A General Loss-Based Nonnegative Matrix Factorization for Hyperspectral Unmixing

Jiangtao Peng, *Member, IEEE*, Weiwei Sun, *Member, IEEE*, Fan Jiang, Hong Chen, Yicong Zhou, *Senior Member, IEEE*, and Qian Du, *Fellow, IEEE*

***Abstract*— Nonnegative matrix factorization (NMF) is a widely used hyperspectral unmixing model which decomposes a known hyperspectral data matrix into two unknown matrices, i.e., endmember matrix and abundance matrix. Due to the use of least-squares loss, the NMF model is usually sensitive to noise or outliers. To improve its robustness, we introduce a general robust loss function to replace the traditional least-squares loss and propose a general loss-based NMF (GLNMF) model for hyperspectral unmixing in this letter. The general loss function is a superset of many common robust loss functions and is suitable for handling different types of noise. Experimental results on simulated and real hyperspectral data sets demonstrate that our GLNMF model is more accurate and robust than existing NMF methods.**

***Index Terms*— General loss, hyperspectral unmixing, nonnegative matrix factorization (NMF).**

I. Introduction

HYPERSPECTRAL image is obtained by simultaneously imaging a targeted region with hundreds of continuous and narrow spectral bands through hyperspectral sensor. Due to the limited spatial resolution of hyperspectral sensors and the mixing of ground materials, the observed spectrum for a pixel may be mixed by several materials' spectra. To deal with the mixed spectra problem, hyperspectral unmixing decomposes a measured mixture spectrum into a collection of endmembers and abundances. One of widely used hyperspectral unmixing techniques is nonnegative matrix factorization (NMF) model [1]–[8], which can decompose a hyperspectral data matrix into the product of two nonnegative matrices, i.e., endmember matrix and abundance matrix.

Although the part-based representation of NMF makes the decomposition matrices more intuitive and interpretable, the solution space of NMF is very large. To reduce the solution space, some additional constraints are imposed either on the endmember matrix or on the abundance matrix, such as minimum volume constraint on endmembers [6], $\ell_{1/2}$ sparsity constraint on abundances [2], and sparsity-constrained deep NMF with total variation [7]. Compared with the original NMF, these modified NMF methods produce much better results. However, when there exists noise in the hyperspectral data (e.g., Gaussian noise and stripes), the performance of these models will dramatically degrade because the least squares objective function in these NMF methods is sensitive to noise [9], [10]. To reduce the effect of noise, robust estimators are introduced to replace the traditional least squares metric and many robust NMF methods have been proposed, such as correntropy loss-based robust NMF (CENMF) [5], $\ell_{2,1}$-norm and $\ell_{1,2}$-norm-based NMF models [11].

In this letter, we introduce a robust general loss function to replace the least squares loss and propose a general loss-based NMF (GLNMF) model for hyperspectral unmixing. The general loss function is a superset of many common robust loss functions, such as Cauchy, Welsch, Geman-McClure, generalized Charbonnier and Huber losses [12]. The robustness and flexibility of the general loss function makes it suitable to handle different types of noise, such as stripe, deadline, and impulse noise. The general loss is mainly controlled by a shape parameter. When the shape parameter becomes more negative, the influence function of the general loss can decrease and tend to zero. Therefore, it can alleviate the influence of a single element, especially for the element with a large noise. Considering that the objective function of GLNMF is nonconvex and nonlinear, we transfer the GLNMF model to an iteratively reweighted NMF problem and design a weight function to suppress the effect of noisy bands. Experimental results demonstrate that our proposed GLNMF model is more accurate and robust than existing NMF methods, especially for hyperspectral data with noisy bands.

