LOW-RANK AND SPARSE NMF FOR JOINT ENDMEMBERS' NUMBER ESTIMATION AND BLIND UNMIXING OF HYPERSPECTRAL IMAGES

Paris V. Giampouras, Athanasios A. Rontogiannis, Konstantinos D. Koutroumbas

IAASARS, National Observatory of Athens, GR-15236, Penteli, Greece

ABSTRACT

Estimation of the number of endmembers existing in a scene constitutes a critical task in the hyperspectral unmixing process. The accuracy of this estimate plays a crucial role in subsequent unsupervised unmixing steps i.e., the derivation of the spectral signatures of the endmembers (endmembers' extraction) and the estimation of the abundance fractions of the pixels. A common practice amply followed in literature is to treat endmembers' number estimation and unmixing, independently as two separate tasks, providing the outcome of the former as input to the latter. In this paper, we go beyond this computationally demanding strategy. More precisely, we set forth a multiple constrained optimization framework, which encapsulates endmembers' number estimation and unsupervised unmixing in a single task. This is attained by suitably formulating the problem via a low-rank and sparse nonnegative matrix factorization rationale, where low-rankness is promoted with the use of a sophisticated ℓ_1/ℓ_2 norm penalty term. An alternating proximal algorithm is then proposed for minimizing the emerging cost function. The results obtained by simulated and real data experiments verify the effectiveness of the proposed approach.

Index Terms— NMF, sparse and low-rank, number of endmembers, unsupervised unmixing.

1. INTRODUCTION

Big imaging data, such as hyperspectral images (HSIs) and video, convey a sheer bulk of information and this makes them valuable tools in a plethora of applications, with remote sensing being probably the most prominent one [1]. One of the HSI processing tasks that has attracted considerable attention lately is that of hyperspectral unmixing (HU). The main goal of HU is to estimate the spectral signatures of the components (called as *endmembers*) that form the spectral information of HSIs, along with the fractions of their contribution (called as *abundances*) at each pixel.

A first critical and indispensable step towards performing unmixing is to uncover the true number of endmembers that exist in a given hyperspectral scene. This challenging task (also known as *rank estimation* or *model order selection*), can be quite daunting in terms of the required computational burden. Several works have come into play for attacking this problem which can be classified into two main categories: a) information theoretic criteria based approaches and b) eigenvalue thresholding methods. As far as the first class is concerned, various approaches have been proposed differing in the criterion used for penalizing an initially overestimated number of endmembers e.g. Akaike's Information Criterion (AIC), Minimum Description Length (MDL), Bayesian Information Criterion (BIC). On the other hand, methods that belong to the second class include PCA based methods, Neyman-Peyrson detection theory based methods etc., [2].

The estimate of the number of endmembers by the above-mentioned algorithms is provided -at a second phase- as input to unsupervised unmixing algorithms, whose goal is to extract the endmembers' spectral signatures along with the abundance fractions of the pixels. A vast amount of works have been published in the literature dealing with unsupervised hyperspectral unmixing. The majority of them hinges on the assumption that the underlying mechanism that describes the mixing process is linear. Indeed, the so-termed *linear mixing model* (LMM) has been proven to be a reliable approximation, although it neglects non-linear effects met in real situations. In the framework of the LMM, there exist both geometrical, [3], and statistical, [4], matrix factorization based approaches for performing unmixing. Among the latter, nonnegative matrix factorization (NMF) based techniques have exhibited a robust behavior offering promising results.

In this paper, we propose a multiple constrained NMF method for *simultaneously* a) determining the number of endmembers, b) extracting the endmembers' spectral signatures and c) estimating the abundance values of the pixels. To accomplish this, we introduce a sophisticated low-rank promoting term, which is based on the group-sparsity ℓ_1/ℓ_2 inducing norm. This term penalizes both endmembers' and abundance matrices by enforcing joint sparsity on their columns. This way, we go one step beyond just revealing the rank, since we further encourage estimation of the true bases of the column spaces of these matrices. At the same time, sparsity is favored on the abundance matrices, as it is physically meaningful. All in all, endmembers' number estimation and unmixing is yielded jointly by the proposed sparse and low-rank NMF approach. To the best of our knowledge, this is the first work that encapsulates those two problems simultaneously in a sin-

gle task. The newly formulated minimization problem is efficiently tackled via an alternating proximal Newton-type algorithm. Results obtained on simulated and real data demonstrate the favorable properties of the proposed algorithm.

2. PROBLEM FORMULATION

Let us assume that the pixels of the HSI are mixed according to the linear mixing model (LMM),

$$\mathbf{Y} = \boldsymbol{\Phi}\mathcal{W}^T + \mathbf{E} \qquad (1)$$

where matrix $\mathbf{Y} \in \mathcal{R}_+^{L \times K}$ (with $\mathcal{R}_+^{L \times K}$ being the $L \times K$ dimensional nonnegative orthant of $\mathcal{R}^{L \times N}$) consists of the K pixels' $L \times 1$ spectral signatures (with L being the number of spectral bands), $\boldsymbol{\Phi} \in \mathcal{R}_+^{L \times N}$ is the endmembers' matrix containing the spectral signatures of the N endmembers and, $\mathcal{W} \in \mathcal{R}_+^{K \times N}$ is the matrix containing the abundance fractions of the pixels of the HSI. Finally, $\mathbf{E} \in \mathcal{R}^{L \times K}$ is additive i.i.d Gaussian noise, which contaminates the pixels' spectral information.

