CAUCHY NMF FOR HYPERSPECTRAL UNMIXING

Jiangtao Peng[1], *Fan Jiang*[1], *Weiwei Sun*[2], *Yicong Zhou*[3]

[1]Faculty of Mathematics and Statistics, Hubei University
[2]Department of Geography and Spatial Information Techniques, Ningbo University
[3]Department of Computer and Information Science, University of Macau

ABSTRACT

Non-negative matrix factorization (NMF) is a classical hyperspectral unmixing model which minimizes the Euclidean distance between the hyperspectral data matrix and its low rank approximation (i.e., the product of endmember matrix and abundance matrix), and it fails when applied to noisy data because the loss function is sensitive to outliers. In this paper, we propose a Cauchy NMF (CauchyNMF) model for hyperspectral unmixing which uses a Cauchy loss function (CLF) to replace the traditional least-squares loss. Compared with the least-squares loss, CLF can penalize the noise term for suppressing the large noise mixed in the real data and thus is much more robust. Experimental results on simulated and real hyperspectral data sets demonstrate that our proposed CauchyNMF method is more accurate and robust than existing NMF methods, especially in the case of heavy noise.

Index Terms— Hyperspectral unmixing, nonnegative matrix factorization, Cauchy loss function

1. INTRODUCTION

Hyperspectral unmixing (HU) is a classical and important task in the field of the hyperspectral image processing. The aim of HU is to decompose a mixed spectrum into pure materials' spectra (endmembers) and their corresponding fractions (abundances) [1]. In practice, the endmember and abundance matrices are usually unknown, it needs to decompose a measured mixture spectrum into two unknown signals. This is similar to the blind source separation (BSS) problem, and can be solved by the nonnegative matrix factorization (NMF) methods [2, 3, 4, 5, 6, 7].

The NMF-based hyperspectral unmixing methods decompose a nonnegative hyperspectral data matrix into a nonnegative endmember matrix and a nonnegative abundance matrix. To make the result physical meaningful, a sum-to-one constraint is usually imposed on the abundance vector for each pixel. In addition, many other constraints either on the endmember matrix or on the abundance matrix have also been added into the NMF framework to improve the unmixing performance, such as minimum volume constraint on the endmembers [6], $\ell_{1/2}$ sparsity constraint on the abundance matrix [3], 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 (i.e., 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 [8, 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 paper, we use a robust Cauchy loss function (CLF) to replace the least squares loss and propose a Cauchy NMF method (CauchyNMF) for hyperspectral unmixing. Compared with the conventional least squares loss, the influence function of CLF has an upper bound. Therefore, it can alleviate the influence of a single element, especially the element with a large noise. In the implementation of CauchyNMF, an auxiliary vector can automatically obtained to indicate the importance of each hyperspectral band. For noisy band, the corresponding weight is relatively small and thus CauchyNMF is more robust to the noise.

2. NMF UNMIXING MODEL

Under the linear mixing mechanism [1, 3], each observed pixel $\mathbf{y} \in \mathcal{R}^{B\times 1}$ can be represented as a linear combination of several spectral signatures called endmembers, i.e., $\mathbf{x}_1, \cdots, \mathbf{x}_P$,

$$\mathbf{y} = \mathbf{x}_1 w_1 + \cdots + \mathbf{x}_P w_P + \mathbf{e} = \mathbf{X}\mathbf{w} + \mathbf{e}, \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_P] \in \mathcal{R}^{B\times P}$ is a nonnegative spectral signature matrix, $\mathbf{w} = [w_1; \cdots; w_P] \in \mathcal{R}^{P\times 1}$ is the abun-

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61871177, 11771130, 41971296, 41671342, by the Science and Technology Development Fund, Macau SAR (File no. 189/2017/A3), and by the University of Macau (File no. MYRG2018-00136-FST), by Zhejiang Provincial Natural Science Foundation of China (LR19D010001), and by the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 18R05. Corresponding author: Weiwei Sum.

dance fraction for each endmember, and $\mathbf{e}$ is the additive noise vector.

Assume that there are N pixels in the HSI. The linear mixing model (LMM) can be written as:

$$\mathbf{Y} = \mathbf{X}\mathbf{W} + \mathbf{E}, \tag{2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N] \in \mathcal{R}^{B\times N}$ is the hyperspectral data, $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_N] \in \mathcal{R}^{P\times N}$ is the endmember abundance matrix, and $\mathbf{E}$ is the noise matrix.