II. NMF Unmixing Model

The linear mixture model assumes that each measured pixel can be considered as a linear mixture of several spectral

Manuscript received May 16, 2020; revised July 11, 2020; accepted August 12, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61871177, Grant 41971296, Grant 11771130, Grant 41671342, and Grant 11671161; in part by the Science and Technology Development Fund, Macau SAR under Grant 189/2017/A3; in part by the University of Macau under Grant MYRG2018-00136-FST; in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR19D010001; and in part by the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 18R05. *(Corresponding author: Weiwei Sun.)*

Jiangtao Peng and Fan Jiang are with the Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China (e-mail: pengjt1982@hubu.edu.cn).

Weiwei Sun is with the Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China (e-mail: sunweiwei@nbu.edu.cn).

Hong Chen is with the College of Science, Huazhong Agricultural University, Wuhan 430070, China (e-mail: chenh@mail.hzau.edu.cn).

Yicong Zhou is with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yicongzhou@um.edu.mo).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

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

Digital Object Identifier 10.1109/LGRS.2020.3017233

signatures called endmembers [2], [7], [13], that is

$$\mathbf{x} = \mathbf{Z}\mathbf{h} + \mathbf{e} \tag{1}$$

where $\mathbf{x} \in \mathcal{R}^{M\times 1}$ is an observed spectral pixel, $\mathbf{Z} \in \mathcal{R}^{M\times P}$ denotes the endmember matrix, $\mathbf{h} \in \mathcal{R}^{P\times 1}$ is the corresponding abundance fraction vector, and $\mathbf{e}$ is the noise. Considering all N pixels in the hyperspectral image, the model (1) can be generalized as

$$\mathbf{X} = \mathbf{Z}\mathbf{H} + \mathbf{E} \tag{2}$$

where the matrices $\mathbf{X} \in \mathcal{R}^{M\times N}$, $\mathbf{H} \in \mathcal{R}^{P\times N}$, $\mathbf{E} \in \mathcal{R}^{M\times N}$ represent the hyperspectral data matrix, abundance matrix, and noise matrix, respectively.

It is obvious that the endmember matrix $\mathbf{Z}$ and abundance matrix $\mathbf{H}$ are nonnegative and only the hyperspectral data matrix $\mathbf{X}$ is determinate. On this account, the NMF model can be used for hyperspectral unmixing [1], [2], [6]. The NMF unmixing model is formulated as

$$\min_{\mathbf{Z},\mathbf{H}} \|\mathbf{X} - \mathbf{Z}\mathbf{H}\|_F^2, \quad \text{s.t.,} \ \ \mathbf{Z} \succeq 0, \mathbf{H} \succeq 0 \tag{3}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. In addition, abundance matrix H should satisfy the sum-to-one constraint: $\forall p, n$, $\sum_{p=1}^{P} \mathbf{H}_{pn} = 1$.

The NMF unmixing model can be solved by the multiplicative update algorithm [1]. However, the solution is not unique. To limit the feasible solution set, various constraints have been incorporated into the NMF framework [2], [5]–[7]. A representative constraint is the sparsity constraint, such as ℓ_q-norm ($q = 1/2, 1, 2$) on the abundance matrix [2]. The $\ell_{1/2}$-norm-based NMF model is

$$\min_{\mathbf{Z},\mathbf{H}} \|\mathbf{X} - \mathbf{Z}\mathbf{H}\|_F^2 + \lambda \|\mathbf{H}\|_{1/2} \tag{4}$$

where λ is a regularization parameter and $\|\mathbf{H}\|_{1/2}$ is the $\ell_{1/2}$-regularizer. In the following, we focus on the $\ell_{1/2}$-NMF model due to the good performance of $\ell_{1/2}$-regularizer [2].

III. GLNMF Unmixing Model

Considering that the least squares objective function in the NMF model (4) is sensitive to noise, we introduce a general robust loss function to suppress the large noise in the real data and propose a robust GLNMF model for hyperspectral unmixing.

