Correntropy-Based Autoencoder-Like NMF With Total Variation for Hyperspectral Unmixing

Xin-Ru Feng, *Graduate Student Member, IEEE* , Heng-Chao Li, *Senior Member, IEEE*, Shuang Liu, and Hongyan Zhang, *Senior Member, IEEE*

***Abstract*— In order to unmix the hyperspectral imagery (HSI) with better performance, this letter proposes a correntropy-based autoencoder-like nonnegative matrix factorization (NMF) (CANMF) with total variation (CANMF-TV) method. NMF is extensively applied to unmix the mixed pixels. However, it only reconstructs the original data from the abundances in endmember space. To directly project the original data space into the endmember space, and then achieve the abundance matrix, we first exploit an autoencoder-like NMF for hyperspectral unmixing, which integrates both *decoder* and *encoder*. Considering that HSI is typically degraded by noise, the correntropy-induced metric (CIM) is introduced to construct a CANMF model. In addition, TV regularizer is imposed into the CANMF model so as to preserve the spatial-contextual information by promoting the piecewise smoothness of abundances. Finally, a series of experiments are conducted on both synthetic and real data sets, demonstrating the effectiveness of the proposed CANMF-TV method over comparison.**

***Index Terms*— Correntropy-induced metric, hyperspectral unmixing, nonnegative matrix factorization, total variation.**

I. Introduction

NONNEGATIVE matrix factorization (NMF) is a promising technique that has the capacity of extracting effective low-dimensional feature representation. It approximates a nonnegative matrix by the product of two nonnegative ones, i.e., one basis matrix and one coefficient matrix. As a consequence, the results are yielded with clear physical interpretation. To this end, NMF has drawn considerable attention in a variety of domains, including, but not limited to, hyperspectral unmixing [1], clustering [2], and face recognition [3].

In particular, NMF has been successfully applied to hyperspectral unmixing based on linear mixture model (LMM). Hyperspectral unmixing aims to obtain a collection of constituent materials (called endmembers) along with their proportions (called abundances) from a hyperspectral imagery (HSI) with mixed pixels [4], thereby benefiting subsequent applications. LMM assumes that each pixel spectrum is a linear combination of endmember signatures and their associated abundances [5]. Based on LMM, there are numerous unmixing approaches, such as vertex component analysis (VCA) [6], sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) [7], and minimum volume simplex analysis (MVSA) [8]. In contrast, the NMF-based methods extract the spectra of endmembers with their abundances simultaneously from the decomposed matrices directly. By making use of prior knowledge, additional constraints are incorporated into NMF for further improving the unmixing performance. Specifically, two properties of hyperspectral data are explored in general. First, the abundance matrix has sparsity owing that the distribution of each endmember is sparse. Consequently, numerous works are developed by imposing sparsity regularizer, such as single sparsity [9], collaborative sparsity [10], and spatial group sparsity [11]. Second, each row of abundance matrix should be piecewise smoothness due to the spatial correlation among the neighboring pixels. Under this assumption, Yang *et al.* [12] used a graph regularizer to preserve the structural information. In [13], hypergraph regularizer was exploited to capture the similarity among the spatial nearby pixels. Moreover, by considering the spatial information and the sparsity, He *et al*. [14] proposed a total variation regularized reweighted sparse NMF (TV-RSNMF) algorithm.

In this letter, we propose a correntropy-based autoencoder-like NMF (CANMF) with TV (CANMF-TV) method to address the problem of hyperspectral unmixing. Our main contributions are threefold: 1) an autoencoder-like NMF is firstly exploited to unmix the mixed pixels, which is motivated by the nonnegative symmetric encoder-decoder model originally proposed by Sun *et al.* [15]. Although the reported NMF-based unmixing methods can achieve satisfactory performance, these algorithms only consider the reconstruction of the original data from the abundances in an endmember space, i.e., *decoder*. The autoencoder-like NMF architecture integrates *decoder* and *encoder* together such that it can directly project the original data into the endmember matrix to obtain the abundance matrix. Different from the autoencoder-based unmixing approaches such as [16], the autoencoder-like NMF does not need to design or select an activation function. That is, it is constructed in a linear way, which is consistent with the assumption of LMM. Besides, our cost function is formed to reduce the sum of residuals for *decoder* and *encoder*, rather than to maximize the similarity between the input and

Manuscript received April 19, 2020; revised June 23, 2020; accepted August 24, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61871335, in part by the Fundamental Research Funds for the Central Universities under Grant 2682020XG02 and Grant 2682020ZT35, and in part by the Postgraduate Scientific Research & Innovation Competition Project of Southwest Jiaotong University. *(Corresponding author: Heng-Chao Li.)*

Xin-Ru Feng, Heng-Chao Li, and Shuang Liu are with the School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China (e-mail: lihengchao_78@163.com).

