Non-negative matrix factorization based unmixing for principal component transformed hyperspectral data

Xiu-rui GENG[‡1], Lu-yan JI[2], Kang SUN[3]

(*1Key Laboratory of Technology in Geo-spatial Information Processing and Application System,*
Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China)
(*2MOE Key Laboratory for Earth System Modeling, Center for Earth System Science,*
Tsinghua University, Beijing 100084, China)
(*3The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050000, China*)
E-mail: gengxr@sina.com; jily@mail.ustc.edu.cn; sunkang-1234@163.com
Received Jan. 19, 2016; Revision accepted Mar. 21, 2016; Crosschecked Apr. 25, 2016

Abstract: Non-negative matrix factorization (NMF) has been widely used in mixture analysis for hyperspectral remote sensing. When used for spectral unmixing analysis, however, it has two main shortcomings: (1) since the dimensionality of hyperspectral data is usually very large, NMF tends to suffer from large computational complexity for the popular multiplicative iteration rule; (2) NMF is sensitive to noise (outliers), and thus the corrupted data will make the results of NMF meaningless. Although principal component analysis (PCA) can be used to mitigate these two problems, the transformed data will contain negative numbers, hindering the direct use of the multiplicative iteration rule of NMF. In this paper, we analyze the impact of PCA on NMF, and find that multiplicative NMF can also be applicable to data after principal component transformation. Based on this conclusion, we present a method to perform NMF in the principal component space, named 'principal component NMF' (PCNMF). Experimental results show that PCNMF is both accurate and time-saving.

Key words: Non-negative matrix factorization (NMF), Principal component analysis (PCA), Endmember, Hyperspectral

http://dx.doi.org/10.1631/FITEE.1600028   **CLC number:** TP751.1

1 Introduction

Spectral unmixing analysis has received increasing interest in hyperspectral remote sensing, since the mixture pixels widely exist in hyperspectral imagery. The linear mixing model (LMM) plays an important role in hyperspectral unmixing analysis (Keshava and Mustard, 2002; Plaza *et al.*, 2004; Bioucas-Dias *et al.*, 2012). It assumes that any pixel in the image can be regarded as a linear combination of several pure spectral signatures (called 'endmembers') weighted by corresponding abundance fractions. Owing to the physical constraints, the abundances satisfy the abundance non-negative constraint (ANC) and abundance sum-to-one constraint (ASC). In the assumption of LMM, when all the endmembers are available, the abundances of the endmembers can be obtained conveniently using linear unmixing methods (Heinz and Chang, 2001; Parente and Plaza, 2010; Heylen *et al.*, 2011). Therefore, some methods concentrate on endmember selection, for instance, pixel purity index (PPI) (Boardman, 1992), N-FINDR (Winter, 1999; Ji *et al.*, 2015), orthogonal bases algorithm (OBA) (Tao *et al.*, 2007a), iterative error analysis (IEA) (Neville *et al.*, 1999), simplex growing algorithm (SGA) (Chang *et al.*, 2006), successive projection algorithm (SPA) (Zhang

et al., 2008), maximum volume by householder transformation (MVHT) (Liu and Zhang, 2012), Gaussian elimination method (GEM) (Geng *et al.*, 2013b), and fast Gram determinant based algorithm (FGDA) (Sun *et al.*, 2014). All these algorithms assume that at least one pure pixel exists for each endmember in the image.

However, the pure-pixel assumption is hard to satisfy for real hyperspectral images. In this case, another alternative approach for unmixing analysis, endmember generation, is required. Methods of this type include minimum volume transform (MVT) (Craig, 1994), minimum volume simplex analysis (MVSA) (Li and Bioucas-Dias, 2008), MINVEST (Hendrix *et al.*, 2012), minimum volume enclosing simplex (MVES) (Chan *et al.*, 2009; Ambikapathi *et al.*, 2011), simplex identification via split augmented Lagrangian (SISAL) (Bioucas-Dias, 2009), iterated constrained endmember (ICE) (Berman *et al.*, 2004), and geometric optimization model (GOM) (Geng *et al.*, 2013a). Take ICE as an example. It formulates an optimization problem with an effort to minimize the reconstruction error regularized by a constrained term, i.e., the sum of variances of the simplex vertices. In each iteration of ICE, the abundance fractions of each pixel can be found by solving a quadratic programming problem, which is very time-consuming. In recent years, non-negative matrix factorization (NMF) has been applied to hyperspectral data unmixing (Miao and Qi, 2007; Zymnis *et al.*, 2007; Jia and Qian, 2009; Huck *et al.*, 2010; Liu *et al.*, 2011; Ji *et al.*, 2013; Zhu *et al.*, 2014). The multiplicative update algorithm for NMF was provided by Lee and Seung (1999), which is demonstrated to be computationally simple and does not need any manually set parameters. Note that NMF suffers from two main problems when used for hyperspectral unmixing analysis. One is that the multiplicative iteration version of NMF is very time-consuming if performed directly to the original hyperspectral data since the dimensionality of hyperspectral data is very high (generally more than 100). The other disadvantage of NMF is that it is very sensitive to noise (outliers). To address the two problems, the operation of dimensionality reduction can be conducted, which can not only reduce the data size but also improve the signal-to-noise ratio (SNR) of the data set. However, after dimensionality reduction, the multiplicative learning rule for NMF cannot be used since dimensionality reduced data generally contain negatives. In this paper, we explore the impact of principal component analysis (PCA) on NMF and find that the multiplicative updating rule is still applicable to the data after the dimensionality reducing process, which contains only the step of rotation and does not involve translation. Then, we present a new approach for NMF in the principal component (PC) space, namely PCNMF.

2 Background

In this section, we briefly review the theory of LMM and NMF.