In reality, only the hyperspectral data $\mathbf{Y}$ is available. So, the unmixing problem can be considered as a blind source separation problem and can be solved by the NMF model [2]. Considering the nonnegative constraints and abundance sum-to-one constraint [1], the NMF-based hyperspectral unmixing model can be written as:

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{W}} \|\mathbf{Y} - \mathbf{X}\mathbf{W}\|_F^2, \\ &\text{s.t.,}\ \mathbf{X}_{bp} \geq 0, \mathbf{W}_{pn} \geq 0, \forall b, p, n \\ &\sum_{p=1}^{P} \mathbf{W}_{pn} = 1,\ \ n = 1, \cdots, N \end{aligned} \tag{3}$$

where $\|\cdot\|_F$ denotes the the Frobenius norm.

The NMF unmixing model (3) can be solved by the multiplicative update rules or additive update rules [2, 12].

3. CAUCHY NMF

3.1. Cauchy loss function

For a loss function ϕ, its influence function is defined as [13]:

$$\psi(x) = \frac{\partial \phi(x)}{\partial x}, \tag{4}$$

which is used to measure the effect of changing the point of the sample on the value of the parameter estimation. For a robust estimator, its influence function should not be sensitive to the increase of the error [13].

For the traditional least squares (LS) loss $\phi(x) = x^2$, its influence function is: $\psi(x) = 2x$. This means that the influence of a sample on the estimate grows linearly as the error increases. So, the least squares estimator is not robust to the noise.

Cauchy loss function (CLF) is a robust estimator which is defined as:

$$\phi(x) = \log\left(1 + (x/c)^2\right), \tag{5}$$

where c is a constant. CLF's influence function is:

$$\psi(x) = \frac{\partial \phi(x)}{\partial x} = \frac{2x}{x^2 + c^2}. \tag{6}$$

The loss function and influence function of LS and CLF are shown in Fig. 1. It can be clearly seen that CLF loss can suppress the large noise and its influence function has the upper bound and tends to zero with the increase of the error.

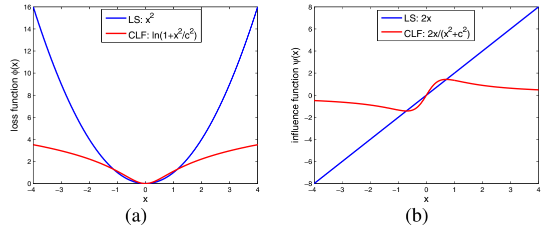

Fig. 1. The comparison of LS and CLF: (a) loss function, (b) influence function.

3.2. Cauchy loss based NMF

Considering the least squares loss function in the $\ell_{1/2}$-NMF unmixing model (3) is sensitive to noise or outliers in the data $\mathbf{Y}$, we introduce the Cauchy loss function to replace the least squares loss in the model (3) and obtain the following objective function:

$$\max -\log\left(1 + \frac{\|\mathbf{Y} - \mathbf{X}\mathbf{W}\|_F^2}{c^2}\right), \tag{7}$$

It can be seen that the objective function in (7) is non-convex and nonlinear. So, it is difficult to optimize directly. Here, we use the half-quadratic technique [14, 8] to solve the optimization problem (7). According to the property of conjugate function, we have:

Proposition 1 *There exists a convex conjugate function g of f, such that,*

$$f(a) = \max_b \{ab - g(b)\}, \tag{8}$$

and the maximizer is $b^* = f'(a)$.

According to Proposition 1, let $f(x) = -\log(1 + x)$ and a_k be the k-th row of residual matrix $\mathbf{E}$, i.e., $a^k = \|\mathbf{Y}^k - (\mathbf{X}\mathbf{W})^k\|_2^2$, the objective function in (7) is changed to:

$$\max \sum_{k=1}^{B} \left(1 + \frac{\|\mathbf{Y}^k - (\mathbf{X}\mathbf{W})^k\|_2^2}{c^2}\right) b_k - g(b_k) \tag{9}$$

In practice, certain sparsity constraint can be performed on the abundance matrix $\mathbf{W}$ to restrict the feasible solution set of model (9) [3, 5, 7]. In this paper, we add an $\ell_{1/2}$-norm constraint on $\mathbf{W}$ and propose a CLF-based $\ell_{1/2}$-NMF (CauchyNMF) as:

$$\begin{aligned} &\max_{\mathbf{X},\mathbf{W}} \sum_{k=1}^{B} \left(1 + \frac{\|\mathbf{Y}^k - (\mathbf{X}\mathbf{W})^k\|_2^2}{c^2}\right) b_k - g(b_k) - \lambda \|\mathbf{W}\|_{\frac{1}{2}} \\ &\text{s.t.,}\ \mathbf{X}_{bp} \geq 0, \mathbf{W}_{pn} \geq 0, \forall b, p, n \\ &\sum_{p=1}^{P} \mathbf{W}_{pn} = 1,\ \ n = 1, \cdots, N \end{aligned} \tag{10}$$

where λ is a regularization parameter.