Hongyan Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China.

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.2020.3020896

output layers; 2) for improving the robustness of NMF, several models have been reported based on a kind of metrics, e.g., L_1-norm regularizer [17], $L_{2,1}$-norm regularizer [18], and correntropy-induced metric (CIM) [19], [20]. Among them, CIM is effective to handle the data with the noise. Actually, HSI inevitably suffers from noise. To mitigate its adverse effect, CIM is introduced to construct a CANMF model; and 3) since more similar abundance values are often presented among the adjacent pixels, the contextual dependence exist in abundance matrix. Therefore, TV regularizer is further imposed to promote the piecewise smoothness of abundances for improving the unmixing performance. Finally, experiments are conducted to testify the validity of the proposed CANMF-TV method on the synthetic and real hyperspectral data sets.

II. Methodology

In this section, the proposed method is described in detail, including the model formulation and its optimization.

A. Formulation of Proposed CANMF-TV Model

Given a nonnegative observation $\mathbf{X}$ with B bands and P pixels, NMF aims at decomposing the matrix $\mathbf{X}$ into the product of two nonnegative matrices, i.e., $\mathbf{X} \approx \mathbf{AS}$, here $\mathbf{A} \in \mathbb{R}^{B\times M}$ and $\mathbf{S} \in \mathbb{R}^{M\times P}$ denote the spectral signatures with M endmembers and their corresponding abundances, respectively. Meanwhile, the abundance matrix $\mathbf{S}$ should satisfy both the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC), i.e., $S_{mp} \geq 0$ and $\sum_m S_{mp} = 1$, where $m = 1, 2, \ldots, M$ and $p = 1, 2, \ldots, P$.

The NMF unmixing is mainly based on reconstructing a network, namely decoder in autoencoder, while lacking an encoder that transforms the observation data into the corresponding endmember space, thus not having the desirable unmixing performance. To address this issue, by following the idea of Sun *et al.* [15], here we make the first attempt to exploit the autoencoder- like NMF for hyperspectral unmixing. Generally, the acquired HSI is inevitably degraded by the noise. In order to improve the robustness of the model, a correntropy-based robust measure is introduced to construct the CANMF model, whose cost function is designed as

$$\begin{aligned}\mathcal{C} = &-\sum_{b,p}\exp\left\{-\frac{(\mathbf{X}-\mathbf{AS})_{bp}^2}{2\sigma_1^2}\right\}\\ &-\alpha\sum_{m,p}\exp\left\{-\frac{\left(\mathbf{S}-\mathbf{A}^T\mathbf{X}\right)_{mp}^2}{2\sigma_2^2}\right\}\\ &\text{s.t. } \mathbf{S}^T\mathbf{1}_M = \mathbf{1}_P, \quad \mathbf{A}\geq\mathbf{0},\ \mathbf{S}\geq\mathbf{0}\end{aligned} \tag{1}$$

where the first term represents the error for *decoder*, the second term is the error for *encoder*, $\mathbf{X}_{bp}$ is an element in the bth row, pth column of matrix $\mathbf{X}$, $(\cdot)^T$ denotes the transpose operation to a matrix, $\mathbf{1}_M$ is column vector of all ones with M elements, σ_1 and σ_2 denote the bandwidth of the Gaussian kernel. α is a tradeoff parameter to control the contribution of the *encoder* term. When $\alpha = 0$, the model (1) is degraded into the basic correntropy-based NMF. The symmetry of model (1) together with the ANC can promote indirectly the sparsity of abundance matrix.