2.1 Linear mixing model

If the assumption of the LMM holds, each spectral vector $\boldsymbol{r}_i$ can be expressed by a linear combination of several endmember vectors $\boldsymbol{e}_j$ $(j = 1, 2, \ldots, p)$ with the physical constrained conditions:

$$\boldsymbol{r}_i = \sum_{j=1}^{p} c_{ij}\boldsymbol{e}_j + \boldsymbol{n} = \boldsymbol{E}\boldsymbol{c}_i + \boldsymbol{n}, \qquad (1)$$

$$\sum_{j=1}^{p} c_{ij} = 1, \ c_{ij} > 0, \qquad (2)$$

where p is the number of endmembers in the image, $\boldsymbol{c}_i = [c_{i1}, c_{i2}, \ldots, c_{ip}]^{\mathrm{T}}$ and c_{ij} is a scalar representing the fractional abundance of endmember vector $\boldsymbol{e}_j$ in the pixel $\boldsymbol{r}_i$. $\boldsymbol{E} = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_p]$ is an $L \times p$ mixing matrix (L is the number of bands for the original data).

2.2 Non-negative matrix factorization

Given an $L \times M$ non-negative matrix $\boldsymbol{R}$ ($M \gg L$ in general) and a positive integer $p < L$, the task of non-negative matrix factorization is to find two non-negative matrices $\boldsymbol{E}_{L \times p}$ and $\boldsymbol{C}_{p \times M}$ such that

$$\boldsymbol{R} \approx \boldsymbol{E}\boldsymbol{C}. \qquad (3)$$

A natural way to solve the NMF problem for $\boldsymbol{E}$ and $\boldsymbol{C}$ is to construct the following optimization problem:

$$\min f(\boldsymbol{E}, \boldsymbol{C}) = \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{E}\boldsymbol{C}\|_{\mathrm{F}}^2$$
$$\text{subject to } \boldsymbol{E} \geq 0, \boldsymbol{C} \geq 0, \qquad (4)$$

where $\| \cdot \|_{\mathrm{F}}$ represents the Frobenius norm. There are many methods that can be used to solve Eq. (4),

among which the most popular one is the multiplicative iteration rule. The learning process for the multiplicative NMF is

$$C = C . * \left(E^{\mathrm{T}} R \right) ./ \left(E^{\mathrm{T}} E C \right),$$ (5)

$$E = E . * \left(R C^{\mathrm{T}} \right) ./ \left(E C C^{\mathrm{T}} \right).$$ (6)

If the initial matrices E and C are strictly non-negative, these matrices can remain non-negative throughout the iterations. If the ASC needs to be considered, we can just replace the matrices R and E by $\bar{R} = \begin{bmatrix} R \\ \delta \mathbf{1}_M^{\mathrm{T}} \end{bmatrix}$, $\bar{E} = \begin{bmatrix} E \\ \delta \mathbf{1}_p^{\mathrm{T}} \end{bmatrix}$, where $\mathbf{1}_M$ is an M-dimensional column vector, and $\mathbf{1}_p$ a p-dimensional column vector, with all elements equal to one, and δ is a positive parameter to control the effect of ASC, which is usually assigned manually. In recent years, NMF has been widely used in un-mixing analysis for hyperspectral data (Miao and Qi, 2007; Zymnis *et al.*, 2007; Jia and Qian, 2009; Huck *et al.*, 2010; Liu *et al.*, 2011; Ji *et al.*, 2013; Zhu *et al.*, 2014). Most of these methods perform NMF directly to the original data, so they are generally time-consuming and sensitive to noise (outliers). Although dimensionality reduction can mitigate these problems well, it leads to another intractable situation; i.e., the non-negative property cannot be satisfied for the dimensionality reduced data. As a result, the multiplicative updating rule cannot be applied in the dimensionality reduced space. Interestingly, we find that the multiplicative updating rule is still applicative for dimensionality reduction where only rotation operation is contained. It will be elaborated in the following section.

3 Impact of PCA on NMF

In this section, taking PCA (Jolliffe, 2002) as an example, we investigate the applicability of NMF in the PC space. As is well known, the standard PCA process contains two main steps, translation and rotation. More specifically, translation means moving the center of the data to the mean vector, while rotation means projecting the mean-removed data to the directions of the eigenvectors derived from its covariance matrix, which is equivalent to rotating the data by the orthogonal matrix (eigenvector matrix). Hence, in the following, we discuss the impact of general translation and rotation on NMF.

3.1 Impact of rotation

In general, the original hyperspectral data are non-negative, i.e., $R \geq 0$. However, the data may contain negatives after the rotation by orthogonal matrix V. Denote $\widehat{R} = V^{\mathrm{T}} R$ and $\widehat{E} = V^{\mathrm{T}} E$ as the rotated data and endmember matrix, respectively. First, we consider the multiplicative learning rule for C, which can be rewritten as

$$\widehat{C} = \widehat{C} . * \left(\widehat{E}^{\mathrm{T}} \widehat{R} \right) ./ \left(\widehat{E}^{\mathrm{T}} \widehat{E} \widehat{C} \right).$$ (7)

Substituting $\widehat{R} = V^{\mathrm{T}} R$ and $\widehat{E} = V^{\mathrm{T}} E$ into Eq. (7), we have

$$\begin{aligned} \widehat{C} &= \widehat{C} . * \left((V^{\mathrm{T}} E)^{\mathrm{T}} (V^{\mathrm{T}} R) \right) ./ \left((V^{\mathrm{T}} E)^{\mathrm{T}} V^{\mathrm{T}} E \widehat{C} \right) \\ &= \widehat{C} . * \left(E^{\mathrm{T}} V V^{\mathrm{T}} R \right) ./ (E^{\mathrm{T}} V V^{\mathrm{T}} E \widehat{C}) \\ &= \widehat{C} . * (E^{\mathrm{T}} R) ./ (E^{\mathrm{T}} E \widehat{C}). \end{aligned}$$ (8)

Clearly, Eq. (8) is absolutely equivalent to Eq. (5). That is to say, the rotation does not change the multiplicative updating rule for the abundance matrix. This is a very interesting conclusion, which implies that the abundance matrix can be obtained directly in the PC space though the rotated data may contain negatives. Considering that the information of hyperspectral data is contained mainly in the top several components, we can multiplicatively update the abundance matrix of hyperspectral data in the space spanned by the top several principal components (PCs). This will greatly reduce the computational complexity. Moreover, the accuracy of endmember and abundance can be improved by performing NMF in the PC space, since PCA has a capacity of suppressing noise and enhancing SNR. Next, let us discuss the computation of endmember matrix E in the PC space. The multiplicative updating rule for E is

$$\widehat{E} = \widehat{E} . * (\widehat{R} \widehat{C}^{\mathrm{T}}) ./ (\widehat{E} \widehat{C} \widehat{C}^{\mathrm{T}}).$$ (9)