According to Proposition 1, for fixed $\mathbf{X}$ and $\mathbf{W}$, the maximizer of (10) is reached at

$$b_k = -\frac{1}{1 + \frac{\|\mathbf{Y}^k - (\mathbf{XW})^k\|_2^2}{c^2}}, k = 1, \cdots, B \tag{11}$$

When the auxiliary vector $\mathbf{b}$ is obtained, the optimization problem (10) can be represented as a weighted $\ell_{1/2}$-NMF problem and the objective function has the form

$$\min_{\widetilde{\mathbf{X}}, \mathbf{W}} \|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_{1/2}, \tag{12}$$

where $\widetilde{\mathbf{Y}} = \mathbf{U}^{\frac{1}{2}}\mathbf{Y}$ and $\widetilde{\mathbf{X}} = \mathbf{U}^{\frac{1}{2}}\mathbf{X}$, and $\mathbf{U}$ is a diagonal matrix whose diagonal element is:

$$\mathbf{U}_{kk} = \frac{1}{c^2 + \|\mathbf{Y}^k - (\mathbf{XW})^k\|_2^2}$$

It is clear that model (12) is also an $\ell_{1/2}$-NMF model and can be solved by the multiplicative update rule. The final endmember matrix is: $\mathbf{X} = \mathbf{U}^{-\frac{1}{2}}\widetilde{\mathbf{X}}$.

4. EXPERIMENTAL RESULTS

The proposed CauchyNMF method is compared with the following unmixing methods: NMF [2], $\ell_{1/2}$-NMF [3], $\ell_{2,1}$-NMF [4], and correntropy based NMF (CENMF) [5] The spectral angle distance (SAD) and the root-mean-square error (RMSE) are used to evaluate the performance of different unmixing methods.

We first test the proposed method on a synthetic data which consists of seven spectral signatures chosen from the United States Geological Survey (USGS) digital spectral library. The abundances of these endmembers are generated based on Refs. [6, 7], and recorded in matrix $\mathbf{W}$. By multiplying the endmember matrix $\mathbf{X}$ and the abundance matrix $\mathbf{W}$, we obtain the synthetic data: $\mathbf{Y} = \mathbf{XW}$.

In order to evaluate the robustness of different algorithms, Gaussian noise is added to the synthetic data $\mathbf{Y}$. The SNRs of bands are generated using the normal distribution $\mathrm{SNR} \sim \mathcal{N}(\overline{\mathrm{SNR}}, \epsilon^2)$, where $\overline{\mathrm{SNR}} \in \{5, 10, 15, 20\}$ and $\epsilon = 5$. Mean of SAD and RMSE results over 10 randomly runs are reported in Tables 1 and 2, respectively.

Table 1. SAD results on the noisy synthetic data.

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CauchyNMF
5	0.2023	0.1365	0.1821	0.1874	**0.0813**
10	0.0758	0.0631	0.0739	0.0683	**0.0448**
15	0.0359	0.0352	0.0360	0.0361	**0.0263**
20	0.0198	0.0193	0.0198	0.0198	**0.0188**

Table 2. RMSE results on the noisy synthetic data.

SNR	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CauchyNMF
5	0.1685	0.1674	0.1734	0.1691	**0.1329**
10	0.1074	0.1018	0.1080	0.1004	**0.0735**
15	0.0633	0.0631	0.0636	0.0637	**0.0415**
20	0.0381	0.0377	0.0381	0.0382	**0.0262**

It can be clearly seen that the performance of each algorithm improves as the increase of SNR. The modified NMF methods, such as $\ell_{1/2}$-NMF and $\ell_{2,1}$-NMF, improve the N-MF in the case of large noise (i.e., small SNR). However, in the case of small noise, these algorithms generate similar results with the original NMF. Our proposed CauchyNMF method provide better results than other methods in terms of both SAD and RMSE. We further show the abundance map of endmember 5 (i.e., 'Erionite+Merlinoit GDS144') obtained by different methods in Fig. 2, where the map of our SpN-MFB is very close to the original ground-truth map.