In view of the fact that stronger spatial information is an important aspect of HSI, which can further improve the unmixing performance by effectively utilizing such information. Specifically, a TV regularization is integrated into the model (1) to obtain abundance maps with piecewise smoothness. Therefore, we require to minimize the following cost function:

$$\begin{aligned}\mathcal{C} = &-\sum_{b,p}\exp\left\{-\frac{(\mathbf{X}-\mathbf{AS})_{bp}^2}{2\sigma_1^2}\right\}\\ &-\alpha\sum_{m,p}\exp\left\{-\frac{(\mathbf{S}-\mathbf{A}^T\mathbf{X})_{mp}^2}{2\sigma_2^2}\right\}+\lambda\|\mathbf{S}\|_{\text{HTV}}\\ &\text{s.t. } \mathbf{S}^T\mathbf{1}_M = \mathbf{1}_P, \quad \mathbf{A}\geq\mathbf{0},\ \mathbf{S}\geq\mathbf{0}\end{aligned} \tag{2}$$

where λ is a regularized parameter, and we have defined

$$\|\mathbf{S}\|_{\text{HTV}} = \sum_{m=1}^{M}\|\mathcal{F}\mathbf{S}^m\|_{\text{TV}} \tag{3}$$

where $\mathbf{S}^m$ represents the mth row of the matrix $\mathbf{S}$, and $\mathcal{F}$ denotes the reshape operation from a vector with P pixels to a matrix with $u \times v$. For a gray-level image $\mathbf{Y}$ of size $u \times v$, the isotropic TV [21] is defined as

$$\|\mathbf{Y}\|_{\text{TV}} = \sum_{u,v}\sqrt{(\mathbf{Y}_{uv}-\mathbf{Y}_{u+1,v})^2+(\mathbf{Y}_{uv}-\mathbf{Y}_{u,v+1})^2}. \tag{4}$$

B. Optimization

The model (2) is nonconvex when optimizing all variables at the same time. Therefore, the gradient descent (GD)-based method is adopted to alternately update the underlying variables. First, an auxiliary variable $\mathbf{U}$ is introduced into the model (2) and equality constraint $\mathbf{S}^T\mathbf{1}_M = \mathbf{1}_P$ is absorbed into the cost function, that is,

$$\begin{aligned}\mathcal{C} = &-\sum_{b,p}\exp\left\{-\frac{(\mathbf{X}-\mathbf{AS})_{bp}^2}{2\sigma_1^2}\right\}\\ &-\alpha\sum_{m,p}\exp\left\{-\frac{(\mathbf{S}-\mathbf{A}^T\mathbf{X})_{mp}^2}{2\sigma_2^2}\right\}+\lambda\|\mathbf{U}\|_{\text{HTV}}\\ &+\frac{\delta}{2}\|\mathbf{S}^T\mathbf{1}_M-\mathbf{1}_P\|_2^2, \quad \text{s.t. } \mathbf{S}=\mathbf{U},\ \mathbf{A}\geq\mathbf{0},\ \mathbf{S}\geq\mathbf{0}\end{aligned} \tag{5}$$

in which $\|\cdot\|_2$ is l_2-norm. The augmented Lagrangian function of the above equation is

$$\begin{aligned}\mathcal{C} = &-\sum_{b,p}\exp\left\{-\frac{(\mathbf{X}-\mathbf{AS})_{bp}^2}{2\sigma_1^2}\right\}+\text{Tr}(\mathbf{\Psi}^T\mathbf{A})+\text{Tr}(\mathbf{\Gamma}^T\mathbf{S})\\ &-\alpha\sum_{m,p}\exp\left\{-\frac{(\mathbf{S}-\mathbf{A}^T\mathbf{X})_{mp}^2}{2\sigma_2^2}\right\}+\lambda\|\mathbf{U}\|_{\text{HTV}}\\ &+\frac{\delta}{2}\|\mathbf{S}^T\mathbf{1}_M-\mathbf{1}_P\|_2^2+\frac{\mu}{2}\left\|\mathbf{U}-\left(\mathbf{S}+\frac{\mathbf{D}}{\mu}\right)\right\|_F^2\end{aligned} \tag{6}$$

where $\|\cdot\|_F$ is the Frobenius norm of the matrix, $\mathbf{D}$, $\mathbf{\Psi}$, and $\mathbf{\Gamma}$ are the Lagrange multipliers in matrix format, and δ and μ

are the penalty parameters. In this case, optimizing problem (6) is split into three subproblems, each of which is solved separately as follows.

