Empirical Automatic Estimation of the Number of Endmembers in Hyperspectral Images

Bin Luo, Jocelyn Chanussot, *Fellow, IEEE*, Sylvain Douté, and Liangpei Zhang, *Senior Member, IEEE*

***Abstract*—In this letter, an eigenvalue-based empirical method is proposed in order to estimate the number of endmembers in hyperspectral data. This method is based on the distribution of the differences of the eigenvalues from the correlation and the covariance matrices, respectively. The eigenvalues corresponding to the noise are identical in the covariance and the correlation matrices, while the eigenvalues corresponding to the signal (the endmembers) are larger in the correlation matrix than in the covariance matrix. The proposed method is totally parameter free and very fast. It is validated by experiments carried on both synthetic and real data sets.**

***Index Terms*—Imaging, spectral analysis.**

I. Introduction

VISIBLE and near-infrared imaging spectroscopy, also called hyperspectral imagery, is a key tool for Earth remote sensing and planetology [1]. As solar light is partially transmitted, reflected, and diffused back by interaction with different constituents of the atmosphere and the surface, the analysis of reflectance spectra may allow the identification and quantification of the chemical species. For this purpose, we consider that each measured hyperspectral spectrum is a linear mixture of the spectra of different chemical species (thereafter called *endmembers*) [2]. Based on this linear mixture model, methods such as the Bayesian positive source separation [3], vertex component analysis (VCA) [4], n-Finder [5], minimum-volume enclosing simplex [6], simplex identification via split augmented Lagrangian [7], minimum-volume constrained nonnegative matrix factorization [8], and region-based method [9] are developed in order to extract the endmembers and estimate the corresponding abundances. This operation is called unmixing. The number of endmembers is usually a crucial parameter. If the estimated number is too small, the spectra of the extracted endmembers could be still mixtures of several chemical species. While if it is too large, the extracted endmembers could be strongly affected by noise.

In order to estimate the number of endmembers, principal-component-analysis-based thresholding is a standard method [3]. However, the cutoff threshold should be manually chosen, which is very difficult to determine since the eigenvalues corresponding to signals and noise are sometimes very similar. This problem is related to the estimation of the number of sources in mixed signals. However, the classical methods for this purpose, such as minimum description length (MDL) [10] and Akaike information criterion (AIC) [11], cannot well estimate the number of endmembers in hyperspectral data because the corresponding noise is not independent and identically distributed, as assumed by MDL and AIC [12]. Therefore, other approaches have been proposed in the literature, which can mainly be divided into two families. One is based on the eigenvalues of the covariance or correlation matrices of the data, such as the Harsanyi–Farrand–Chang (HFC) method [12]. This approach requires fixing the false-alarm probability α, which could affect the estimated number of endmembers. The other is based on the minimization of error when the data are projected into a subspace, such as the Hyperspectral Signal Identification by Minimum Error (HySIME) method [13]. This approach requires the estimation of the noise of each spectral band in order to compute the noise covariance matrix. The estimation of noise is very computationally demanding.

In this letter, we propose a new empirical method—eigenvalue likelihood maximization (ELM)—based on the likelihood function of the distribution of the eigenvalue differences for determining the number of endmembers, which does not need any parameter and does not require the estimation of the noise.

This letter is organized as follows. In Section II, the method proposed for estimating the number of endmembers is presented. The experimental study includes Section III, with various synthetic data sets, and Section IV, with real hyperspectral data sets.

II. ELM

In hyperspectral imagery, the measured spectra are usually considered as linear mixtures of the spectra of constituting chemical species (*endmembers*) with different abundances.

