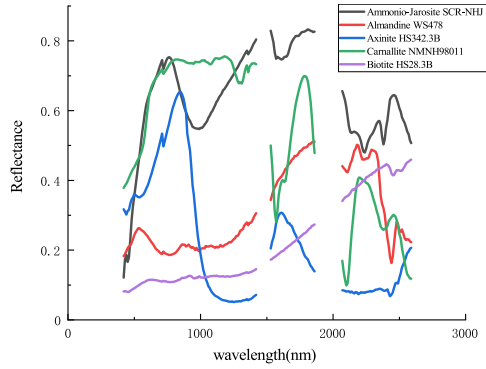Larger SFF indicates higher similarity of two spectral curves.

Fig. 3.   Several endmembers used in synthetic data.

Fig. 4.   Comparisons of convergence rate.

The proposed mHuber method is compared with several widely used M-estimators and HU NMF algorithms throughout this article, including L_2-NMF, $L_{1/2}$-NMF, $L_{2,1}$-NMF, CIM-NMF, Cauchy-NMF, Huber-NMF. All the algorithms are implemented by ourselves. The best results are highlighted in bold in all the tables in this paper.

A. Experiments on Synthetic Data

We refer to the common method to generate the synthetic dataset. As in [12], 10 pure spectra (i.e., endmembers) are arbitrarily chosen form USGS spectral library. Among them, we display the first 5 spectra in Fig. 3. The image is of size 64×64, in which each pixel has 224 spectral bands. Their wavelengths range from 0.38 to 2.5 μm with a resolution of 10 nm. In order to linearly mix them, the entire image would be divided into several 8×8 blocks and then be simulated into an image with mixed pixels by a spatial low-pass filter. The low-pass filter we utilize here is a simple 7×7 averaging filter, which is able to determine the degree of mixing. To enhance the mixture of the obtained dataset, we replace each pixel with the abundance larger than 80% by a new pixel with of all endmembers having equal abundances in it; so we can see that, the abundance of each endmember in these pixels turn out to be $1/c$. The example of endmembers in synthetic dataset are shown in Fig. 3.

Gaussian noise with zero mean is added into the synthetic data having different levels of signal-to-noise ratio (SNR) as

$$\text{SNR} = 10 \log_{10} \frac{E\left(X^T X\right)}{e^T e} \tag{30}$$

in which X represents the original data and e represents the noise in a pixel. E is the expectation.

1) Convergence Rate Analysis: Now, we perform convergence rate experiments of original Huber function and mHuber function under two algorithms, i.e., the existing half-quadratic algorithm and our proposed improved half-quadratic algorithm. All the experiments are set on the condition of SNR = 35 dB, endmember number 4, and VCA and FCLS for initialization. We record the first 500 iterations of each experiment. In general, the MU are often criticized to be very slowly convergent especially with large dataset. The conjugated gradient method searches orthogonal directions at each iteration to ensure that every descent
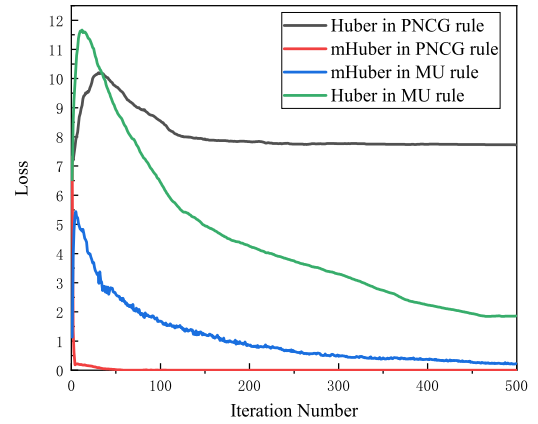
direction is different from others. As shown in Fig. 4, the red line shows that our proposed mHuber model under our PNCG rule in half-quadratic scheme enjoys extremely fast convergence rate in the first 20–50 iterations. mHuber model under MU rule (blue line) does not converge until almost first 300 iterations. Turning to Huber model, it (black line) nearly converges at the 120th iteration under our proposed algorithm, while it (green line) needs more than 500 iterations to converge under MU rule. Both the results show our superiority in convergence rate. Moreover, the fitting effects of our proposed model are much better than the Huber model.