1) *Update Rule for* $\mathbf{A}$*:* To find the optimal endmember matrix, the partial derivative for (6) with respect to $\mathbf{A}$ is

$$\frac{\partial \mathcal{C}}{\partial \mathbf{A}} = \left[\mathbf{K}_1 \odot \frac{(\mathbf{AS}-\mathbf{X})}{\sigma_1^2}\right]\mathbf{S}^T + \alpha\mathbf{X}\left[\mathbf{K}_2 \odot \frac{(\mathbf{A}^T\mathbf{X}-\mathbf{S})}{\sigma_2^2}\right]^T + \boldsymbol{\Psi} \quad (7)$$

with

$$\mathbf{K}_1 = \exp\left\{-\frac{(\mathbf{X}-\mathbf{AS})\odot(\mathbf{X}-\mathbf{AS})}{2\sigma_1^2}\right\} \quad (8a)$$

$$\mathbf{K}_2 = \exp\left\{-\frac{(\mathbf{S}-\mathbf{A}^T\mathbf{X})\odot(\mathbf{S}-\mathbf{A}^T\mathbf{X})}{2\sigma_2^2}\right\} \quad (8b)$$

where $\odot$ denotes the elementwise multiplication. Thus, by using Karush-Kuhn–Tucker (KKT) condition $\mathbf{A}\odot\boldsymbol{\Psi}=\mathbf{0}$ and setting $\partial\mathcal{C}/\partial\mathbf{A}=0$, update rule for $\mathbf{A}$ is derived as

$$\mathbf{A} \leftarrow \mathbf{A}\odot\left\{\left(\mathbf{K}_1\odot\left(\mathbf{X}/\sigma_1^2\right)\right)\mathbf{S}^T + \alpha\mathbf{X}\left(\mathbf{K}_2\odot\left(\mathbf{S}/\sigma_2^2\right)\right)^T\right\} \oslash \left\{\left[\mathbf{K}_1\odot\left(\mathbf{AS}/\sigma_1^2\right)\right]\mathbf{S}^T + \alpha\mathbf{X}\left[\mathbf{K}_2\odot\left(\mathbf{A}^T\mathbf{X}/\sigma_2^2\right)\right]^T\right\} \quad (9)$$

where $\oslash$ denotes the elementwise division.

2) *Update Rule for* $\mathbf{S}$*:* For abundance matrix, the partial derivative for (6) with respect to $\mathbf{S}$ is expressed as

$$\frac{\partial \mathcal{C}}{\partial \mathbf{S}} = \mathbf{A}^T\left[\mathbf{K}_1\odot\frac{(\mathbf{AS}-\mathbf{X})}{\sigma_1^2}\right] + \alpha\left[\mathbf{K}_2\odot\frac{(\mathbf{S}-\mathbf{A}^T\mathbf{X})}{\sigma_2^2}\right] + \mathbf{D} + \mu(\mathbf{S}-\mathbf{U}) + \delta\left(\mathbf{1}_M\mathbf{1}_M^T\mathbf{S} - \mathbf{1}_M\mathbf{1}_P^T\right) + \boldsymbol{\Gamma}. \quad (10)$$

Similarly, update rule for $\mathbf{S}$ is given as

$$\mathbf{S} \leftarrow \mathbf{S}\odot\left\{\mathbf{A}^T\left(\mathbf{K}_1\odot\left(\mathbf{X}/\sigma_1^2\right)\right) + \alpha\left[\mathbf{K}_2\odot\left(\mathbf{A}^T\mathbf{X}/\sigma_2^2\right)\right] + \mu\mathbf{U} + \delta\mathbf{1}_M\mathbf{1}_P^T\right\} \oslash \left\{\mathbf{A}^T\left[\mathbf{K}_1\odot\left(\mathbf{AS}/\sigma_1^2\right)\right] + \alpha\left(\mathbf{K}_2\odot\left(\mathbf{S}/\sigma_2^2\right)\right) + \mathbf{D} + \mu\mathbf{S} + \delta\mathbf{1}_M\mathbf{1}_M^T\mathbf{S}\right\}. \quad (11)$$

3) *Update Rules for* $\mathbf{U}$*:* In addition, the auxiliary variable $\mathbf{U}$ can be updated by

$$\mathbf{U} = \arg\min_{\mathbf{U}} \frac{\lambda}{\mu}\|\mathbf{U}\|_{\mathrm{HTV}} + \frac{1}{2}\left\|\mathbf{U} - \left(\mathbf{S} + \frac{\mathbf{D}}{\mu}\right)\right\|_F^2. \quad (12)$$