A. General Robust Loss Function

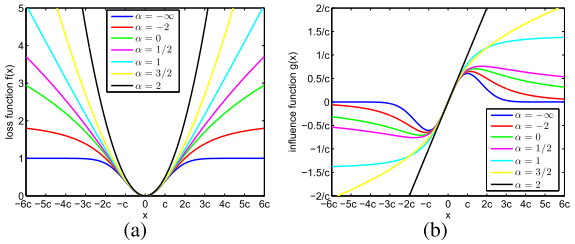
Derived from the "generalized Charbonnier" loss, a general loss function is defined as [12]

$$f(x, \alpha, c) = \frac{|2-\alpha|}{\alpha}\left(\left(\frac{(x/c)^2}{|2-\alpha|} + 1\right)^{(\alpha/2)} - 1\right) \tag{5}$$

where α is a shape parameter that controls the robustness of the loss and c is a scale parameter.

When $\alpha = 2$, the loss approaches the ℓ_2 loss

$$\lim_{\alpha\to 2} f(x, \alpha, c) = \frac{1}{2}(x/c)^2. \tag{6}$$



(a) (b)

Fig. 1. General loss function (a) and its influence function (b) for different values of shape parameter α.

When α approaches zero, we can obtain the Cauchy loss

$$\lim_{\alpha\to 0} f(x, \alpha, c) = \log\left(\frac{1}{2}(x/c)^2 + 1\right). \tag{7}$$

The loss (5) reproduces the Welsch loss in the limit as α approaches negative infinity

$$\lim_{\alpha\to -\infty} f(x, \alpha, c) = 1 - \exp\left(-\frac{1}{2}(x/c)^2\right). \tag{8}$$

By continuously changing the value of shape parameter α, the loss (5) is a generalization of the Cauchy, Geman-McClure, Welsch, Charbonnier, generalized Charbonnier, pseudo-Huber, ℓ_1, and ℓ_2 loss functions. To define the loss function everywhere, the final loss function can be rewritten as

$$f(x, \alpha, c) = \begin{cases} \frac{1}{2}(x/c)^2, & \text{if } \alpha = 2 \\ \log\left(\frac{1}{2}(x/c)^2 + 1\right), & \text{if } \alpha = 0 \\ 1 - \exp\left(-\frac{1}{2}(x/c)^2\right), & \text{if } \alpha = -\infty \\ \frac{|2-\alpha|}{\alpha}\left(\left(\frac{(x/c)^2}{|2-\alpha|} + 1\right)^{\alpha/2} - 1\right), & \text{otherwise.} \end{cases} \tag{9}$$

The influence function of a loss f is defined as [12], [14]

$$g(x) = \frac{\partial f(x)}{\partial x}. \tag{10}$$

For a robust estimator, its influence function should not be sensitive to the increase of the error [14].

The general loss function and its influence function for different values of shape parameter α are shown in Fig. 1. For all values of α, the influence function is approximately linear when $|x| < c$. This means that the effect of a small residual is always linearly proportional to that residual's magnitude [12]. When $\alpha = 2$ (i.e., ℓ_2 loss), the influence function is also a linear function (unbounded) and larger residuals have correspondingly larger effects. So, ℓ_2 loss is not a robust loss and is sensitive to noise. When $\alpha < 1$, the influence function begins to decrease as $|x|$ grows larger than c. As the decrease of α, the effect of outlier is diminished, so the corresponding losses (i.e., ℓ_1 loss with $\alpha = 1$, Cauchy loss with $\alpha = 0$, Geman-McClure loss with $\alpha = -2$) are robust.

Compared with a single robust loss, the general loss is more flexibility. It has a shape parameter α to control the robustness of the loss and a scale parameter c to control the size of the loss's nonrobust quadratic bowl near $x = 0$ [12].

This flexibility in shaping a loss function is useful because of non-Gaussian noise. For complex noise in hyperspectral data, the combination of α and c can ensure the contribution of inliers and suppress the negative effect of outliers.