Similarly, plugging $\widehat{C} = C$ and $\widehat{R} = V^{\mathrm{T}} R$ into Eq. (9), we have

$$\begin{aligned} \widehat{E} &= \widehat{E} . * (\widehat{R} C^{\mathrm{T}}) ./ (\widehat{E} C C^{\mathrm{T}}) \\ &= \widehat{E} . * (V^{\mathrm{T}} R C^{\mathrm{T}}) ./ (\widehat{E} C C^{\mathrm{T}}). \end{aligned}$$ (10)

Unfortunately, the effect of V cannot be eliminated in the multiplicative updating rule for the endmember matrix. Both $\widehat{R}$ and $\widehat{E}$ in Eq. (10) may contain

negatives in the PC space. Therefore, this is the main obstacle to prevent us from applying NMF in the PC space.

3.2 Impact of translation

From the above analysis, it can be found that the rotation will not change the update rule of $\boldsymbol{C}$, but the update rule of $\boldsymbol{E}$ may not hold, since $\widehat{\boldsymbol{R}}$ in Eq. (10) may contain negatives. To make the multiplicative update rule applicable to the rotated data, a simple trick to make all data non-negative is by translation. Here, supposing the data and endmembers contain negatives, we can select a vector $\boldsymbol{r}_0$ (L-dimensional column vector), such that $(\boldsymbol{R} - \boldsymbol{r}_0\mathbf{1}_M^{\mathrm{T}}) \geq 0$ and $(\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}}) \geq 0$. Then the corresponding multiplicative updating rules become

$$\boldsymbol{C} = \boldsymbol{C}.* \left((\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{R} - \boldsymbol{r}_0\mathbf{1}_M^{\mathrm{T}}) \right)$$
$$./ \left((\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}})\boldsymbol{C} \right), \quad (11)$$

$$\boldsymbol{E} = (\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}}).* \left((\boldsymbol{R} - \boldsymbol{r}_0\mathbf{1}_M^{\mathrm{T}})\boldsymbol{C}^{\mathrm{T}} \right)$$
$$./ \left((\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}})\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} \right) + \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}}. \quad (12)$$

Then the objective function of NMF in Eq. (4) becomes

$$f(\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}}, \boldsymbol{C}) = \|(\boldsymbol{R} - \boldsymbol{r}_0\mathbf{1}_M^{\mathrm{T}}) - (\boldsymbol{E} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}})\boldsymbol{C}\|_{\mathrm{F}}^2$$
$$= f(\boldsymbol{E}, \boldsymbol{C}) + 2\mathrm{tr}\left((\boldsymbol{R} - \boldsymbol{E}\boldsymbol{C})(\boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}}\boldsymbol{C} - \boldsymbol{r}_0\mathbf{1}_M^{\mathrm{T}})^{\mathrm{T}} \right)$$
$$+ \|\boldsymbol{r}_0\mathbf{1}_M^{\mathrm{T}} - \boldsymbol{r}_0\mathbf{1}_p^{\mathrm{T}}\boldsymbol{C}\|_{\mathrm{F}}^2, \quad (13)$$

where $\mathrm{tr}(\cdot)$ is the trace operator. We can see from Eq. (13) that, when the abundance matrix $\boldsymbol{C}$ satisfies ASC, i.e., $\mathbf{1}_M^{\mathrm{T}} = \mathbf{1}_p^{\mathrm{T}}\boldsymbol{C}$, the translation does not change the value of the objective function. However, the ASC of $\boldsymbol{C}$ cannot be guaranteed in the multiplicative learning process (see Eqs. (5), (6), (11), and (12)). Therefore, the translation of the data will change the final solution of NMF. Moreover, the number of such $\boldsymbol{r}_0$ which can achieve the non-negativity of $\boldsymbol{R}$ and $\boldsymbol{E}$ is infinite, and different $\boldsymbol{r}_0$ will lead to different $\boldsymbol{E}$ and $\boldsymbol{C}$ even with the same initialization. Therefore, although the way of translation can make $\boldsymbol{R} \geq 0, \boldsymbol{E} \geq 0$, it is unnatural and its use in real applications is not suggested.

4 Principal component non-negative matrix factorization

In Section 3, we learn that both steps of the PCA process could cause the failure of applying mul-

tiplicative update rules to NMF in the PC space. For the rotation step, since the effect of rotation matrix $\boldsymbol{V}$ can be completely eliminated, as Eq. (8), the update rule for the abundance matrix remains the same. Yet, the rotation matrix $\boldsymbol{V}$ cannot be eliminated in the update rule of $\boldsymbol{E}$, as Eq. (10), so the update rule of $\boldsymbol{E}$ is no longer universally applicable. Although data translation can solve the problem caused by $\boldsymbol{V}$, the analysis in Section 3.2 indicates that the translation will change the final solution, and different $\boldsymbol{r}_0$ will lead to different results.

Interestingly, based on our observation, the maximum spectral angle between pixels in real hyperspectral data is mostly small (generally less than 45°). In addition, the rotation operation will not change the spectral angles between data points. Both facts motivate us to apply the orthogonal procrustes (OP) technique to solve the non-negative problem of $\boldsymbol{R}$ and $\boldsymbol{E}$ in the PC space. That is, forcibly rotating all the data points into the first quadrant of the PC space will make the update rule still work for $\boldsymbol{E}$.