The optimization of (12) can be split into M subproblems, and each subproblem is to solve

$$\mathbf{U}^m = \arg\min_{\mathbf{U}^m} \frac{\lambda}{\mu}\|\mathcal{F}\mathbf{U}^m\|_{\mathrm{TV}} + \frac{1}{2}\left\|\mathcal{F}\mathbf{U}^m - \left(\mathcal{F}\mathbf{S}^m + \frac{\mathcal{F}\mathbf{D}^m}{\mu}\right)\right\|_F^2 \quad (13)$$

here, $\mathbf{U}^m$ denotes the mth row of matrix $\mathbf{U}$. In this letter, generic proximal algorithm presented in [21] is adopted to solve the problem (13).

Algorithm 1 Proposed CANMF-TV Algorithm

Input: Observation matrix $\mathbf{X}$.
Number of endmembers M.
Parameters α, λ.
Initialize: $\mathbf{A}$ and $\mathbf{S}$ by VCA and FCLS, $\mathbf{U}=\mathbf{0}$, $\mathbf{D}=\mathbf{0}$, $\mu=0.01$, $\rho=1.1$, $\mu_{\max}=1000$, $T_{\max}=200$, $\varepsilon=10^{-3}$, $\sigma_1=1$, and $\sigma_2=10$.
Repeat
Calculate $\mathbf{K}_1$ and $\mathbf{K}_2$ by (8a) and (8b).
Update $\mathbf{A}$ by (9).
Recalculate $\mathbf{K}_1$ and $\mathbf{K}_2$ by (8a) and (8b).
Update $\mathbf{S}$ by (11).
Update $\mathbf{U}$ by (13).
Update $\mathbf{D}$ by $\mathbf{D}\leftarrow\mathbf{D}+\mu(\mathbf{S}-\mathbf{U})$.
Update μ by $\mu=\min(\rho\mu,\mu_{\max})$.
until $\|\mathbf{S}-\mathbf{U}\|_\infty<\varepsilon$.
Output: optimal solutions $\mathbf{A}$ and $\mathbf{S}$.

The optimization process for the CANMF-TV model is summarized in Algorithm 1.

III. Experimental Results and Discussion

In this section, the performance of the proposed method is evaluated by using synthetic and real hyperspectral data sets. The CANMF-TV is compared with untied denoising autoencoder with sparsity (uDAS) [16], $L_{1/2}$-RNMF [18], and $L_{1/2}$-NMF [5], where VCA-FCLS is used for initialization and the results are obtained over 10 runs for each method so as to make a reliable comparison. Moreover, all experiments are conducted under the environment of MATLAB R2015b software and computer configuration Intel Core i5 CPU at 2.80 GHz and 8.00 GB RAM. Spectral angle distance (SAD) and root-mean-square error (RMSE) are respectively utilized to endmembers and abundances for quantitative assessment, which can be referred to [5].

A. Synthetic Data Experiments

In order to obtain a synthetic data, we first choose six spectral signatures (i.e., Annite WS660, Illite IMt-1.a, Limonite HS41.3, Meionite WS701, Muscovite GDS108, and Olivine HS285.4B) from the United States Geological Survey (USGS) spectral library, which comprises nearly 500 typical minerals and is available online at http://speclab.cr.usgs.gov/spectral.lib06. Following the generation process described in [9], the synthetic data is created with a size of 64×64 pixels and 224 bands changed from 0.4 to 2.5 μm. Besides, the synthetic data is contaminated by zero-mean Gaussian noise to be more realistic. The noise level can be described by signal-to-noise ratio (SNR).

1) Experiment 1 (Parameter Analysis): First, our proposed method is tested in terms of different values of α and λ, which are tuned in the range $\{0, 0.01, 0.02, 0.05, 0.1, 0.2\}$ and {1e-4, 1e-3, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1}, respectively. In this experiment, the Gaussian noise is added into the synthetic data with SNR = 40 dB. The results yielded by CANMF-TV when α and λ vary are shown in Fig. 1. Overall, the proposed

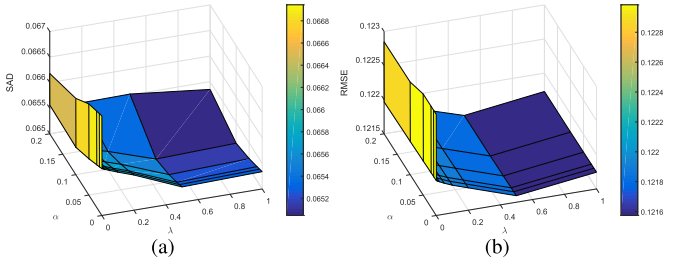

Fig. 1. Performance of CANMF-TV with respect to parameters α and λ in terms of (a) SAD and (b) RMSE.