2) Impact of Different Kinds of Noise: In this experiment, we compare SAD and RMSE of different NMF methods on different kinds of noise including salt and pepper noise, Rayleigh noise, gamma noise, Poisson noise. All the algorithms use VCA and FCLS methods for initialization. The Rayleigh noise is elementwise which obeys Rayleigh probability density function. The salt and pepper noise: The amplitude is almost the same, but the location is randomly distributed. The Poisson noise: The mean of its distribution equals to deviation. The gamma noise: It subjects to gamma distribution, which is the composition of several exponential distributions. It is observed from Table IV, our mHuber NMF model enjoys two lowest SAD for gamma noise and Rayleigh noise. For the other two noises, our results keep stable, not fluctuating dramatically. And it is easy to see from Table V, our proposed method enjoys all best RMSE results and low deviations. For Tables IV and V, our mHuber method gets 6 best results out of all 8 results, this indicates our general robustness to different kinds of noise. Note that our proposed model surpasses the original Huber model in all four kinds of noise.

B. Experiments on Real-World Data

Here, we conduct experiments on real-world data. Two datasets are used, the Samson dataset and the Cuprite dataset acquired by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor over Cuprite, Nevada. These two datasets are widely applied to evaluate the performance of hyperspectral unmixing algorithms. Moreover, in order to validate the robustness

TABLE IV
SAD ($\times 10^{-2}$) OF DIFFERENT KINDS OF NOISE FOR SYNTHETIC DATASET

	L_2-NMF		$L_{1/2}$-NMF		$L_{2,1}$-NMF		CIM-NMF		Cauchy NMF		Huber NMF		mHuber NMF	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
Salt & Pepper	16.81	3.22	11.29	2.85	9.71	2.97	11.29	3.17	**8.74**	2.54	11.49	2.83	9.77	1.05
Rayleigh	12.71	1.85	13.44	0.59	9.34	4.04	8.62	2.16	8.82	5.19	11.54	3.65	**8.60**	3.51
Poisson	5.74	3.22	**4.97**	3.58	6.14	3.11	6.76	1.76	7.48	2.52	11.67	3.22	6.73	1.56
Gamma	30.39	6.23	29.55	4.63	28.70	4.21	28.18	0.95	29.01	2.85	29.01	5.17	**27.67**	3.49

TABLE V
RMSE ($\times 10^{-2}$) OF DIFFERENT KINDS OF NOISE FOR SYNTHETIC DATASET

	L_2-NMF		$L_{1/2}$-NMF		$L_{2,1}$-NMF		CIM-NMF		Cauchy NMF		Huber NMF		mHuber NMF	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
Salt & Pepper	16.85	6.22	13.42	3.54	10.53	2.56	10.79	3.47	10.42	1.65	11.01	2.29	**10.02**	0.55
Rayleigh	15.87	3.90	12.76	5.12	12.49	1.95	10.47	6.44	12.80	1.71	8.91	2.55	**8.16**	2.89
Poisson	7.16	2.31	5.79	3.22	6.14	3.11	6.68	4.77	6.92	2.04	6.38	3.46	**5.77**	2.10
Gamma	22.36	5.45	21.46	3.78	20.70	1.12	19.70	3.67	22.93	1.36	20.01	5.33	**19.62**	1.21

TABLE VI
RUNNING TIME (SECOND) OF DIFFERENT NMF METHODS FOR PURE SAMSON DATASET UNDER PNCG AND MU RULE

	$L_{2,1}$	CIM	Cauchy	Huber	mHuber
MU	1.35	2.76	3.77	3.64	3.47
PNCG	1.27	0.21	2.75	3.31	2.54

of our proposed method, we not only conduct experiments on the two datasets but also on the two "noisy" datasets by adding Gaussian noise. Each experiment is initialized by VCA and FCLS method, and each result used here is the mean values after running each experiment five times.

1) *Samson Dataset*: Each pixel in Samson dataset has 156 spectral bands with the wavelengths ranging from 401 to 889 nm. Its spectral resolution is 3.13 nm. There are three substances in the dataset including tree, rock, and water. Here we compare SAD and RMSE of different NMF methods on both pure Samson dataset and noisy Samson dataset (adding Gaussian noise). The noisy Samson dataset is created by adding Gaussian noise in it. The SNR is 15 dB here, which is relatively large. First, Table VI shows the running time of different methods with M-estimators to get their best results (SAD), and compares their running time under PNCG rule and MU rule. It clearly shows the evident decrease of running time in all compared methods by using PNCG rule. As shown from Table VII, mHuber has the lowest mean SAD of 0.0629 among all seven methods. Huber NMF, Cauchy NMF, and CIM-NMF are also better than the common NMF methods in HU. As for noisy Samson dataset, the advantages of our mHuber NMF is apparent, we not only have the lowest mean SAD of 0.1190 but also have all the best SADs of three endmembers. Similarly, L_2-NMF and $L_{1/2}$-NMF behave less satisfying than the models with better robustness, such as $L_{2,1}$-NMF, CIM-NMF, Cauchy NMF, Huber NMF, and