B. General Loss Based NMF for Unmixing (GLNMF)

The least squares loss function in the $\ell_{1/2}$-NMF model (4) can be rewritten as

$$J(\mathbf{Z},\mathbf{H}) = \|\mathbf{X}-\mathbf{Z}\mathbf{H}\|_F^2 = \sum_{i=1}^{M} \left\|\mathbf{X}^i - (\mathbf{Z}\mathbf{H})^i\right\|_2^2 = \sum_{i=1}^{M} {e_i}^2$$

where $\mathbf{X}^i$ denotes the ith row of the matrix $\mathbf{X}$, and $e_i = \|\mathbf{X}^i - (\mathbf{Z}\mathbf{H})^i\|_2$ measures the row-wise band error. It is clear that $J(\mathbf{Z},\mathbf{H})$ is sensitive to outliers in the hyperspectral data $\mathbf{X}$. To reduce the effect of noisy bands or outliers, we introduce a general robust loss function to replace the least squares loss and propose a GLNMF model for hyperspectral unmixing as

$$\min_{\mathbf{Z},\mathbf{H}} \sum_{i=1}^{M} f\left(e_i^2\right) + \lambda\|\mathbf{H}\|_{1/2} \tag{11}$$

where f is the general robust loss function defined in (9).

Because f is a nonconvex and nonlinear function, it is difficult to solve the model (11) directly. However, we can transfer the optimization problem to an iteratively reweighted NMF problem

$$\min_{\mathbf{Z},\mathbf{H}} \sum_{i=1}^{M} w_i e_i^2 + \lambda\|\mathbf{H}\|_{1/2} \tag{12}$$

where the weight has the form [12]

$$w_i = \frac{1}{e_i}\frac{\partial f}{\partial e_i} = \begin{cases} \dfrac{1}{c^2}, & \text{if } \alpha = 2 \\ \dfrac{2}{e_i^2 + 2c^2}, & \text{if } \alpha = 0 \\ \dfrac{1}{c^2}\exp\left(-\dfrac{1}{2}(e_i/c)^2)\right), & \text{if } \alpha = -\infty \\ \dfrac{1}{c^2}\left(\dfrac{(e_i/c)^2}{|2-\alpha|}+1\right)^{(\alpha/2)-1}, & \text{otherwise.} \end{cases} \tag{13}$$

Substituting the weight and error terms into the model (12), the objective function of GLNMF can be expressed as

$$\begin{aligned} &\sum_{i=1}^{M} w_i \left\|\mathbf{X}^i - (\mathbf{Z}\mathbf{H})^i\right\|_2^2 + \lambda\|\mathbf{H}\|_{1/2} \\ &\quad= \sum_{i=1}^{M} \left\|\sqrt{w_i}\mathbf{X}^i - \left(\sqrt{w_i}\mathbf{Z}\mathbf{H}\right)^i\right\|_2^2 + \lambda\|\mathbf{H}\|_{1/2} \\ &\quad= \left\|\widetilde{\mathbf{X}} - \widetilde{\mathbf{Z}}\mathbf{H}\right\|_F^2 + \lambda\|\mathbf{H}\|_{1/2} \end{aligned} \tag{14}$$

where $\widetilde{\mathbf{X}} = \mathbf{W}^{1/2}\mathbf{X}$ and $\widetilde{\mathbf{Z}} = \mathbf{W}^{1/2}\mathbf{Z}$, and $\mathbf{W}$ is a diagonal matrix whose diagonal element is $\mathbf{W}_{ii} = w_i$.

It is obvious that the GLNMF model is a variant of the original $\ell_{1/2}$-NMF model, and the multiplicative update rule for the $\ell_{1/2}$-NMF can also be applied here. The whole process of GLNMF model is summarized in Algorithm 1.

Algorithm 1 GLNMF for Hyperspectral Unmixing

Input: hyperspectral data matrix $\mathbf{X}$,
initial endmember $\mathbf{Z}_0$ and abundance $\mathbf{H}_0$,
the shape parameter α and scale parameter c.

Output: endmember and abundance matrices.