We note $\mathbf{X}$ as the matrix representing the hyperspectral image cube, where $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_a}\}$, where $\mathbf{x}_k = \{x_{1,k}, x_{2,k}, \ldots, x^{\mathrm{T}}_{N_s,k}\}$ with $x_{l,k}$ as the value of the kth pixel at the lth spectral band. We assume that the spectrum of each pixel is a linear mixture of the spectra of N_c endmembers, leading to the following model:

$$\mathbf{X} = \mathbf{MS} + \mathbf{n} \tag{1}$$

Manuscript received August 10, 2011; revised December 5, 2011; accepted December 28, 2011. This work was supported in part by the Agence Nationale de la Recherche (ANR) through Project VAHINE and Project XIMRI/NSFC41061130553, by the Chinese 973 Program under Grant 2011CB707100, and by the National Natural Science Foundation of China under Project 61102129.

B. Luo and L. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 430079, Wuhan 430072, China (e-mail: robinlb2002@gmail.com; zlp62@public.wh.hb.cn).

J. Chanussot is with the GIPSA-Lab, Grenoble Institute of Technology, 38402 Grenoble, France (e-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

S. Douté is with the Institut de Planétologie et d'Astrophysique de Grenoble, Centre National de la Recherche Scientifique, 38041 Grenoble, France (e-mail: sylvain.doute@obs.ujf-grenoble.fr).

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

Digital Object Identifier 10.1109/LGRS.2012.2189934

where $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{N_c}\}$ is the mixing matrix, where $\mathbf{m}_n$ denotes the spectral signature of the nth endmember. $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{N_c}\}^{\mathrm{T}}$ is the abundance matrix, where $\mathbf{s}_n = \{s_{n,1}, s_{n,2}, \ldots, s_{n,N_a}\}$ ($s_{n,k} \in [0,1]$ is the abundance of the nth endmember at the kth pixel). $\mathbf{n}$ stands for the additive noise of the image.

We note K as the sample covariance matrix of $\mathbf{X}$ and R as its correlation matrix. Suppose that λ_i and $\hat{\lambda}_i$ are the ith eigenvalues of K and R, respectively, with $i \geq 0$, $\lambda_i > \lambda_{i+1}$, and $\hat{\lambda}_i > \hat{\lambda}_{i+1}$. Theoretically, if there are N_c endmembers present in $\mathbf{X}$, the eigenvalues $\hat{\lambda}_i$ $(i > N_c)$ and $\lambda_i (i > N_c)$ correspond to the noise variance; therefore, we have

$$\begin{cases} \hat{\lambda}_i - \lambda_i > 0, & i \leq N_c \\ \hat{\lambda}_i - \lambda_i = 0, & i > N_c. \end{cases} \tag{2}$$

Noting $z_i = \hat{\lambda}_i - \lambda_i$, a Neyman–Person test can be used to threshold the z_i value in order to estimate the number of endmembers [12]. However, one has to fix the false-alarm value α in order to determine the threshold for z_i, which can affect the estimated number of endmembers.

In this letter, we propose to use the distribution of z_i values for estimating the number of endmembers without any parameter, which can give a precise number of endmembers even if the *SNR* is very low. The distribution of z_i can be modeled by [14]

$$\begin{aligned} z_i &\sim \aleph\left(\mu_i, \sigma_i^2\right), & i \leq N_c \\ z_i &\sim \aleph(0, \sigma_i^2), & i > N_c \end{aligned} \tag{3}$$

where the mean μ_i is unknown and the standard deviation σ_i can be given by $\sigma_i^2 \approx (2/N)(\hat{\lambda}_i^2 + \lambda_i^2)$, if the number of samples is sufficiently large (which is usually the case for hyperspectral images) [12], [14].

According to (3), we define a joint likelihood function

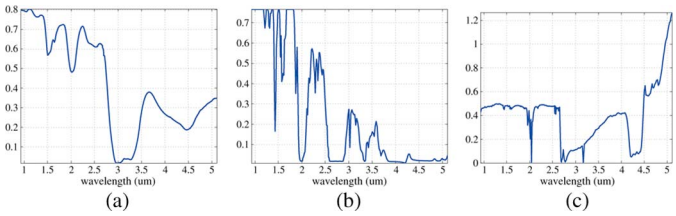
$$H(i) = \Pi_{l=i}^{N_s} \frac{1}{\sigma_l} \exp\left(-\frac{z_l^2}{2\sigma_l^2}\right) \tag{4}$$

and take its logarithmic value

$$\begin{aligned} \tilde{H}(i) &= \log H(i) \\ &= -\sum_{l=i}^{N_s} \frac{z_l^2}{2\sigma_l^2} - \sum_{l=i}^{N_s} \log \sigma_l. \end{aligned} \tag{5}$$