mHuber NMF. From Fig. 5, we show our obtained endmember spectra both on pure Samson and noisy Samson, compared with the ground truth. In (a) of rock, three spectral curves are quite close, with the one in noisy Samson a little farther than other two. In (b) of tree, three spectral curves are almost overlapped. But in (c) for water, the effect of noise has gradually released after reaching the top. The blue curve decreases sharply after reaching the top, and then shows large difference with other two curves.

2) *Cuprite Dataset*: Cuprite dataset was captured over Cuprite in Southern Nevada by the AVIRIS sensor, which has been widely used in examining HU algorithms. 224 bands are included in the Cuprite image from 400 to 2500 nm with size of 250×191. After removing low SNR bands and water vapor absorption bands (i.e., 1-2104-113, 148-167, and 221-224 bands), 188 bands are left. In most papers, the endmember number is set to be 12.

First, Table IX shows the running time of different methods with M-estimators to get their best results (SAD) on pure Cuprite dataset, and compares their running time under PNCG rule and MU rule. The results show a big improvement in running time on all the compared methods. Due to the similar computational complexity, we do not give the results on noisy Cuprite dataset. The noisy Cuprite dataset is also created by adding Gaussian noise in it. The SNR is 15 dB. We can see from Table X, our method maintains its excellent performance in the mean SAD value as 0.1221 and get 5 best individual SADs out of all 12 ones. But we note that, there is not any obvious difference between the results of L_2-NMF, $L_{1/2}$-NMF, and other four methods with better robustness. We guess it may derive from the fine initialization of VCA and FCLS methods or low noise in Cuprite dataset. It needs further exploration. Then from Table XII, we can see that the proposed method also gets the largest SFF value of 4.75 among all the compared methods, and in six of twelve minerals, the proposed method achieves the best fit in all the

TABLE VII
SAD ($\times 10^{-2}$) OF DIFFERENT NMF METHODS FOR PURE SAMSON DATASET

	L_2-NMF	$L_{1/2}$-NMF	$L_{2,1}$-NMF	CIM-NMF	Cauchy NMF	Huber NMF	mHuber NMF
Rock	7.38	4.90	3.68	3.26	3.85	3.72	**2.87**
Tree	4.91	3.89	5.73	**5.46**	6.68	5.31	5.66
Water	13.94	13.49	11.40	11.44	**9.55**	11.21	10.22
Mean	8.75	7.43	6.94	6.72	6.69	6.75	**6.29**

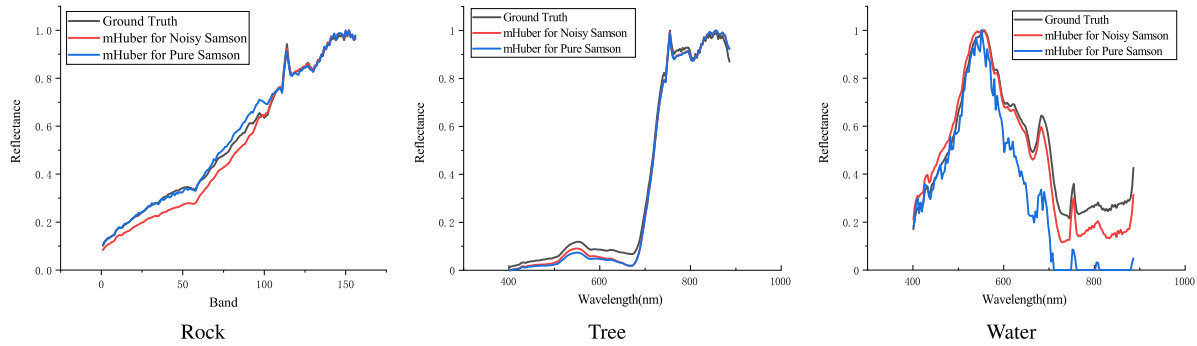
Fig. 5. Endmembers estimated by our method for pure and noisy Samson and ground truth.