1. Initialize $\mathbf{Z}^{(0)} = \mathbf{Z}_0$, $\mathbf{H}^{(0)} = \mathbf{H}_0$, set $k = 1$.
2. Run the following steps until convergence:
   (a) calculate the errors:
   $$({e_i}^2)^{(k)} = \|\mathbf{X}^i - (\mathbf{Z}^{(k-1)}\mathbf{H}^{(k-1)})^i\|_2^2$$
   (b) compute the weight of each entry:
   $$w_i = w(({e_i}^2)^{(k)}, \alpha, c)$$
   (c) update weight: $\mathbf{W}^{(k)} = \mathrm{diag}\{w_1^{(k)}, \ldots, w_M^{(k)}\}$
   (d) update matrix:
   $$\widetilde{\mathbf{X}} = (\mathbf{W}^{(k)})^{\frac{1}{2}}\mathbf{X}$$
   $$\widetilde{\mathbf{Z}}^{(k-1)} = (\mathbf{W}^{(k)})^{\frac{1}{2}}\mathbf{Z}^{(k-1)}$$
   (e) solve endmember matrix and abundance matrix:
   $$(\widetilde{\mathbf{Z}}^{(k)}, \mathbf{H}^{(k)}) = \mathrm{L}_{1/2}\mathrm{NMF}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Z}}^{(k-1)}, \mathbf{H}^{(k-1)})$$
   $$\mathbf{Z}^{(k)} = (\mathbf{W}^{(k)})^{-\frac{1}{2}}\widetilde{\mathbf{Z}}^{(k)}$$
   (g) $k = k + 1$

IV. Experiments

In this section, we use synthetic and real-world hyperspectral data sets to evaluate our GLNMF model and compare it to other methods, such as standard NMF, $\ell_{1/2}$-NMF [2], $\ell_{1,2}$-NMF [8], correntropy-induced metric-based NMF (CIMNMF), Huber-NMF, and CENMF [5]. The spectral angle distance (SAD) and abundance maps are used to evaluate the accuracy of the extracted endmembers and corresponding abundances. The vertex component analysis (VCA) and fully constrained least squares (FCLS) methods are used to initialize the endmember matrix and abundance matrix, respectively.

A. Experiments on Synthetic Data

To construct the synthetic data, we first select seven spectral signatures (i.e., "Carnallite NMNH98011," "Actinolite NMNHR16485," "Andradite WS487," "Diaspore HS416.3B," "Erionite + Merlinoit GDS144," "Halloysite NMNH106236," "Hypersthene NMNHC2368") from the United States Geological Survey (USGS) spectral library to construct the endmember matrix. Then, these seven spectra are mixed according to the method in [7] to form the corresponding abundances, and guarantee the nonnegative and sum-to-one constraints. The synthetic data $\mathbf{X}$ is generated by the product of the endmember matrix and the abundance matrix, which has 224 bands and 4096 pixels. To investigate the robustness of the proposed method, we add different types of noise to the synthetic data.

1) IID Gaussian Noise: The synthetic data $\mathbf{X}$ is degraded by the independent identical distribution (iid) Gaussian noise with distribution SNR $\sim \mathcal{N}(\overline{\mathrm{SNR}}, \epsilon^2)$, where $\overline{\mathrm{SNR}} \in \{5, 10, 15, 20\}$ and $\epsilon = 5$. The mean of SAD results over 20 random tests are shown in Table I, where the shape

TABLE I
SAD RESULTS IN THE CASE OF GAUSSIAN NOISE

$\overline{\text{SNR}}$	NMF	$\ell_{1/2}$-NMF	$\ell_{1,2}$-NMF	CENMF	CIM-NMF	Hub-NMF	GLNMF
5	0.1984	0.1330	0.1796	0.1890	0.2145	0.1941	**0.0768**
10	0.0742	0.0604	0.0738	0.0739	0.0804	0.0756	**0.0335**
15	0.0357	0.0350	0.0353	0.0360	0.0396	0.0367	**0.0215**
20	0.0221	0.0219	0.0220	0.0224	0.0248	0.0226	**0.0192**