Next, we consider a real world Jasper hyperspectral data set. This data set consists of 224 spectral bands and has 512 $\times$ 614 pixels. The experiment uses a subimage of 100 $\times$ 100 pixels, which contain four targets, i.e., road, soil, water, and tree. Due to dense water vapor and atmospheric effects, there exists some noisy bands, including the bands 1–3, 108–112, 154–166 and 220–224. We show the performance of different unmixing methods on this noisy Jasper data set in Table 3, where our proposed CauchyNMF shows the best overall result. In particular, on the endmember 'Road', the existing N-MF methods show very poor results while CauchyNMF show relatively better result.

Table 3. SAD of different methods on the JASPER data set.

	NMF	$\ell_{1/2}$-NMF	$\ell_{2,1}$-NMF	CENMF	CauchyNMF
Tree	0.3666	0.3226	0.3443	0.3087	**0.0905**
Water	0.3024	**0.0949**	0.2818	0.2955	0.2184
Soil	0.2956	0.1198	0.2462	0.4288	**0.0633**
Road	0.6839	0.5381	0.7476	0.6914	**0.2563**
Mean	0.4121	0.2689	0.4050	0.4311	**0.1571**

5. CONCLUSION

In this paper, we have proposed a Cauchy loss function based robust nonnegative matrix factorization (CauchyNM-F) method for the unmixing of hyperspectral images. The Cauchy loss function is more robust than traditional lease squares loss and can be effectively suppress large noise in the real data. Experimental results on simulated and real

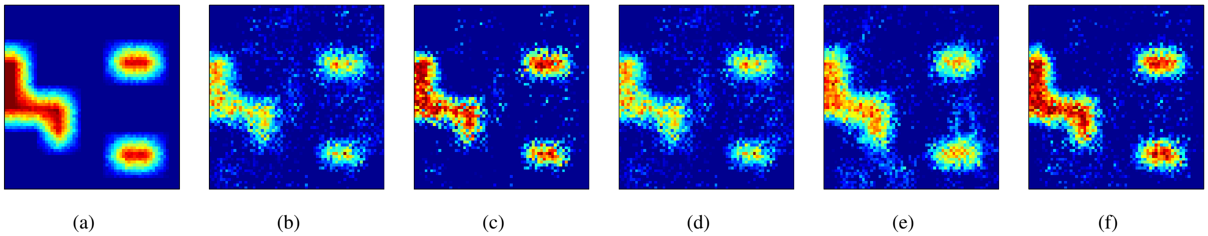

Fig. 2. The abundance maps for endmember 5. (a) Ground-truth, (b) NMF, (c) $\ell_{1/2}$-NMF, (d) $\ell_{2,1}$-NMF, (e) CENMF, (f) CauchyNMF.

hyperspectral data sets have demonstrated that the proposed CauchyNMF method is more accurate and robust than existing NMF methods.

6. REFERENCES

[1] J. M. Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot, "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Top. Appl. Ear. Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] D. Danie Lee and H. Sebastian Seung, "Algorithms for nonnegative matrix factorization," in *NIPS*, 2000, pp. 556–562.

[3] Yuntao Qian, Sen Jia, Jun Zhou, and Antonio 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.

[4] Deguang Kong, Chris Ding, and Heng Huang, "Robust nonnegative matrix factorization using ℓ_{21}-norm," in *ACM CIKM*, 2011, pp. 673–682.

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

[6] Lidan Miao and Hairong 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] Xinru Feng, Hengchao Li, Jun Li, Qian Du, Antonio Plaza, and William 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] Jiangtao Peng and Qian Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.

[9] Jiangtao Peng, Weiwei Sun, and Qian 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] Jiangtao Peng, Luoqing Li, and Yuan Yan 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] Yong Ma, Chang Li, Xiaoguang Mei, Chengyin Liu, and Jiayi Ma, "Robust sparse hyperspectral unmixing with $\ell_{2,1}$ norm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1227–1239, 2016.

[12] O. Patrik Hoyer, "Non-negative sparse coding," in *IEEE Workshop Neural Netw. Signal Process. XII*, 2002, pp. 557–565.

[13] Xuelong Li, Quanmao Lu, Yongsheng Dong, and Dacheng Tao, "Robust subspace clustering by cauchy loss function," *IEEE Trans. Neur. Netw. Learn. Syst.*, vol. 30, no. 7, pp. 2067–2078, Jul. 2019.

[14] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Scientific computing*, vol. 27, no. 3, pp. 937–966, 2005.