The OP problem was first proposed by Green (1952), and the general solution to this problem was given by Schönemann (1966). The problem is defined as the least-squares problem of transforming a given matrix $\boldsymbol{B}$ to a given matrix $\boldsymbol{A}$ by an orthogonal transformation matrix $\boldsymbol{Q}$, which maps $\boldsymbol{B}$ to $\boldsymbol{A}$ as closely as possible. Mathematically, this problem can be stated as follows:

$$\min_{\boldsymbol{Q}} f(\boldsymbol{Q}) = \|\boldsymbol{A} - \boldsymbol{B}\boldsymbol{Q}\|_{\mathrm{F}}$$
$$\text{subject to} \quad \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{I}. \quad (14)$$

According to Schönemann (1966), the optimal solution of Eq. (14) is

$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{W}^{\mathrm{T}}, \quad (15)$$

where matrices $\boldsymbol{U}$ and $\boldsymbol{W}$ satisfy the singular value decomposition (SVD) equation $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{W}^{\mathrm{T}}$.

Here, a 2D example is shown in Fig. 1 to illustrate the OP process. The original data are distributed in the second quadrant. Let $\boldsymbol{A} = [1, 1]$, and $\boldsymbol{B}$ be the mean vector of the data set. Then we can calculate $\boldsymbol{Q}$ by Eq. (15). It can be seen that all data are transformed to the first quadrant, being non-negative. As mentioned before, to achieve the non-negativity requirement for both $\boldsymbol{R}$ and $\boldsymbol{E}$, the data should satisfy one premise; that is, the maximum spectral angle of the data set should be no
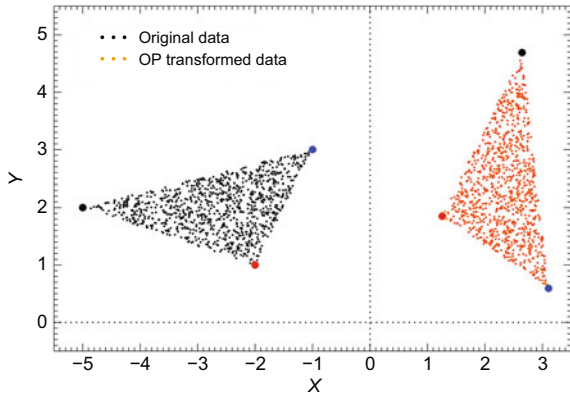
Fig. 1 Illustration of the orthogonal procrustes (OP) process in the 2D space

larger than 90°. Since digital number, radiance, or reflectance values are all non-negative, real hyperspectral data all exist in the first quadrant. As a result, the maximum spectral angle of the real data set is no larger than 90°. Moreover, we compute a series of real hyperspectral data sets, and find that the maximum spectral angle is usually less than 45° in practice. Therefore, by OP transformation, we can perform the multiplicative update rules in the PC space without involving the translation process. Clearly, the PC transformation without translation and the OP transformation can be combined into one process, named 'PC-OP transformation' here. Overall, PC-OP transformation has the following advantages: (1) reduce the dimensionality of the data set and thus the computational complexity, (2) make all data non-negative, so that multiplicative update rules of NMF can still work in the PC space, and (3) improve the SNR of the data set by choosing only the first $p - 1$ PC bands, to improve the final accuracy.

After obtaining the non-negative data and endmember matrices, we can directly perform NMF on them for the final endmember extraction, which is summarized in Algorithm 1. Some practical considerations need to be stated here:

1. Initialization: Due to the non-convexity of the objective function (4), the random initialization on E and C will definitely lead to the local minima problem, which makes the solution meaningless. Taking a 2D situation as an example, the scatter points in Fig. 2 are the highly mixed data cloud. Points A, B, and C are three real endmembers, and obviously ΔABC is the most compacted triangle that encloses all these points. Because of the non-convexity of NMF, any triangle in the plane

---

Algorithm 1 Principal component non-negative matrix factorization

Input: data matrix, $R_{L \times M}$; number of endmembers, p; maximum iteration number, maxiter; ASC weight, δ.

Output: endmember matrix, $E_{L \times p}$; abundance matrix, $C_{p \times M}$.

// PC-OP transformation

1: Calculate the autocorrelation matrix, $\Sigma_{L \times L} = RR^{\mathrm{T}}/M$

2: Calculate the eigenvector matrix for Σ which satisfies $\Sigma = VDV^{\mathrm{T}}$, with the PC rotation matrix being $V = [V]_{:,1:(p-1)}$

3: Calculate the mean vector of the PC rotated data, $\bar{r} = V^{\mathrm{T}}\mathrm{mean}(R)$

4: Let $A_{(p-1) \times 1} = [1, 1, \cdots, 1]^{\mathrm{T}}$, $B = \bar{r}$, and calculate the OP matrix Q based on Eq. (15)

5: Calculate the PC-OP transformed data, $\widetilde{R} = (VQ)^{\mathrm{T}}R$

// NMF initialization

6: $\mathrm{index}_0 = \mathrm{FGDA}(\widetilde{R}, p)$, $\widetilde{E}_0 = \widetilde{R}_{:,\mathrm{index}_0}$

7: $\bar{R} = \begin{bmatrix} \widetilde{R} \\ \delta\mathbf{1}_M^{\mathrm{T}} \end{bmatrix}$, $\bar{E}_0 = \begin{bmatrix} \widetilde{E}_0 \\ \delta\mathbf{1}_p^{\mathrm{T}} \end{bmatrix}$

8: $C_0 = \mathrm{lsqnonneg}(\bar{R}, \bar{E}_0)$

// NMF based endmember extraction

9: $[E, C] = \mathrm{NMF}(\bar{R}, \bar{E}_0, C_0, \mathrm{maxiter})$

---

that can enclose the data set will be a local solution of NMF, such as $\Delta A_1 B_1 C_1$ and $\Delta A_2 B_2 C_2$. However, the vertices A_1, B_1, C_1 or A_2, B_2, C_2 are far different from the true endmembers A, B, and C, so they have no physical meanings.

To avoid unreasonable local minima, Tao *et al.* (2007b) used the N-FINDR outputs as the endmember initialization. We can employ the same strategy. We use the results of FGDA (Sun *et al.*, 2014), which are independent of the dimensionality of data, as the initialization for E. For C, it is estimated by solving the non-negative least-squares constraint problem (lsqnonneg) on the augmented data and endmember matrices, $\bar{R}$ and $\bar{E}_0$.

2. Stopping conditions: There are two widely used criteria, maximum iteration number and minimum error tolerance. Here, we employ the maximum iteration number.