TABLE VIII
SAD ($\times 10^{-2}$) OF DIFFERENT NMF METHODS FOR NOISY SAMSON DATASET

	L_2-NMF	$L_{1/2}$-NMF	$L_{2,1}$-NMF	CIM-NMF	Cauchy-NMF	Huber NMF	mHuber NMF
Rock	8.08	6.25	2.80	3.09	2.82	2.85	**2.63**
Tree	6.28	5.00	6.28	6.57	6.49	6.51	**5.01**
Water	35.50	36.38	28.93	29.06	29.76	29.19	**28.26**
Mean	16.62	15.88	12.67	12.91	13.02	12.85	**11.90**

TABLE IX
RUNNING TIME (SECOND) OF DIFFERENT NMF METHODS FOR PURE CUPRITE DATASET UNDER PNCG AND MU RULE

	$L_{2,1}$	CIM	Cauchy	Huber	mHuber
MU	11.22	10.73	53.66	31.49	26.85
PNCG	8.03	4.97	32.56	19.61	17.78

methods. Then for Table XI, the proposed method does not behave as good as before. Cauchy NMF has the best SAD results among all compared methods. But for Table XIII, the proposed method obtained the largest mean SFF of 8.88, slightly bigger than 8.85 of $L_{2,1}$-NMF. The advantage is not as obvious as in the noisy Cuprite dataset.

From Fig. 6, we compared our obtained endmember spectra both on pure Cuprite and noisy Cuprite with the ground truth. The spectral curves on noisy Cuprite (blue line) show high similarity to almost all the ground truth but undergo sharp fluctuation in almost all the endmembers. But for (a) Alunite, the difference between blue line and other four is large, especially during the bands from 2000 to 2500 nm. The purple and green

curves are obtained by evaluating the similarity of two spectral curves by SFF during the range of 2000 and 2500 nm. From the figure of Muscovite, SFF-based method shows absorption features around 2200 nm. As SAD focuses on global shape similarity while SFF highlights the spectral absorption feature in some specific band range in identifying different minerals, then, it would be better to combine these two indicator together to evaluate the performance of unmixing.

In order to evaluate the performance of identification of endmembers, we compare the abundance maps of the 12 minerals matched by SFF and SAD method on pure Cuprite dataset (Table XIV) with the classification map created by by Tetracoder 3.3 software [36]. Since the widely used reference Fig. 7 does not represent all the materials included in the groundtruth, such as Andradite, Pyrope, Sphene, Dumortierite, we just compare some of them. For example, the white region of Fig. 7 represents "Chalcedony". Table XIV visualized our abundance maps by SFF and SAD. SFF method in the revised version used 50 bands from 2.0 to 2.48 μm while SAD method used all bands.

It is easy to see from Table XIV, by SFF method, the identification of Alunite, Chalcedony, Kaolinite 1, and Kaolinite 2