We note $A(i) = -\sum_{l=i}^{N_s}(z_l^2/2\sigma_l^2)$ and $B(i) = -\sum_{l=i}^{N_s} \log \sigma_l$. Generally, if the signal abundance is large enough (which is usually the case in hyperspectral imagery), μ_i should not be close to zero when $i < N_c$. Therefore, z_l $(l \leq N_c)$ should also be large, which yields $\exp(-(z_l^2/2\sigma_l^2))$ to be close to zero. When $l > N_c$, z_l is very close to zero, and $\exp(-(z_l^2/2\sigma_l^2))$ will be very large when compared to the case where $l \leq N_c$. Therefore, the likelihood function $H(N_c + 1) > H(i)$ when $i \leq N_c$. In practice, we have also found that, when $i \geq N_c + 1$, $A(i)$ will change very slightly with i. Moreover, if we normalize all the values of $\mathbf{X}$ into [0, 1], $-\log \sigma_i$ will be positive, and $B(i)$ decreases as a function of i when $i \geq N_c + 1$, as well as $\tilde{H}(i)$. There is a global *maximum* of $\tilde{H}(i)$ at $i = N_c + 1$. Therefore, the number of endmembers can be estimated by

$$\hat{N}_c = \arg\max_i \left\{\tilde{H}(i)\right\} - 1. \tag{6}$$



(a) (b) (c)

Fig. 1. Reference spectra of the three endmembers. (a) H_2O. (b) CO_2. (c) Dust.

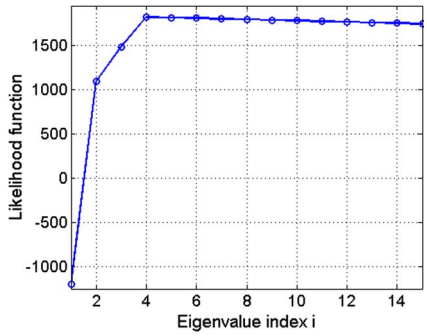


Fig. 2. Likelihood function computed on synthetic data (see text).

III. Experiments on Synthetic Data

A. Synthetic Data Without Artifact

In this section, we carry out experiments on synthetic data simulated by using the spectra of three typical endmembers (H_2O, CO_2, and dust) on the south pole of planet Mars provided by the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG), which are shown in Fig. 1. Each spectrum has 256 spectral bands. The reflectance spectra of these endmembers are calculated with radiative transfer models and laboratory data under the atmospheric condition of the planet Mars.

The abundances of these endmembers are simulated by Dirichlet distributions, and the maximal abundance of each endmember $\sup \mathbf{S}$ is set from 0.4 to 1 with an increment of 0.1. The dimension of the abundance of each endmember is 96×96 pixels. Gaussian zero-mean noise is added, with *SNR* varying from 10 to 50 dB. Therefore, the dimension of one synthetic hyperspectral image is 9216×256.

In Fig. 2, the likelihood function $\tilde{H}(i)$(6) computed on synthetic data when $SNR = 10$ dB and the maximal abundance values of endmembers are 0.4. It can be seen that, when $i = 4$, there is a global maximum. The estimated number of endmembers is hence $N_c = 3$(6).

Our experiment shows that all the three methods—ELM, HySIME, and HFC—always give the correct number of endmembers.

For the HFC method, three false-alarm values are used ($\alpha = 10^{-5}, 10^{-4}$, and 10^{-3}), always leading to the same result. In this experiment, all the three methods (ELM, HySIME, and HFC) correctly estimate the number of endmembers. The estimation results by using HFC with $\alpha = 10^{-4}$ and $\alpha = 10^{-5}$ are the same as the results when $\alpha = 10^{-3}$.