5 Experiments

We have conducted tests based on both simulated and real data to evaluate the performance of PCNMF. Both endmember accuracy and
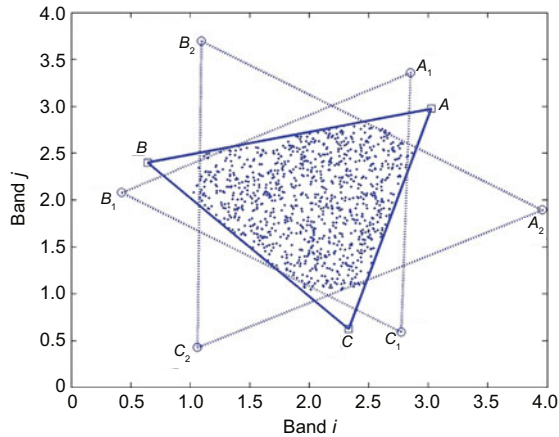
Fig. 2 Illustration of the orthogonal procrustes (OP) process in the 2D space

computation time have been evaluated. For fair comparison, we set the same initializations of E and C, maximum iteration number maxiter, and ASC weight δ for both NMF and PCNMF in all experiments.

5.1 Simulated data

Here, we design two experiments to evaluate the performance of FGDA, NMF, and PCNMF. The spectra of three minerals (Alunite, Calcite, and Kaolinite) from the U.S. Geological Survey (USGS) Digital Spectral Library are selected as endmember signatures. Then 2000 mixture vectors are generated according to Eqs. (1) and (2) with abundance fractions following a Dirichlet distribution. To ensure that no pure pixel exists, fractions are not allowed to be larger than 0.9. The selection of the maximum iteration number, maxiter, may be different for different data sets. For the simulated data sets used in this study, the NMF algorithms can generally produce acceptable results when the iteration number is around 4000, so we set maxiter to 4000. The ASC weight is set to $\delta = 13$ for all experiments. In addition, since the data are randomly generated, the average result of 10 runs is presented in the following. To evaluate the performance of these methods, the metrics of rmsSAD, rmsSID (Nascimento and Bioucas-Dias, 2005), and the relative abundance error (RAE) (Geng *et al.*, 2015) are used, where SAD stands for the spectral angle distance, and SID stands for the spectral information divergence. The abundance matrix of FGDA is estimated by a least-squares method with the ASC weight also set to 13.

5.1.1 Experiment 1: accuracy test for noiseless data

Fig. 3 shows that NMF-based methods NMF and PCNMF obtain exactly the same performance in terms of accuracy. Compared to FGDA (rmsSAD about 0.81°), the two NMF-based methods can obtain more accurate endmembers (rmsSAD about 0.49°), which can be attributed to the fact that there are no 'pure' pixels in the data cloud.
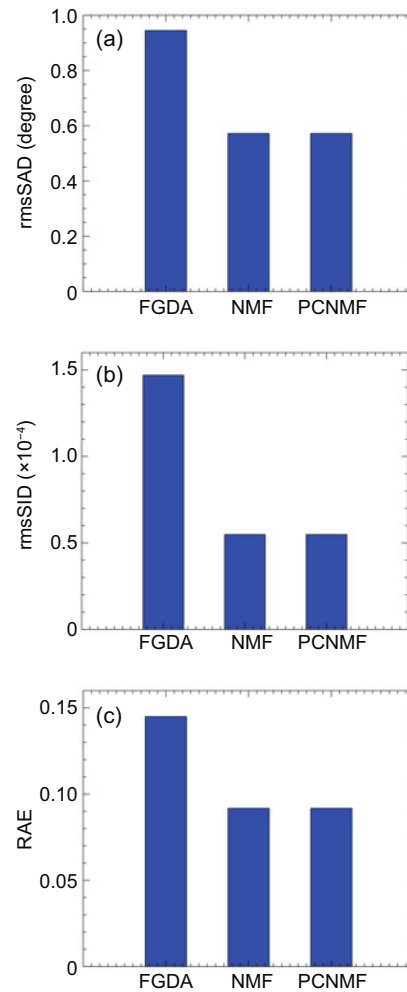

Fig. 3 The rmsSAD (a), rmsSID (b), and relative abundance error (RAE) (c) for different methods with noise-free data

5.1.2 Experiment 2: accuracy test for noisy data

Real data always contain noise. Here we investigate the performances of these methods for noisy data. The Gaussian noise at different SNR levels (SNR = 10, 15, 20, 25, and 30 dB) is added to the simulated data. The accuracy results are shown in

Fig. 4.

As can be seen from Fig. 4, our proposed method PCNMF is always better than FGDA and NMF with the original dimensional data. When SNR≥15 dB, FGDA has the worst performance. When SNR= 10 dB, NMF has the lowest accuracy in all metrics, which indicates that NMF is sensitive to noise. The superiority of PCNMF can be attributed to the fact that PCA contributes to the improvement of SNR and suppression of noise. NMF in the PC space is more accurate, particularly when SNR is low.
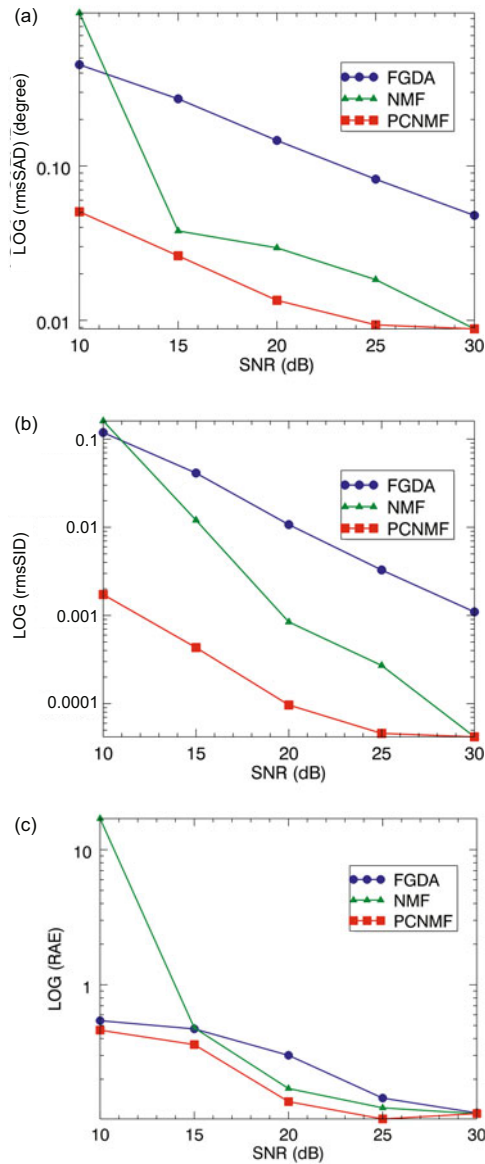