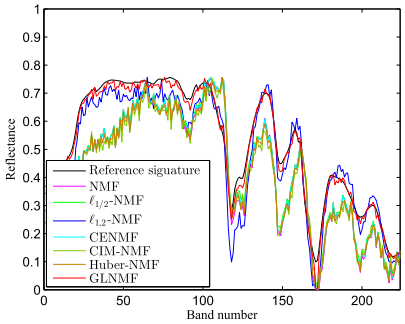


Fig. 2. Reference and estimated spectra for endmember 1.

and scale parameters in the GLNMF are set as: $\alpha = -1$ and $c = 1$.

As can be seen from these tables, the unmixing performance of each algorithm improves significantly as the SNR increases. In the case of a larger SNR (i.e., 20), each algorithm has good unmixing performance, and our model also performs well. When the noise level is high (i.e., lower SNR), other algorithms show marginal improvement over the standard NMF, while our GLNMF model still shows excellent unmixing performance. In general, GLNMF provides the best results in different levels of noise.

To visualize the results, we present the reference and estimated spectra for the endmember 1 (i.e., "Carnallite NMNH98011") in the case of $\overline{\text{SNR}} = 10$ in Fig. 2. It is clear that GLNMF and $\ell_{1,2}$-NMF provide relatively better results than other methods especially in the first 100 bands, and the estimated spectral curve by our GLNMF is more close to the reference spectral curve than $\ell_{1,2}$-NMF. Fig. 3 shows the abundance map of endmember 1 obtained by different methods, where the map from the GLNMF can approximate to the original ground-truth map well. When the hyperspectral data are corrupted by noise, the robustness of GLNMF helps to improve both endmember and abundance estimation results.

2) Non-IID Noise: In real situations, some spectral bands may be corrupted by different types of noise [15]. Here, we randomly choose 40 bands to add following non-iid noise.

1) Non-iid Gaussian noise (NG): $G(0, s_t)$, s_t is different for different bands and is generated from uniform distribution $U(0, 0.5)$.
2) Impulse noise (I): "salt and pepper" noise with intensity 0.2 is added in the image of each chosen band.
3) Deadline noise (D): 20 deadlines are added in the image of each chosen band.
4) Stripe noise (S): 10 stripes per band are added and the width of stripe is between 1 and 3.
5) Gaussian + Impulse noise (GI): both 1) and 2).
6) Gaussian + Deadline noise (GD): both 1) and 3).
7) Gaussian + Stripe noise (GS): both 1) and 4).

TABLE II
SAD RESULTS IN THE CASE OF NON-IID NOISE

	NMF	$\ell_{1/2}$-NMF	$\ell_{1,2}$-NMF	CENMF	CIM-NMF	Hub-NMF	GLNMF
NG	0.2492	0.2254	0.2178	0.1627	0.2095	0.2127	**0.1114**
I	0.1496	0.1420	0.1373	0.1289	0.1375	0.1284	**0.0943**
D	0.2423	0.2280	0.2237	0.2065	0.1754	0.1826	**0.1670**
S	0.2845	0.2668	0.2566	0.2671	**0.1840**	0.2285	0.2044
GI	0.2515	0.2244	0.2124	0.1659	0.2208	0.2140	**0.1278**
GD	0.2782	0.2480	0.2475	0.2100	0.2108	0.2166	**0.1503**
GS	0.3432	0.3120	0.2926	0.2970	0.2382	0.2688	**0.1843**

TABLE III
SAD OF DIFFERENT METHODS ON THE JASPER DATA SET WITH ALL NOISY BANDS

	NMF	$\ell_{1/2}$-NMF	$\ell_{1,2}$-NMF	CENMF	CIM-NMF	Hub-NMF	GLNMF
Tree	0.2370	0.1839	0.2514	0.2310	0.1603	0.2378	**0.0858**
Water	0.2972	**0.1221**	0.2995	0.2909	0.2758	0.2725	0.2175
Soil	0.3960	0.1472	0.1075	0.3971	0.2736	0.2061	**0.0632**
Road	0.6500	0.5241	0.7676	0.6599	0.7033	0.6930	**0.2541**
Mean	0.3950	0.2443	0.3565	0.3947	0.3532	0.3523	**0.1551**