In this case, the ELM method is as robust as the state-of-the-art approaches (HFC and HySIME) even when the noise level is very high ($SNR = 10$ dB) and the maximal abundances of endmembers are very low (0.4). However, the average computation times of ELM, HySIME, and HFC for estimating the

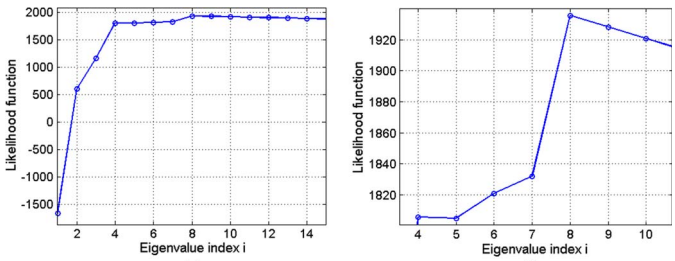


Fig. 3. (a) Likelihood function computed on synthetic data with artifacts.

number of endmembers on one synthetic hyperspectral image are 0.169, 2.488, and 0.313 s, respectively (Matlab, 2.4-GHz CPU, 4-GB RAM). The computational cost of the ELM method is considerably lower than those of the other methods, particularly when compared to HySIME which needs to estimate the noise of each spectral band.

B. Synthetic Data With Artifacts

In practice, hyperspectral data are usually severely corrupted by artifacts. Although the artifacts can be considered as noise, they are not Gaussian and nonzero mean.

In this section, we replace the 10th, 20th, 30th, and 40th spectral bands of the synthetic data by four images with a straight line (with width of five lines) at different positions in order to simulate the artifact, particularly the stripe artifact. Since four bands are corrupted by the artifacts, we note $N_{\text{art}} = 4$.

In Fig. 3(a), the likelihood function $\tilde{H}(i)$ when $SNR = 10$ dB and the maximal abundances of the endmembers are 0.7. In order to show the maxima of the likelihood function, we show the zoomed region of $\tilde{H}(i)$ in Fig. 3(b). There are one local maximum at $i = 4$ and a global maximum at $i = 8$. The first maximum corresponds to the real number of endmembers $N_c = 3$, while the global maximum corresponds to the number of endmembers plus the number of the spectral bands which are corrupted by artifacts, i.e., $N_c + N_{\text{art}} = 7$.

Therefore, in practice, if there are several maxima on the function $\tilde{H}(i)$, the number of endmembers N_c is defined by the first local maximum, i.e.,

$$\hat{N}_c = \min_i \left\{ \tilde{H}(i-1) \leq \tilde{H}(i) \quad \text{and} \quad \tilde{H}(i+1) \leq \tilde{H}(i) \right\} - 1. \tag{7}$$

In Table I, the estimated numbers of endmembers by using ELM, HySIME, and HFC with different levels of noise and different maximal abundance values of endmembers are shown. For the results of ELM method, the numbers corresponding to the first local maximum [see (7)] and the global maximum [in parentheses; see (6)] are both shown. It can be seen that, when the maximal abundances of the endmembers are larger than 0.4, the estimated numbers by (7) are always equal to the real number of endmembers, and the global maxima nearly always correspond to $N_c + N_{\text{art}} = 7$; except when the maximal abundance is 0.5 and the *SNR* is too low ($SNR = 10$ dB), there is only one global maximum which corresponds to the real number of endmembers. When the maximal abundances of the endmembers are too low (0.4), the ELM method does not estimate the correct number of endmembers.