TABLE X
THE PROPOSED ABUNDANCE MAPS IDENTIFIED BY SAD AND SFF FOR PURE CUPRITE DATASET
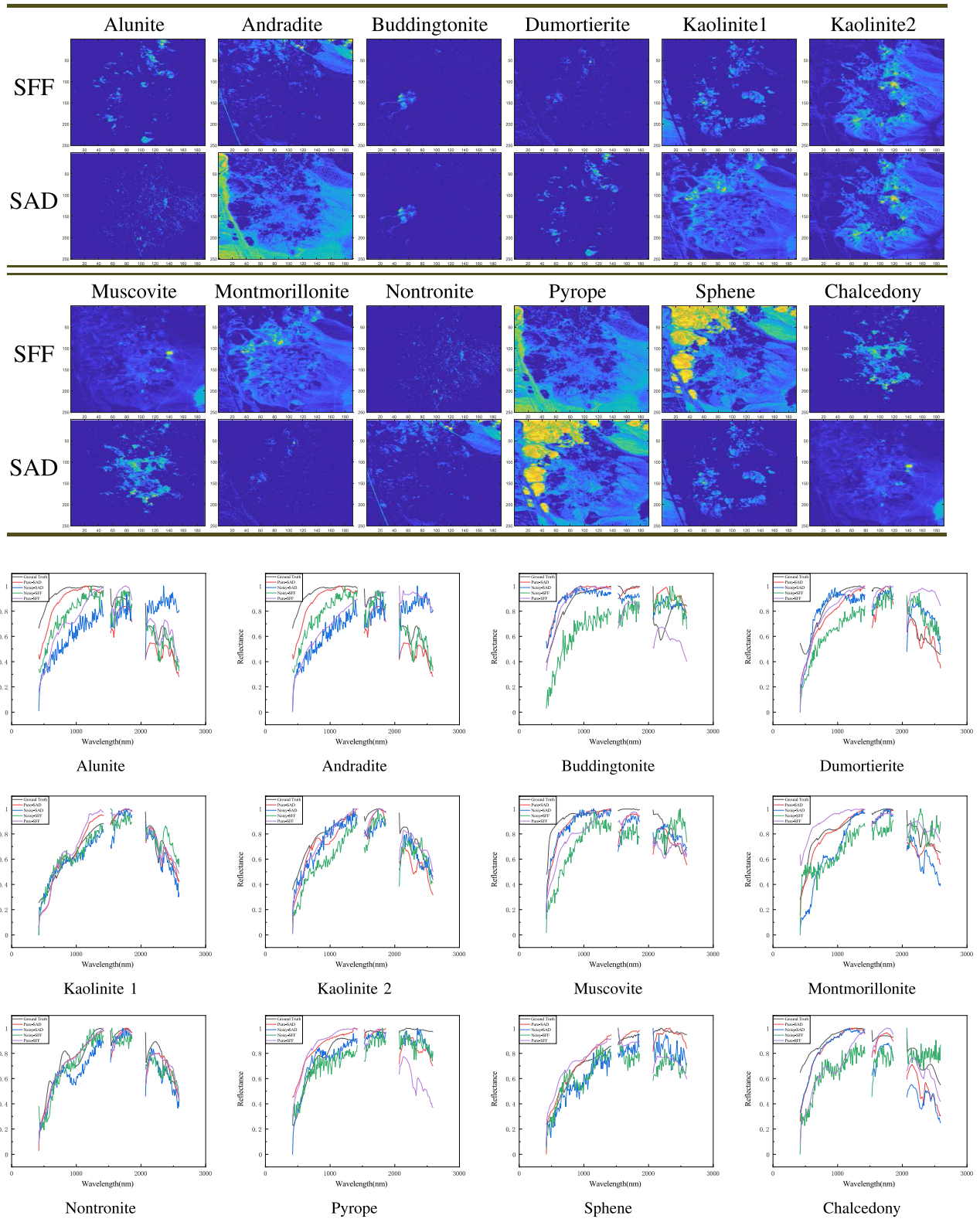
Fig. 6. Endmembers estimated by our method for pure and noisy Cuprite and ground truth. The black line represents the ground truth. The red line represents the spectral curves of our proposed method in pure Cuprite evaluated by SAD, while the blue line indicates the spectral curves of our proposed method in noisy cuprite by SAD. The purple line represents the spectral curves of our proposed method in pure Cuprite evaluated by SFF, while the green line indicates the spectral curves of our proposed method in noisy Cuprite by SFF.

TABLE XI
SAD ($\times 10^{-2}$) ON PURE CUPRITE DATASET OF DIFFERENT METHODS

	L_2-NMF	$L_{1/2}$-NMF	$L_{2,1}$-NMF	CIM-NMF	Cauchy NMF	Huber NMF	mHuber NMF
Alunite	**11.53**	23.44	11.87	**11.53**	11.63	11.75	22.08
Andradite	7.51	7.56	7.38	7.51	**7.25**	7.35	7.59
Buddingtonite	12.15	11.93	12.17	12.15	**9.26**	9.31	11.19
Dumortierite	11.29	16.73	**11.27**	11.29	11.28	11.28	11.91
Kaolinite 1	9.61	9.01	8.53	8.61	**8.52**	8.60	8.73
Kaolinite 2	11.11	**6.17**	11.07	11.11	11.07	11.19	10.15
Muscovite	10.92	15.03	10.97	10.91	11.08	11.05	**10.86**
Montmorillonite	7.99	11.90	7.92	7.99	**6.80**	6.92	7.58
Nontronite	8.45	9.55	**8.36**	8.45	8.50	8.47	8.63
Pyrope	14.42	10.95	14.59	14.42	17.14	17.04	**5.49**
Sphene	6.05	**5.61**	6.31	6.05	6.55	6.35	10.34
Chalcedony	14.04	**11.65**	14.15	14.04	14.05	14.21	12.29
Mean	10.34	11.63	10.39	10.34	**10.26**	10.29	10.57