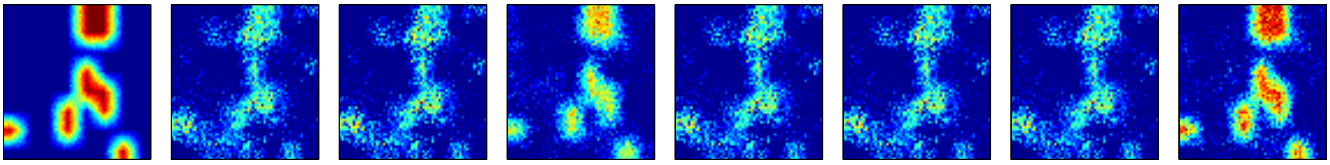
The shape and scale parameters in the GLNMF are set as: $\alpha = -1$ and $c = 1$ for case 1), 4), 7), and set as $\alpha = -1$ and $c = 4$ for case 2), 3), 5), 6). The SAD results in the case of non-iid noise are shown in Table II. When the data are corrupted by only one type of noise, the proposed GLNMF provides the best results in the cases of non-iid Gaussian, impulse, and deadline noise, and the second best result in the case of stripe noise. In the cases of mixture noise, the GLNMF shows the best results. It demonstrates that the GLNMF is also effective for non-iid noise or outliers.

B. Experiments on Jasper Data

The Jasper data set consists of 224 spectral bands ranging from 380 to 2500 nm and has the size of 512 × 614 pixels. In the experiments, a subimage of 100 × 100 pixels is used. In the selected region, there are four targets, i.e., road, soil, water, and tree. Due to dense water vapor and atmospheric effects, there exists 26 noisy bands, i.e., bands 1–3, 108–112, 154–166, and 220–224. To test the robustness of different unmixing methods, we investigate the performance of algorithms on data containing noisy band images (i.e., the original data with $B = 224$) and also data without noisy band images (i.e., $B = 198$).

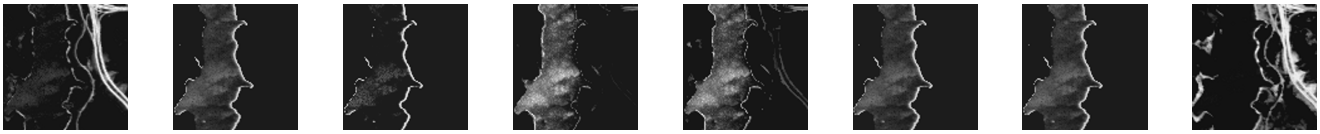
Table III shows the SAD results obtained by different methods on the Jasper data set with noisy bands. It can be seen that all modified NMF methods improve the original NMF at a certain extent, and the proposed GLNMF provides the best result. Except for our GLNMF, other NMF methods shows very bad results on the endmember "Road." The abundance maps corresponding to endmember "Road" obtained by different methods are shown in Fig. 4. GLNMF provides more consistent results with the ground-truth abundance while other methods lose key information.

Fig. 5(a) shows the weights of different bands estimated by GLNMF, where the known noisy bands, such as bands 1–3, 108–112, 154–166, and 220–224, are assigned small weights. By adaptively assigning small weights to noisy bands, our proposed GLNMF is much more robust to noise than traditional NMF models. In the proposed GLNMF, the parameter