TABLE I
NUMBERS OF ENDMEMBERS ESTIMATED BY ELM, HFC, AND HYSIME ON SYNTHETIC DATA WITH ARTIFACTS

Numbers of endmembers estimated by ELM The estimated numbers corresponding to the first local maximum and the global maximum (in parenthesis) are shown.					
SNR Maximal abundance	10dB	20dB	30dB	40dB	50dB
1.0	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)
0.9	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)
0.8	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)
0.7	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)
0.6	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)
0.5	**3**	3, (**7**)	3, (**7**)	3, (**7**)	3, (**7**)
0.4	2, (**7**)	2, (**7**)	2, (**7**)	2, (**7**)	2, (**7**)
Numbers of endmembers estimated by HySIME)					
SNR Maximal abundance	10dB	20dB	30dB	40dB	50dB
0.4-1.0	3	3	3	3	3
Numbers of endmembers estimated by HFC when $\alpha = 10^{-3}$, 10^{-4} and 10^{-5}, respectively					
SNR Maximal abundance	10dB	20dB	30dB	40dB	50dB
1.0	6, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6
0.9	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6
0.8	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6
0.7	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6
0.6	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6
0.5	5, 4, 4	7, 6, 6	7, 6, 6	7, 6, 6	7, 6, 6
0.4	3, 3, 3	5, 5, 5	4, 4, 4	5, 5, 4	5, 4, 4

HySIME is as robust as in the previous experiment. The estimated number is always equal to three.

For the HFC method, we have used the same three different false-alarm values. HFC provides $N_c + N_{\text{art}} = 7$ only when $\alpha = 10^{-3}$. For other values, the estimated number by HFC is between N_c and $N_c + N_{\text{art}}$. The artifacts have a very strong influence on the HFC method and will further affect the unmixing procedure.

IV. EXPERIMENTS ON REAL DATA

In this section, the proposed ELM algorithm is validated on two real hyperspectral data: the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite data and the data acquired by the Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité (OMEGA) instrument on the south pole of planet Mars.

A. Cuprite Image

We at first validate our method on the Cuprite data set acquired by the AVIRIS sensor on June 19, 1997. A subset of 200×200 pixels with 50 spectral bands from 2.00 to 2.48 μm are taken into consideration [see Fig. 4(a)]. We compute the likelihood function defined by (6), of which the result is shown in Fig. 4(b). It can be seen that, when $i = 2$, there is a local maximum. According to (7), this local maximum leads to only one endmember, which is not reasonable. Therefore, the number of endmembers is determined by the global maximum, which is at $i = 8$. According to (6), the number of endmembers estimated by ELM is $N_c = 7$. For comparison, we have used HySIME and HFC (with false-alarm values $\alpha = 10^{-3}$, 10^{-4}, and 10^{-5}). The results are shown in Table II. The numbers of endmembers estimated by the three methods are very close (seven by ELM, eight by HySIME, and nine by HFC). According to these results, the performance of the ELM method is quite close to those of

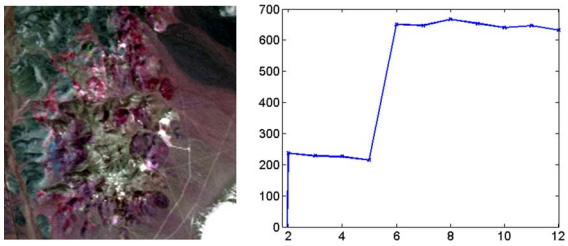


Fig. 4. (a) Cuprite data (R—2.10 μm; G—2.20 μm; B—2.34 μm). (b) Likelihood function when $i \in [4, 11]$. The global maximum is at $i = 8$.

TABLE II
NUMBERS OF ENDMEMBERS ESTIMATED BY THE ELM, HYSIME, AND HFC METHODS ON AVIRIS CUPRITE IMAGE

Method	ELM	HySIME	HFC (10^{-5})	HFC (10^{-4})	HFC (10^{-3})
Estimated number	7	8	9	9	9

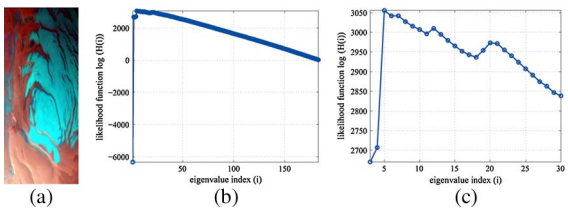


Fig. 5. OMEGA ORB0041 image. (R—band 198; G—band 55; B—band 14). (a) Likelihood function $\tilde{H}(i)$ calculated on the image of Mars. (b) Zoom for $i \in [3, 30]$. A global *maximum* is reached at $i = 5$.

the state-of-the-art methods. The unmixing results by extracting seven endmembers from this data set can be found in [15].