Fig. 4 The rmsSAD (a), rmsSID (b), and relative abundance error (RAE) (c) as functions of SNR for different methods with noisy data

5.1.3 Experiment 3: computation time

We have also recorded the computation time for these methods (Table 1). As shown, NMF is very time-consuming. Benefiting from the computation of E and C in the low-dimensional PC space, PCNMF costs much less time than NMF.

Table 1 Computation time for different methods

Method	Time (s)	Method	Time (s)
FGDA	0.0093	PCNMF	0.7324
NMF	12.5288		

5.2 Real data

In this subsection, we evaluate the performance of PCNMF using the well-known AVIRIS Cuprite data set (Green *et al.*, 1998). AVIRIS is a high-quality, low-noise hyperspectral instrument that acquires data in 224 contiguous spectral bands ranging from 365 to 2500 nm. The selected subscence is shown in Fig. 5 with a size of 191×250 pixels. Due to water absorption or low SNR, bands 1–3, 104–113, 148–167, and 221–224 are removed. The maximum SADs of the original, PC transformed, and OP transformed data are listed in Table 2. The maximum SAD of the original data is around 23°, greatly less than 90°, and the PC and PC-OP transformations have little effect on SADs between pixels.

According to the related references (Chan *et al.*, 2009) and the ground-truth (Swayze *et al.*, 1992), we set the number of endmembers as $p = 14$. For fair comparison, FGDA is also conducted on the data after dimensionality reduction by Algorithm 1. Also, the 14 endmembers extracted by FGDA are used as the initial endmembers for PCNMF. For the maximum iteration number and ASC weight, the same values used in the simulations are applied, i.e., maxiter=4000, δ=13. The resampled spectra from the USGS Digital Spectral Library are selected as the ground-truth for comparison. For each mineral, the library spectrum, which has both small SAD and small SID with the endmember extracted from FGDA, is selected as the reference spectrum. The comparison of FGDA and PCNMF in terms of SAD is shown in Table 3. It can be seen that 10 out of 14 endmembers extracted by PCNMF outperform those by FGDA and it has smaller average SAD compared to FGDA.
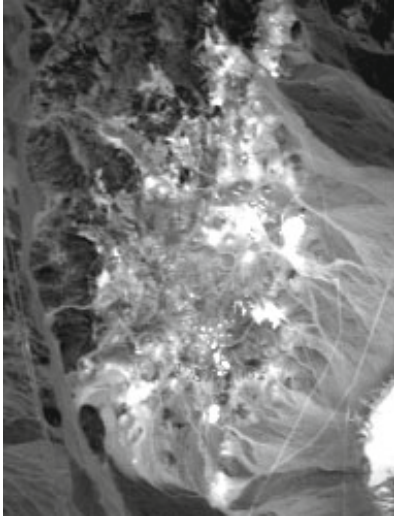
Fig. 5 Band 8 (λ=0.463 μm) of the real hyperspectral scene

Table 2 The maximum SAD of the original, PC transformed, and OP transformed data

Data	Maximum SAD (degree)
Original	23.7468
PC transformed	28.9216
OP transformed	28.9216

Table 3 The SAD between USGS reference spectra and extracted endmembers by FGDA and PCNMF

No.	Substance	SAD (degree)	
		FGDA	PCNMF
1	Muscovite IL107	**5.3472**	5.9231
2	Desert Vanish GDS141	9.4029	**8.1041**
3	Alunite GDS84 Na03	4.6300	**3.9203**
4	Kaolin/Smect KLF508	**5.2682**	5.8254
5	Montmorillonite SWy-1	6.9298	**6.6978**
6	Kaolinite CM7	**4.8221**	4.8635
7	Buddingtonite NHB2301	5.4664	**5.2410**
8	Alunite GDS82 Na82	12.4593	**11.3928**
9	Montmorillonite+Illi CM42	6.2363	**5.6179**
10	Chalcedony CU91-6A	5.5653	**5.1812**
11	Alunite AL706	7.7288	**7.6880**
12	Montmorillonite+Illi CM37	4.8154	**4.7617**
13	Kaolin/Smect KLF511	**3.8877**	4.3025
14	Kaolin/Smect H89-FR-5	4.2887	**5.2605**
	Average	6.2034	**6.0557**