TABLE XII
SFF ON NOISY CUPRITE DATASET OF DIFFERENT METHODS

	L_2-NMF	$L_{1/2}$-NMF	$L_{2,1}$-NMF	CIM-NMF	Cauchy NMF	Huber NMF	mHuber NMF
Alunite	6.40	6.37	6.32	**6.6**	6.36	6.41	6.46
Andradite	4.67	4.69	4.64	4.79	4.79	4.90	**5.77**
Buddingtonite	6.19	6.12	5.83	6.31	5.05	5.68	**6.33**
Dumortierite	2.19	0.47	1.97	0.39	2.37	0.53	**2.66**
Kaolinite 1	6.99	6.55	6.21	**7.09**	5.74	5.86	5.73
Kaolinite 2	**1.93**	1.65	0.40	1.56	0.54	1.22	0.52
Muscovite	12.16	12.06	11.78	**12.27**	11.74	12.04	10.83
Montmorillonite	6.28	6.71	5.71	5.91	6.38	6.86	**7.77**
Nontronite	2.37	2.34	2.21	**2.80**	2.09	2.23	2.27
Pyrope	3.02	3.10	2.72	2.96	3.04	3.11	**3.51**
Sphene	0.40	2.40	1.60	1.83	1.78	**2.45**	1.86
Chalcedony	3.04	3.10	2.74	2.83	3.11	3.15	**3.30**
Mean	4.64	4.63	4.34	4.61	4.42	4.54	**4.75**

TABLE XIII
SFF ON PURE CUPRITE DATASET OF DIFFERENT METHODS

	L_2-NMF	$L_{1/2}$-NMF	$L_{2,1}$-NMF	CIM-NMF	Cauchy NMF	Huber NMF	mHuber NMF
Alunite	2.77	3.01	3.06	2.92	3.17	3.08	**3.31**
Andradite	8.13	10.18	**11.68**	10.75	10.37	11.01	9.38
Buddingtonite	7.63	7.65	7.54	**7.87**	7.27	7.38	7.14
Dumortierite	**19.31**	17.62	13.76	17.45	17.21	17.19	17.51
Kaolinite 1	7.29	7.43	7.54	7.53	7.54	7.49	**7.67**
Kaolinite 2	12.38	13.50	13.76	13.37	13.95	13.85	**14.01**
Muscovite	0.37	0.57	0.60	0.28	0.65	0.56	**0.77**
Montmorillonite	0.58	0.749	1.01	1.07	**1.09**	0.121	0.91
Nontronite	2.01	1.90	**1.98**	2.09	1.92	1.97	1.75
Pyrope	3.74	3.82	3.79	3.83	3.81	3.79	**3.86**
Sphene	27.40	28.26	28.95	27.91	28.34	27.97	**29.88**
Chalcedony	9.00	9.57	9.42	9.09	9.72	9.49	**10.41**
Mean SFF	8.38	8.69	8.85	8.62	8.75	8.73	**8.88**

TABLE XIV
SAD ($\times 10^{-2}$) ON NOISY CUPRITE DATASET OF DIFFERENT METHODS

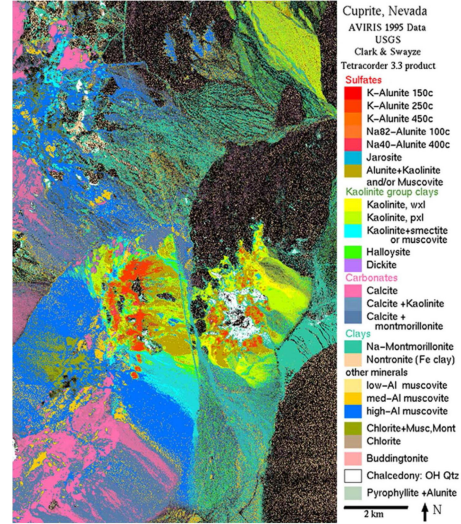|	L_2-NMF	$L_{1/2}$-NMF	$L_{2,1}$-NMF	CIM-NMF	Cauchy NMF	Huber NMF	mHuber NMF
Alunite	25.20	31.36	24.69	25.68	24.90	39.37	**21.12**
Andradite	11.44	14.77	**11.11**	11.82	11.25	16.79	14.91
Buddingtonite	11.05	**10.61**	10.88	11.47	10.95	11.58	11.60
Dumortierite	16.20	17.47	15.89	16.37	16.00	16.49	**15.83**
Kaolinite 1	9.62	10.91	9.56	10.86	9.56	8.95	**7.58**
Kaolinite 2	9.88	10.56	**9.36**	10.32	9.55	10.06	11.44
Muscovite	17.08	18.55	16.34	17.98	16.63	**11.17**	13.50
Montmorillonite	10.45	11.56	9.72	11.20	10.04	19.01	**7.55**
Nontronite	12.49	11.33	12.28	10.81	12.38	**9.51**	13.35
Pyrope	9.83	10.10	10.34	9.45	10.14	**7.69**	11.49
Sphene	11.15	9.95	12.31	**9.87**	11.87	8.88	11.53
Chalcedony	14.90	16.01	14.07	15.69	14.36	17.51	**6.55**
Mean	13.27	14.43	13.04	13.46	13.14	14.75	**12.21**