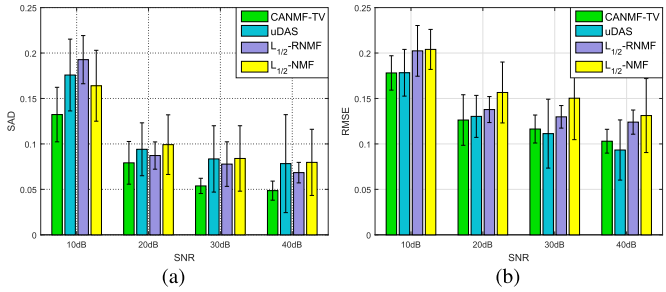

Fig. 2. Comparison of (a) SAD and (b) RMSE for different methods versus the Gaussian noise levels.

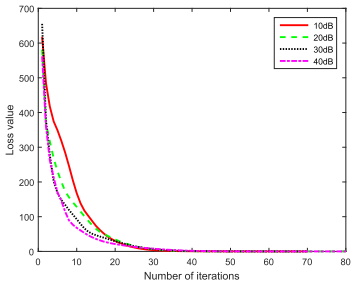

Fig. 3. Convergence curves of CANMF-TV on the synthetic data set.

method performs good for the above varied values of α and λ. It can be found that the optimal performance is achieved at $\alpha = 0.1$ and $\lambda = 0.5$. The performance is improved gradually as the α increases, but deteriorated when α is larger than 0.1. That is, the CANMF-TV performs better than that with $\alpha = 0$, showing the effectiveness of *encoder*. In contrast, both SAD and RMSE values decrease obviously when λ is increased to 0.5. This reveals that it is significant to incorporate the TV regularizer.

2) Experiment 2 (Robustness Analysis to Noise): To verify the robustness to noise contamination of our method, we assign SNR $\in \{10, 20, 30, 40\}$ (dB). Fig. 2(a) illustrates the average SAD values along with standard deviation under different noise levels for each method, and Fig. 2(b) gives the corresponding RMSE results. As can be seen, all methods yield smaller SAD and RMSE values under lower Gaussian noise levels. Nevertheless, CANMF-TV obtains better overall unmixing performance than other considered methods, indicating its significant advantage.

3) Experiment 3 (Convergence Observation and Running Time): In order to demonstrate the convergence of CANMF-TV, Fig. 3 gives the convergence curves of the proposed method on the synthetic data with four Gaussian noise levels, i.e., 10, 20, 30, and 40 dB. From Fig. 3, as the number of iterations increases, all curves are monotonically decreasing

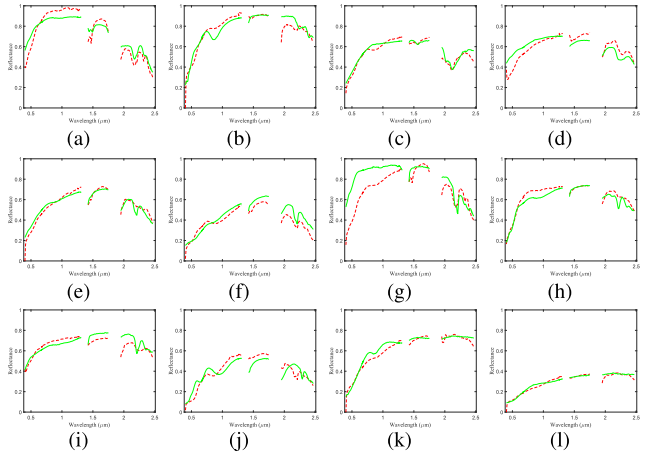

Fig. 4. Comparison of the USGS library spectra (green solid line) with the endmember signatures estimated by CANMF-TV (red dashed line) on the AVIRIS Cuprite data set. (a) #1. (b) #2. (c) #3. (d) #4. (e) #5. (f) #6. (g) #7. (h) #8. (i) #9. (j) #10. (k) #11. (l) #12.

and approach convergence. In addition, we also investigate the average running time of ten times on synthetic data with 30 dB for CANMF-TV, uDAS, $L_{1/2}$-RNMF, and $L_{1/2}$-NMF, which are 29.12, 6.05, 22.94, and 2.05 s, respectively. Although more running time is required, our method is implemented still in second as other compared methods, which is acceptable for small-scale data. Besides, for the much larger data, the graphic processing unit (GPU) parallel computing technique can be adopted to shorten the computational time significantly.