B. OMEGA Image

The OMEGA ORB0041 image [16] [see Fig. 5(a)] taken in late summer covers a large part of the south polar permanent cap and layered deposits of Mars. We reduce the spatial extent of the image to this region of interest (of which the spatial dimension is 128 $\times$ 300 pixels). Previous studies [17] detected three principal chemical species on the surface: water ice, CO_2 ice, and mineral dust (of which the spectra are shown in Fig. 1). The spatial resolution of the image is approximately 1 km/pixel. After classically removing the bands corresponding to dead spectels and saturated atmospheric absorption, 183 spectral bands remain in the visible and infrared range. In [3], the authors applied to the same OMEGA observation a blind source separation consisting in an independent component analysis, followed by a Bayesian scheme that implements positivity constraints, where the number of endmembers is manually fixed to three.

In Fig. 5(b) and (c), the likelihood $\tilde{H}(i)$ calculated on the image is shown. The number of endmembers estimated by ELM is $\hat{N}_c = 4$. This number is quite close to the expected number (which is three) of endmembers existing in this region.

In Table III, the estimated numbers of endmembers by using ELM, HySIME, and HFC on the OMEGA data are shown. The numbers estimated by both HySIME and HFC methods are too high than the real number. In addition, the computation times of ELM, HySIME, and HFC are 0.359, 6.894, and 0.807 s, respectively. The ELM method continues to show its efficiency as in the synthetic experiments.

TABLE III
NUMBERS OF ENDMEMBERS ESTIMATED BY THE ELM, HYSIME, AND HFC METHODS ON THE OMEGA IMAGE

Method	ELM	HySIME	HFC (10^{-5})	HFC (10^{-4})	HFC (10^{-3})
Estimated number	4	10	19	21	22