Fig. 7. Classification map of Cuprite dataset obtained by Tetracoder 3.3 software [36].

When we consider the reason of the misidentification, two points might be relevant. First of all, even though spectral absorption feature is essential in identifying minerals, the mixing of minerals may still lead to wavelength shift, since different minerals can have different absorption feature ranges. For example, in the "Kaolinite and Alunite and/or Muscovite," Muscovite's range is 2.082 to 2.395 μm; Alunite's range is 1.99 to 2.36 μm; Kaolinite's range is 2.20 to 2.22 μm. This phenomenon might bring obstacles in extracting endmembers for unmixing work.

Second, since the proposed algorithm aims to get the smallest mean SFF or SAD, it cannot guarantee the best matching of each minerals. And as 12 estimated endmembers have to be matched to 12 reference spectra, one wrong misidentification would cause chain reaction. Thus, there remains more research to do to promote the performance. Future work will pay more attention on this problem.

V. CONCLUSION

In this article, a mHuber-based NMF algorithm is proposed for HU task. This method explores a more stable huber loss for NMF, which also induces better reconstruction. Compared with original Huber NMF, the mHuber NMF algorithm enjoys less sensitivity and better stability. The improved half-quadratic algorithm we propose not only accelerates the convergence rate but also gives a new tuning parameter adapting to the mHuber NMF model. Experimental results on synthetic datasets confirms our advantages on convergence rate, SAD and RMSE estimates, robustness to different kinds of noise. SFF indicators on Cuprite dataset also confirm our superiority over other methods in HU performance and better identification of endmembers than SAD indicators. Experiments on real-world datasets also implies the efficiency of our proposed method especially on the noisy datasets.

were considered to be right. Montmorillonite and Muscovite were misidentified. However, SAD method only identified Alunite and Kaolinite 2 correctly, while misidentifying at least Chalcedony, Muscovite, Dumortierite, Kaolinite 1. Therefore, SFF method shows much better identification accuracy than SAD method. But without more abundance groundtruth information, there are still several identifications that cannot be evaluated.

References

[1] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[3] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data," in *Proc. SPIE*, vol. 3753, pp. 266–277, Oct. 1999.

[4] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.

[5] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process. Evol. Remote Sens.*, 2009, pp. 1–4.

[6] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, vol. 3, 2008, pp. 250–253.

[7] C. Y. Zheng, H. Li, Q. Wang, and C. P. Chen, "Reweighted sparse regression for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 479–488, Jan. 2016.

[8] X. Fu, W. K. Ma, T. H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1128–1141, Sep. 2015.

[9] S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 482–496, Jan. 2018.

[10] X. Wei, X. Liu, B. Wang, and L. Zhang, "Independent component analysis for blind unmixing of hyperspectral imagery with additional constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2165–2179, Jun. 2011.

[11] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4282–4297, Nov. 2011.

[12] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[13] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[14] Z. Guo, T. Wittman, and S. Osher, "L1 unmixing and its application to hyperspectral image enhancement," in *Proc. SPIE*, vol. 7334, 2009, Art. no. 73341M.

[15] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016.

[16] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017.

[17] H. C. Li, S. Liu, X. R. Feng, and S. Zhang, "Sparsity-constrained coupled nonnegative matrix-tensor factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5061–5073, 2020.