B. Real Data Experiments

Here, the proposed CANMF-TV method is applied to unmix a real hyperspectral image. Thus, experiments are conducted on the AVIRIS Cuprite data set. It consists of 250 × 191 pixels and remains 188 bands after removing several low SNR bands. This data set is extensively used in hyperspectral research and 12 minerals are generally considered, including "Alunite GDS82 Na82 (#1)," "Andradite WS487 (#2)," "Buddingtonite GDS85 D-206 (#3)," "Chalcedony CU91-6A (#4)," "Kaolin/Smect H89-FR-5 30K (#5)," "Kaolin/Smect KLF508 85%K (#6)," "Kaolinite KGa-2 (#7)," "Montmorillonite + Illi CM37 (#8)," "Muscovite IL107 (#9)," "Nontronite NG-1.a (#10)," "Pyrope WS474 (#11)," and "Sphene HS189.3B (#12)."

Fig. 4 depicts the comparison of the reference signatures collected in USGS library with the estimated ones by CANMF-TV. Obviously, the estimated spectra of each endmember is in accordance with the library one, which confirms that the proposed method is effective to address the unmixing problem. In Fig. 5, we present the abundance maps produced by our proposed method. Furthermore, the SAD results are listed in Table I. From Table I, we can find that all methods achieve satisfactory results for several minerals, such as #2, #8, and #12. Furthermore, our proposed method obtains the minimum value in most cases, especially the average value. Therefore, the proposed CANMF-TV outperforms other methods over comparison.

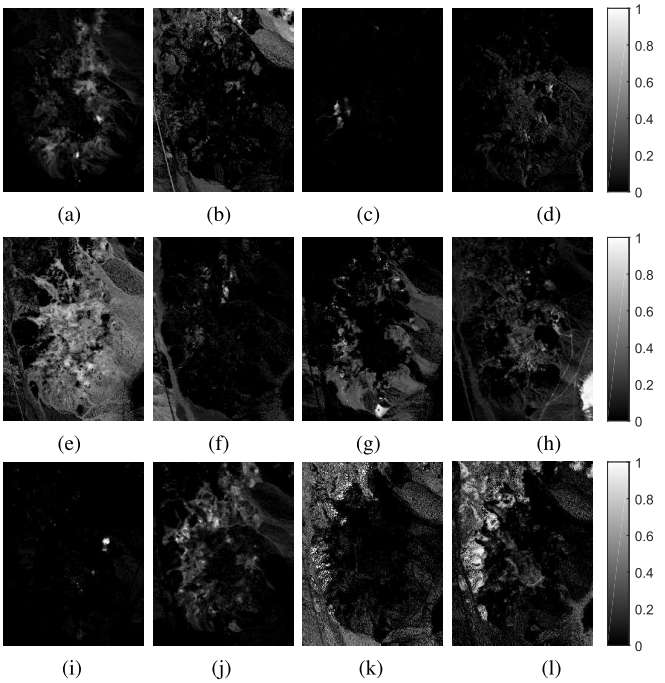

Fig. 5. Fractional abundance maps estimated by CANMF-TV on the AVIRIS Cuprite data set. (a) #1. (b) #2. (c) #3. (d) #4. (e) #5. (f) #6. (g) #7. (h) #8. (i) #9. (j) #10. (k) #11. (l) #12.