Bold text indicates the better SAD

Fig. 6 shows the abundance images of six main endmembers from PCNMF, which present high level of similarity to the published geologic maps.
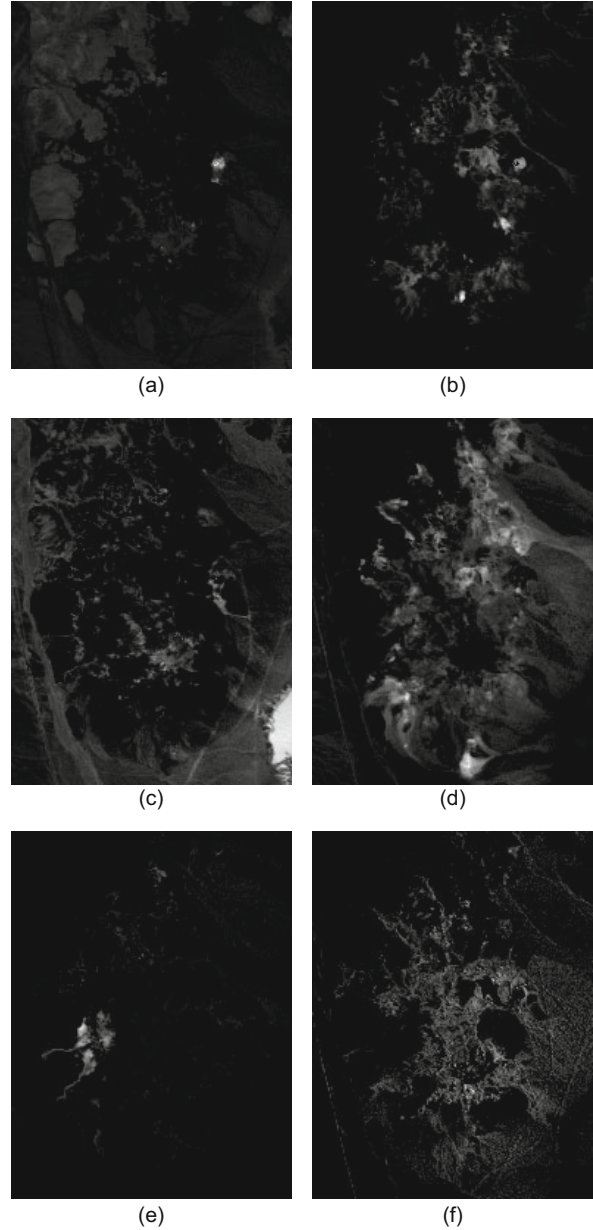

Fig. 6 Abundance maps of different endmembers from PCNMF: (a) E1, Muscovite; (b) E3, Alunite1; (c) E5, Montmorillonite; (d) E6, Kaolinite; (e) E7, Buddingtonite; (f) E10, Chalcedony

6 Conclusions

In this paper, we have explored the possibility of applying the PCA dimensionality reduction technique to the multiplicative update rules of NMF. The main obstacle is that data after PC transformation may contain negatives, which could be caused by both steps of PCA (i.e., translation and rotation). We have proved that the rotation matrix of PCA can be eliminated using the multiplicative learning

rule for $\boldsymbol{C}$, but not for $\boldsymbol{E}$. To solve the non-negative problem of PCA data, one possible way is to add a large positive vector, $\boldsymbol{r}_0$, to all data. However, different $\boldsymbol{r}_0$ will lead to different $\boldsymbol{E}$ and $\boldsymbol{C}$, so translation is not an advisable way for practical applications. According to our observation, the maximum SAD of real hyperspectral data is not large (generally less than $45°$). Since rotation operation will not change the SADs between data vectors, we thereby introduce the OP transformation to forcibly rotate the data cloud into the first quadrant in the PC space. This NMF-based unmixing analysis method for data after PC rotation and OP transformations is named PCNMF. Compared to the original NMF, PCNMF is more robust to noise. Moreover, since the data dimensionality is reduced to $p - 1$, our method can greatly save computational time.

References

Ambikapathi, A., Chan, T.H., Ma, W.K., et al., 2011. Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens., **49**(11):4194-4209. http://dx.doi.org/10.1109/TGRS.2011.2151197

Berman, M., Kiiveri, H., Lagerstrom, R., et al., 2004. ICE: a statistical approach to identifying endmembers in hyperspectral images. IEEE Trans. Geosci. Remote Sens., **42**(10):2085-2095. http://dx.doi.org/10.1109/TGRS.2004.835299

Bioucas-Dias, J.M., 2009. A variable splitting augmented Lagrangian approach to linear spectral unmixing. 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, p.1-4. http://dx.doi.org/10.1109/WHISPERS.2009.5289072

Bioucas-Dias, J.M., Plaza, A., Dobigeon, N., et al., 2012. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., **5**(2):354-379. http://dx.doi.org/10.1109/JSTARS.2012.2194696

Boardman, J.W., 1992. Automated spectral unmixing of AVIRIS data using convex geometry concepts. Summaries of the 4th Annual JPL Airborne Geoscience Workshop, p.11-14.

Chan, T.H., Chi, C.Y., Huang, Y.M., et al., 2009. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. IEEE Trans. Signal Process., **57**(11):4418-4432. http://dx.doi.org/10.1109/TSP.2009.2025802

Chang, C.I., Wu, C.C., Liu, M., et al., 2006. A new growing method for simplex-based endmember extraction algorithm. IEEE Trans. Geosci. Remote Sens., **44**(10):2804-2819. http://dx.doi.org/10.1109/TGRS.2006.881803

Craig, M.D., 1994. Minimum-volume transforms for remotely sensed data. IEEE Trans. Geosci. Remote Sens., **32**(3):542-552. http://dx.doi.org/10.1109/36.297973

Geng, X.R., Ji, L.Y., Zhao, Y.C., et al., 2013a. A new endmember generation algorithm based on a geometric optimization model for hyperspectral images. IEEE Geosci. Remote Sens. Lett., **10**(4):811-815. http://dx.doi.org/10.1109/LGRS.2012.2224635

Geng, X.R., Xiao, Z.Q., Ji, L.Y., et al., 2013b. A Gaussian elimination based fast endmember extraction algorithm for hyperspectral imagery. ISPRS J. Photogr. Remote Sens., **79**:211-218. http://dx.doi.org/10.1016/j.isprsjprs.2013.02.020

Geng, X.R., Sun, K., Ji, L.Y., et al., 2015. Optimizing the endmembers using volume invariant constrained model. IEEE Trans. Image Process., **24**(11):3441-3449. http://dx.doi.org/10.1109/TIP.2015.2446196

Green, B.F., 1952. The orthogonal approximation of an oblique structure in factor analysis. Psychometrika, **17**(4):429-440. http://dx.doi.org/10.1007/BF02288918

Green, R.O., Eastwood, M.L., Sarture, C.M., et al., 1998. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). Remote Sens. Environ., **65**(3):227-248. http://dx.doi.org/10.1016/S0034-4257(98)00064-9

Heinz, D.C., Chang, C.I., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Trans. Geosci. Remote Sens., **39**(3):529-545. http://dx.doi.org/10.1109/36.911111