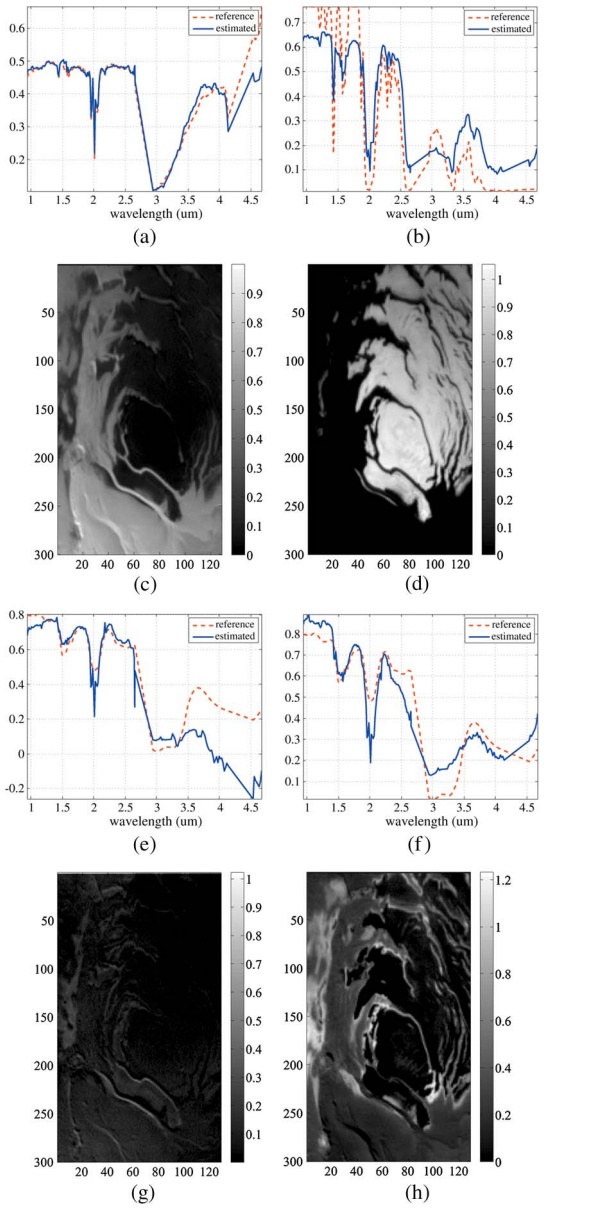


Fig. 6. (a), (b), (e), (f) (Solid line) Four extracted spectra and (dashed line) the most similar references. (c), (d), (g), (h) Abundances of the endmembers computed on the OMEGA ORB0041 image.

The number estimated by the ELM method $(N_c = 4)$ is retained in order to unmix this image. The VCA [4] is used for extracting the spectra of the endmembers, which are shown in Fig. 6(a)–(d). The extracted spectra are compared with the reference spectra of these minerals provided by IPAG by computing the correlation coefficients between them (see Table IV). It can be seen that the spectra of the first, second, and fourth endmembers can easily be identified as dust, CO_2 ice, and H_2O ice since their spectra are strongly correlated.

The abundances of the endmembers are estimated by a non-negative least squares estimator based on the extracted spectra,

TABLE IV
CORRELATION COEFFICIENTS BETWEEN THE REFERENCE SPECTRA AND THE SPECTRA OF THE EXTRACTED ENDMEMBERS

Endmember	H_2O ice	CO_2 ice	Dust
1st	0.835	0.503	**0.959**
2nd	0.833	**0.959**	0.540
3rd	0.922	0.789	0.548
4th	**0.933**	0.896	0.677

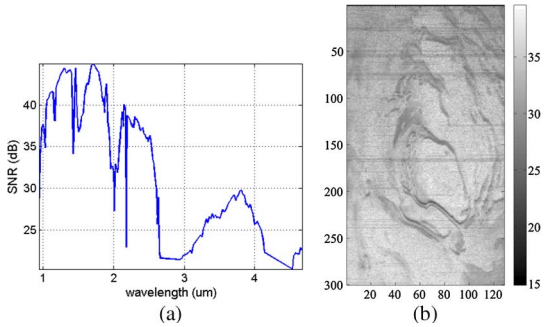


Fig. 7. (a) Spectral SNR of reconstruction. (b) Spatial SNR of reconstruction.

i.e., the abundances of the endmembers $\hat{\mathbf{S}}$ are obtained by the following equation:

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{MS}\|^2, \quad s.t. \quad \mathbf{S} \geq 0 \tag{8}$$

where $\mathbf{M}$ is the spectra of the extracted endmembers. The estimated abundances are shown in Fig. 6(e)–(h).

When compared to the results obtained by the state-of-the-art method in [3], the abundance maps of the CO_2, H_2O, and dust endmembers obtained by linear unmixing are very similar. Although the fourth endmember cannot be identified by its spectrum, its abundance reveals that it corresponds to the line shift of the L and C channels of the OMEGA instrument.

The spectral and the spatial SNR of reconstruction by using the spectra and the abundances of the four extracted endmembers are shown in Fig. 7. The spectral SNR is relatively low in the absorptions of H_2O ice and CO_2 ice (around 1.5, 2.0, and 2.65 μm) because nonlinear effects such as the influence of grain size or of the presence of intimate mixtures are more pronounced in those bands. The radiometric accuracy of the L channel of the instrument is significantly degraded, thus explaining the relatively low SNRs of reconstruction in this spectral range (2.67–4 μm). On the other bands, the spectral SNRs are quite high. The spatial SNRs of reconstruction are quite high (higher than 30 dB), except on the regions where the line shift of L- and C-bands is significant.