TABLE I
PERFORMANCE ABOUT SAD ALONG WITH STANDARD DEVIATION ON THE AVIRIS CUPRITE DATA SET FOR DIFFERENT METHODS. BOLDFACED DENOTES THE BEST RESULTS UNDER EACH CONDITION

	CANMF-TV	uDAS	$L_{1/2}$-RNMF	$L_{1/2}$-NMF
#1	0.1056±2.10%	0.1220±1.27%	**0.0891±2.04%**	0.1340±6.29%
#2	0.0763±0.88%	0.0939±2.93%	0.0832±2.14%	**0.0731±1.66%**
#3	**0.0752±1.80%**	0.1133±0.95%	0.1009±2.48%	0.1032±3.20%
#4	0.1378±2.68%	**0.1349±2.40%**	0.1500±1.59%	0.1504±1.53%
#5	0.0860±3.93%	0.0711±2.47%	**0.0699±1.39%**	0.1004±3.83%
#6	**0.1090±3.99%**	0.1152±4.11%	0.1091±3.34%	0.1163±3.78%
#7	**0.1239±1.84%**	0.1322±1.06%	0.1300±3.62%	0.1520±5.67%
#8	0.0553±1.15%	**0.0502±0.80%**	0.0528±1.07%	0.0583±1.19%
#9	**0.0878±1.43%**	0.1332±2.22%	0.1212±4.11%	0.1045±2.59%
#10	0.1326±1.21%	0.1374±1.49%	0.1292±1.10%	**0.1261±1.34%**
#11	0.0883±3.05%	**0.0768±1.35%**	0.0993±4.47%	0.1134±2.55%
#12	0.0640±0.84%	**0.0595±0.78%**	0.0829±2.00%	0.0680±1.69%
Mean	**0.0951±0.52%**	0.1033±0.35%	0.1015±0.51%	0.1083±0.72%

IV. CONCLUSION

In this letter, a novel CANMF-TV method has been proposed for hyperspectral unmixing. It not only considers the reconstruction of the original data from the abundances in endmember space, but also projects directly the original data into the endmember matrix to obtain the abundance matrix. Meanwhile, the CIM measure is adopted to handle noise effectively. Moreover, considering that more similar abundance values are often presented among the adjacent pixels, TV regularizer is further incorporated to promote the piecewise smoothness of abundances for improving the unmixing performance. Comparing with the existing uDAS, $L_{1/2}$-RNMF, and $L_{1/2}$-NMF, the experimental results validate that the superiority of the proposed CANMF-TV algorithm.

REFERENCES

[1] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 38–42, Jan. 2015.

[2] X. Zhang, L. Zong, X. Liu, and J. Luo, "Constrained clustering with nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1514–1526, Jul. 2016.

[3] T. Zhang, B. Fang, Y. Y. Tang, G. He, and J. Wen, "Topology preserving non-negative matrix factorization for face recognition," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 574–584, Apr. 2008.

[4] 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.

[5] 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.

[6] J. M. P. Nascimento and J. M. B. 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.

[7] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[8] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[9] L. Tong, J. Zhou, Y. Qian, X. Bai, and Y. Gao, "Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6531–6544, Nov. 2016.

[10] 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.

[11] L. Yang, J. Peng, H. Su, L. Xu, Y. Wang, and B. Yu, "Combined nonlocal spatial information and spatial group sparsity in NMF for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, early access, Dec. 4, 2019, doi: 10.1109/LGRS.2019.2954335.

[12] S. Yang, X. Zhang, Y. Yao, S. Cheng, and L. Jiao, "Geometric nonnegative matrix factorization (GNMF) for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2696–2703, Jun. 2015.

[13] W. Wang, Y. Qian, and Y. Y. Tang, "Hypergraph-regularized sparse NMF for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 681–694, Feb. 2016.

[14] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

[15] B.-J. Sun, H. Shen, J. Gao, W. Ouyang, and X. Cheng, "A non-negative symmetric encoder-decoder approach for community detection," in *Proc. CIKM*, 2017, pp. 597–606.

[16] Y. Qu and H. Qi, "uDAS: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1698–1712, Mar. 2019.

[17] L. Zhang, Z. Chen, M. Zheng, and X. He, "Robust non-negative matrix factorization," *Frontiers Elect. Electron. Eng. China*, vol. 6, no. 2, pp. 192–200, Jun. 2011.

[18] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, Sep. 2016.

[19] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[20] L. Du, X. Li, and Y.-D. Shen, "Robust nonnegative matrix factorization via half-quadratic minimization," in *Proc. IEEE 12th Int. Conf. Data Mining*, Dec. 2012, pp. 201–210.

[21] L. Condat, "A generic proximal algorithm for convex optimization—Application to total variation minimization," *IEEE Signal Process. Lett.*, vol. 21, no. 8, pp. 985–989, Aug. 2014.