Hendrix, E.M.T., Garcia, I., Plaza, J., et al., 2012. A new minimum-volume enclosing algorithm for endmember identification and abundance estimation in hyperspectral data. IEEE Trans. Geosci. Remote Sens., **50**(7):2744-2757. http://dx.doi.org/10.1109/TGRS.2011.2174443

Heylen, R., Burazerovic, D., Scheunders, P., 2011. Fully constrained least squares spectral unmixing by simplex projection. IEEE Trans. Geosci. Remote Sens., **49**(11):4112-4122. http://dx.doi.org/10.1109/TGRS.2011.2155070

Huck, A., Guillaume, M., Blanc-Talon, J., 2010. Minimum dispersion constrained nonnegative matrix factorization to unmix hyperspectral data. IEEE Trans. Geosci. Remote Sens., **48**(6):2590-2602. http://dx.doi.org/10.1109/TGRS.2009.2038483

Ji, L.Y., Geng, X.R., Yu, K., et al., 2013. A new non-negative matrix factorization method based on barycentric coordinates for endmember extraction in hyperspectral remote sensing. Int. J. Remote Sens., **34**(19):6577-6586. http://dx.doi.org/10.1080/01431161.2013.804223

Ji, L.Y., Geng, X.R., Sun, K., et al., 2015. Modified N-FINDR endmember extraction algorithm for remote-sensing imagery. Int. J. Remote Sens., **36**(8):2148-2162. http://dx.doi.org/10.1080/01431161.2015.1034895

Jia, S., Qian, Y.T., 2009. Constrained nonnegative matrix factorization for hyperspectral unmixing. IEEE Trans. Geosci. Remote Sens., **47**(1):161-173. http://dx.doi.org/10.1109/TGRS.2008.2002882

Jolliffe, I.T., 2002. Principal Component Analysis. Springer.

Keshava, N., Mustard, J.F., 2002. Spectral unmixing. IEEE Signal Process. Mag., **19**(1):44-57. http://dx.doi.org/10.1109/79.974727

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**(6755):788-791. http://dx.doi.org/10.1038/44565

Li, J., Bioucas-Dias, J.M., 2008. Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data. IEEE Int. Geoscience and Remote Sensing Symp., p.250-253. http://dx.doi.org/10.1109/IGARSS.2008.4779330

Liu, J.M., Zhang, J.S., 2012. A new maximum simplex volume method based on householder transformation for endmember extraction. *IEEE Trans. Geosci. Remote Sens.*, **50**(1):104-118. http://dx.doi.org/10.1109/TGRS.2011.2158829

Liu, X.S., Xia, W., Wang, B., *et al.*, 2011. An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, **49**(2):757-772. http://dx.doi.org/10.1109/TGRS.2010.2068053

Miao, L.D., Qi, H.R., 2007. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.*, **45**(3):765-777. http://dx.doi.org/10.1109/TGRS.2006.888466

Nascimento, J.M.P., Bioucas-Dias, J.M., 2005. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, **43**(4):898-910. http://dx.doi.org/10.1109/TGRS.2005.844293

Neville, R.A., Staenz, K., Szeredi, T., *et al.*, 1999. Automatic endmember extraction from hyperspectral data for mineral exploration. Canadian Symp. on Remote Sensing, p.21-24.

Parente, M., Plaza, A., 2010. Survey of geometric and statistical unmixing algorithms for hyperspectral images. 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, p.1-4. http://dx.doi.org/10.1109/WHISPERS.2010.5594929

Plaza, A., Martinez, P., Perez, R., *et al.*, 2004. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, **42**(3):650-663. http://dx.doi.org/10.1109/TGRS.2003.820314

Schönemann, P.H., 1966. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, **31**(1):1-10. http://dx.doi.org/10.1007/BF02289451

Sun, K., Geng, X.R., Wang, P.S., *et al.*, 2014. A fast endmember extraction algorithm based on Gram determinant. *IEEE Geosci. Remote Sens. Lett.*, **11**(6):1124-1128. http://dx.doi.org/10.1109/LGRS.2013.2288093

Swayze, G., Clark, R.N., Kruse, F., *et al.*, 1992. Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada. Summaries of the 3rd Annual JPL Airborne Geoscience Workshop, p.47-49.

Tao, X.T., Wang, B., Zhang, L.M., *et al.*, 2007a. A new endmember extraction algorithm based on orthogonal bases of subspace formed by endmembers. IEEE Int. Geoscience and Remote Sensing Symp., p.2006-2009. http://dx.doi.org/10.1109/IGARSS.2007.4423223

Tao, X.T., Wang, B., Zhang, L.M., *et al.*, 2007b. A new scheme for decomposition of mixed pixels based on nonnegative matrix factorization. IEEE Int. Geoscience and Remote Sensing Symp., p.1759-1762. http://dx.doi.org/10.1109/IGARSS.2007.4423160

Winter, M.E., 1999. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. *SPIE*, **3753**:266-275.

Zhang, J.K., Rivard, B., Rogge, D.M., 2008. The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data. *Sensors*, **8**(2):1321-1342. http://dx.doi.org/10.3390/s8021321

Zhu, F.Y., Wang, Y., Xiang, S.M., *et al.*, 2014. Structured sparse method for hyperspectral unmixing. *ISPRS J. Photogr. Remote Sens.*, **88**:101-118.

Zymnis, A., Kim, S.J., Skaf, J., *et al.*, 2007. Hyperspectral image unmixing via alternating projected subgradients. 41st Asilomar Conf. on Signals, Systems and Computers, p.1164-1168. http://dx.doi.org/10.1109/ACSSC.2007.4487406

Dr. Xiu-rui GENG, first author of this invited research article, received his bachelor and master degrees in applied mathematics from Beihang University in 1999 and 2002, respectively, and the Ph.D. degree in hyperspectral remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2005. Currently, he is an Associate Professor in the Key Laboratory of Technology in Geo-spatial Information Process and Application Systems, Institute of Electronics, Chinese Academy of Sciences. His research interests are hyperspectral data understanding and algorithm development in feature extraction, mixed pixel analysis, image matching, target detection and classification by introducing new algebra, geometry, analysis, and statistics viewpoints.