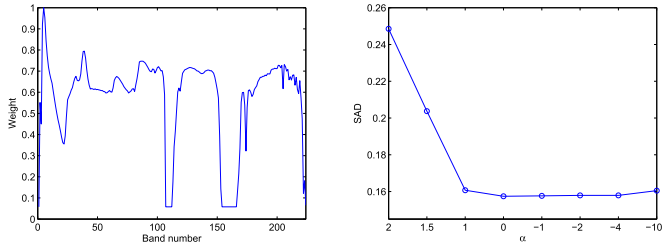
(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3. Abundance maps for endmember 1. (a) Ground-truth, (b) NMF, (c) $\ell_{1/2}$-NMF, (d) $\ell_{1,2}$-NMF, (e) CENMF, (f) CIMNMF, (g) Huber-NMF, and (h) GLNMF.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4. Abundance maps for endmember "Road." (a) Ground-truth, (b) NMF, (c) $\ell_{1/2}$-NMF, (d) $\ell_{1,2}$-NMF, (e) CENMF, (f) CIMNMF, (g) Huber-NMF, and (h) GLNMF.



(a) (b)

Fig. 5. Band weights and the effect of shape parameter on Jasper with the noisy bands: (a) Band weights and (b) SAD versus the shape parameter α.

TABLE IV

SAD OF DIFFERENT METHODS ON THE JASPER DATA SET WITHOUT THE NOISY BANDS

	NMF	$\ell_{1/2}$-NMF	$\ell_{1,2}$-NMF	CENMF	CIM-NMF	Hub-NMF	GLNMF
Tree	0.2694	0.2385	0.2059	0.2597	0.2091	0.2566	**0.0828**
Water	0.3051	**0.1408**	0.2983	0.3089	0.2946	0.2875	0.2005
Soil	0.4606	0.1561	0.3150	0.4768	0.2915	0.2173	**0.0625**
Road	0.5529	0.5436	0.6339	0.5616	0.7587	0.7021	**0.1981**
Mean	0.3970	0.2697	0.3633	0.4018	0.3885	0.3659	**0.1359**

α decides the shape of the general loss. Fig. 5(b) shows the performance of GLNMF at different shape parameters. It can be seen that GLNMF shows good performance when α is smaller than 1.

The SAD results obtained by different methods on the Jasper data set without noisy bands are shown in Table IV, where our proposed GLNMF also shows the overall best result. By deleting 26 noisy bands, the results on the remaining 198 bands are slightly better than those on the full 224 bands.

V. CONCLUSION

In this letter, we have proposed a GLNMF model for the unmixing of hyperspectral images. The general loss function is a superset of many common robust loss functions and is very flexible in handling different types of noise. By choosing a negative shape parameter, the influence function of the general loss is bounded and decreases, and hence, the general loss becomes a very robust loss. Experimental results on simulated and real hyperspectral data sets demonstrate that our proposed GLNMF model can effectively suppress large noise in the real data and is robust for noisy bands.

REFERENCES

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2000, pp. 556–562.

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

[3] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using L21-norm," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2011, pp. 673–682.

[4] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 554–569, Apr. 2013.

[5] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy-based metric," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4027–4039, Nov. 2015.

[6] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[7] X.-R. Feng, H.-C. Li, J. Li, Q. Du, A. Plaza, and W. J. Emery, "Hyperspectral unmixing using sparsity-constrained deep nonnegative matrix factorization with total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6245–6257, Oct. 2018.

[8] R. Huang, X. Li, and L. Zhao, "Spectral–spatial robust nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8235–8254, Oct. 2019.

[9] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.

[10] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation-based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.

[11] Y. Ma, C. Li, X. Mei, C. Liu, and J. Ma, "Robust sparse hyperspectral unmixing with $\ell_{2,1}$ norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1227–1239, Mar. 2017.

[12] J. T. Barron, "A general and adaptive robust loss function," 2017, *arXiv:1701.03077*. [Online]. Available: http://arxiv.org/abs/1701.03077

[13] 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.

[14] X. Li, Q. Lu, Y. Dong, and D. Tao, "Robust subspace clustering by cauchy loss function," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2067–2078, Jul. 2019.

[15] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, Mar. 2018.