[18] Z. Shu, J. Zhou, T. Lei, B. Xiao, and C. Zhao, "Multilayer manifold and sparsity constrainted nonnegative matrix factorization for hyperspectral unmixing," in *Proc. IEEE Conf. Image Process.*, 2015, pp. 2174–2178.

[19] J. Peng, W. Sun, F. Jiang, H. Chen, Y. Zhou, and Q. Du, "A general loss-based nonnegative matrix factorization for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, to be publihed, doi: 10.1109/LGRS.2020.3017233.

[20] L. Du, X. Li, and Y. D. Shen, "Robust nonnegative matrix factorization via half-quadratic minimization," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2012, pp. 201–210.

[21] H. Wang, W. Yang, and N. Guan, "Cauchy sparse NMF with manifold regularization: A robust method for hyperspectral unmixing," *Knowl.-Based Syst.*, vol. 184, Aug. 2019, Art. no. 104898.

[22] J. Peng, Y. Zhou, W. Sun, Q. Du, and L. Xia, "Self-paced nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1501–1515, Feb. 2021.

[23] P. J. Huber, *Robust Statistics*. vol. 523, Hoboken, NJ, USA: Wiley, 2004.

[24] W. J. J. Rey, *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin, Germany: Springer-Verlag, 1983.

[25] M. Nikolova and R. H. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process*, vol. 16, no. 6, pp. 1623–1627, Jun. 2007.

[26] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM J. Optim.*, vol. 10, no. 1, pp. 177–182, Nov. 1999.

[27] M. R. Hestenes *et al.*, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Standard*, vol. 49, no. 6, pp. 409–436, 1952.

[28] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Commun. Statist.-Theory Methods*, vol. 6, no. 9, pp. 813–827, Jun. 2007.

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

[30] C. I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[31] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[32] Y. H. Dai and Y. Yuan, "An efficient hybrid conjugate gradient method for unconstrained optimization," *Ann. Oper. Res.*, vol. 103, no. 1/4, pp. 33–47, Jan. 2001.

[33] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image Vision Comput.*, vol. 1, no. 5, pp. 59–76, 1997.

[34] Z. Pan, J. Huang, and F. Wang, "Multi range spectral feature fitting for hyperspectral imagery in extracting oilseed rape planting area," *Int. J. Appl. Earth Observation Geoinformation*, vol. 25, pp. 21–29, 2013.

[35] R. D. M. Scafutto, H. van der Werff, W. H. Bakker, F. van der Meer, and C. R. de Souza Filho, "An evaluation of airborne SWIR imaging spectrometers for CH4 mapping: Implications of band positioning, spectral sampling and noise," *Int. J. Appl. Earth Observation Geoinformation*, vol. 94, 2021, Art. no. 102233.

[36] X. Xu, Z. Shi, and B. Pan, "$\ell 0$-based sparse hyperspectral unmixing using spectral information and a multi-objectives formulation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 141, pp. 46–58, 2018.

Ziyang Guo received the B.Sc. and M.Sc. degrees in applied mathematics from Huazhong University of Science and Technology, Wuhan, China, in 2011 and 2014, respectively, where he is currently working toward the Ph.D. degree in computational mathematics with the School of Mathematics and Statistics.

His research interests include pattern recognition and machine learning.

Anyou Min received the B.S. degree in computational mathematics from Huazhong University of Science and Technology, Wuhan, China, in 2017, where he is currently working toward the Ph.D. degree in statistics with the School of Mathematics and Statistics.

His research interests include pattern recognition and machine learning.

Bing Yang received the B.Sc. degree in information and computing science from Henan University, Kaifeng, China, in 2015. He is currently working toward the Ph.D. degree in statistics with the School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China.

His research interests include pattern recognition and machine learning.

Hong Li (Member, IEEE) received the M.Sc. degree in mathematics and the Ph.D. degree in pattern recognition and intelligence control from Huazhong University of Science and Technology, Wuhan, China, in 1986 and 1999, respectively.

She is currently a Professor with the School of Mathematics and Statistics, Huazhong University of Science and Technology. Her research interests include approximation theory, wavelet analysis, learning theory, neural networks, signal processing, and pattern recognition.

Junhong Chen received the B.S. degree in computational mathematics from Huazhong University of Science and Technology, Wuhan, China, in 2017. He is currently working toward the Ph.D. degree in computational mathematics with the School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China.

His research interests include pattern recognition and deep learning.