It has to be remarked that, in [3], the number of the chemical species in the image is a key parameter, which is manually fixed. On the other hand, in this letter, the number is automatically estimated by the ELM method. The spectra and the abundances of the endmembers are computed in a totally unsupervised parameter-free way.

V. CONCLUSION

An eigenvalue-based method (ELM) has been proposed in order to estimate the number of endmembers in a hyperspectral image. When compared to the state-of-the-art methods, ELM is fast and parameter free and does not need to estimate the noise of each band. Even though the image is corrupted by artifacts, ELM can estimate not only the correct number of endmembers but also the number of bands corrupted by the artifacts. The experiments carried on synthetic data have validated the efficiency of the method, while the experiments carried on real data set acquired by the Observatoire pour la Minlogie, l'Eau, les Glaces et l'Activité (OMEGA) instrument on planet Mars show that the ELM method is the only one which can estimate a reasonable number of endmembers existing in this image. When performing unmixing with the ELM-estimated number of endmembers, physically consistent results are obtained.

REFERENCES

[1] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. S110–S122, Sep. 2009.

[2] A. Plaza, G. Martin, J. Plaza, M. Zortea, and S. Sanchez, "Recent developments in spectral unmixing and endmember extraction," in *Optical Remote Sensing—Advances in Signal Processing and Exploitation Techniques*, S. Prasad, L. Bruce, and J. Chanunussot, Eds. Berlin, Germany: Springer-Verlag, 2011, ch. 12, pp. 235–267.

[3] S. Moussaoui, H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, and J. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation," *Neurocomput., Neurocomput. Vis. Res.; Adv. Blind Signal Process.*, vol. 71, no. 10–12, pp. 2194–2208, Jun. 2008.

[4] J. Nascimento and J. Bioucas 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.

[5] M. E. Winter, "Fast autonomous spectral end-member determination," in *Proc. 13th Int. Conf. Appl. Geol.*, Vancouver, BC, Canada, 1999, vol. 2, pp. 337–344.

[6] T. H. Chan, C. Y. Chi, Y. M. Huang, and W. K. Ma, "A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[7] J. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st IEEE GRSS WHISPERS*, Grenoble, France, 2009, pp. 1–4.

[8] 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.

[9] G. Martin and A. Plaza, "Region-based spatial preprocessing for endmember extraction and spectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 745–749, Jul. 2011.

[10] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[11] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

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

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

[14] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* 2nd ed. New York: Springer-Verlag, 1984.

[15] H. Li and L. Zhang, "A hybrid automatic endmember extraction algorithm based on a local window," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4223–4237, Nov. 2011.

[16] J. P. Bibring, A. Soufflot, M. Berthé, Y. Langevin, B. Gondet, P. Drossart, M. Bouyé, M. Combes, P. Puget, A. Semery, G. Bellucci, V. Formisano, V. Moroz, V. Kottsov, G. Bonello, S. Erard, O. Forni, A. Gendrin, N. Manaud, F. Poulet, G. Poulleau, T. Encrenaz, T. Fouchet, R. Melchiori, F. Altieri, N. Ignatiev, D. Titov, L. Zasova, A. Coradini, F. Capacionni, P. Cerroni, S. Fonti, N. Mangold, P. Pinet, B. Schmitt, C. Sotin, E. Hauber, H. Hoffmann, R. Jaumann, U. Keller, R. Arvidson, J. Mustard, and F. Forget, "OMEGA: Observatoire pour la minéralogie, l'eau, les glaces et l'activit'e," in *Proc. Mars Exp.—The Scientific Payload*, Aug. 2004, pp. 37–49, ESA SP-1240.

[17] J. P. Bibring, Y. Langevin, F. Poulet, A. Gendrin, B. Gondet, M. Berthé, A. Soufflot, P. Drossart, M. Combes, G. Bellucci, V. Moroz, N. Mangold, and B. Schmitt, "Perennial water ice identified in the south polar cap of Mars," *Nature*, vol. 428, no. 6983, pp. 627–630, Apr. 2004.