As mentioned earlier, traditional hyperspectral unmixing algorithms assume the number N of endmembers known beforehand. Typically, this number N is provided by an independent procedure which is applied before the unmixing process. In this work, we depart from this rationale, accounting in our methodology also for the lack of knowledge of the actual number of endmembers existing in a hyperspectral scene. More specifically, we overstate the latent dimension N of the matrix product in (1) and the LMM is subsequently expressed as,

$$\mathbf{Y} = \boldsymbol{\Phi}\mathbf{W}^T + \mathbf{E}, \qquad (2)$$

where $\boldsymbol{\Phi} \in \mathcal{R}_+^{L \times r}$ and $\mathbf{W} \in \mathcal{R}_+^{K \times r}$ are now the endmembers' and abundance matrices respectively, with their relevant dimension overestimated i.e. $r \geq N$.

Non-negative matrix factorization has been widely applied for attacking HU under the LMM, formulating the problem as follows:

$$\min_{\boldsymbol{\Phi} \geq 0, \mathbf{W} \geq 0} \|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{W}^T\|_F^2 \qquad (3)$$

where $\|\cdot\|_F$ denotes the Frobenious norm. Several other constraints may be included in the NMF cost function, giving rise to disparate algorithms. All these constraints aim at capturing inherent structures on the sought matrices such as sparsity, structured sparsity etc, [5], stemming from physical properties e.g. spatial correlation. In the case of the rank overestimated LMM of (2), it is obvious that this approach calls for an appropriate regularization term which should be added in (3). Concretely, such a term should penalize properly the rank of the data representation matrix $\boldsymbol{\Phi}\mathbf{W}^T$.

In light of the above, the ubiquitous tight upper bound of the nuclear norm i.e.,

$$\|\boldsymbol{\Phi}\mathbf{W}^T\|_* = \inf \frac{1}{2} \left(\|\boldsymbol{\Phi}\|_F^2 + \|\mathbf{W}\|_F^2 \right) \qquad (4)$$

has come into play in several works in the literature for imposing low-rankness in bilinear terms of the form $\boldsymbol{\Phi}\mathbf{W}^T$. The minimization of this upper bound though favors low-rankness by inducing smoothness on matrices $\boldsymbol{\Phi}$ and $\mathbf{W}$, i.e., it produces matrices with linear dependent columns/rows. Since HU is an instance of blind source separation, the outcome of the above-mentioned bound minimization lies in the opposite direction of what we actually desire. Ideally, we need an algorithm that returns only distinct endmembers' spectral signatures and the respective abundance fractions' of the pixels.

Considering this and the fact that only a subset of the endmembers in $\boldsymbol{\Phi}$ composes the spectral signature of each pixel (i.e., $\mathbf{W}$ is expected to be sparse), we formulate HU as the following *sparse and low-rank* NMF-type optimization problem,

$$\min_{\boldsymbol{\Phi} \geq 0, \mathbf{W} \geq 0} \|\mathbf{Y} - \boldsymbol{\Phi}\mathbf{W}^T\|_F^2 + \delta \sum_{i=1}^{r} \sqrt{\|\boldsymbol{\phi}_i\|_2^2 + \|\boldsymbol{w}_i\|_2^2}$$
$$+ \lambda_1\|\mathbf{W}\|_1. \qquad (5)$$

The first term in (5) performs the fitting between data matrix $\mathbf{Y}$ and its bilinear representation $\boldsymbol{\Phi}\mathbf{W}^T$, the second term is a novel low-rank promoting penalty and the last term in (5) is the ℓ_1 sparsity inducing norm. Parameters δ and λ_1 are the regularization coefficients of the low-rank and sparsity terms respectively. In a few words, (5) defines the cost function of a multiple constrained optimization problem, which accounts simultaneously for a) non-negativity of the factors $\boldsymbol{\Phi}, \mathbf{W}$, b) low-rank on the product $\boldsymbol{\Phi}\mathbf{W}^T$ and c) sparsity on the abundance matrix $\mathbf{W}$.

Remark 1: The introduced low-rank promoting term in (5) is non-smooth and induces non-separability w.r.t. columns $\boldsymbol{\phi}_i$ and $\boldsymbol{w}_i$ of $\boldsymbol{\Phi}$ and $\mathbf{W}$ respectively. Actually, it is tantamount to applying the column-sparsity promoting ℓ_1/ℓ_2 norm on the augmented matrix $\left[\begin{smallmatrix} \boldsymbol{\Phi} \\ \mathbf{W} \end{smallmatrix} \right]$.

From Remark 1, it can be easily understood that the minimization of the low-rank promoting term of (5) results to zeroing jointly columns of $\boldsymbol{\Phi}$ and $\mathbf{W}$. This way, we can claim that the overestimated dimension r decreases and the remaining non-zero columns of the matrices constitute the bases of the subspaces spanned by $\boldsymbol{\Phi}$ and $\mathbf{W}$.

Next we present a proximal alternating minimization algorithm for solving (5).

3. PROPOSED MINIMIZATION ALGORITHM

The minimization of (5) is by no means a straightforward task. First, exact minimization w.r.t. matrices $\boldsymbol{\Phi}$ and $\mathbf{W}$ is impossible due to the induced non-separability of the low-rank

promoting term (Remark 1). Second, joint non-smoothness and non-separability of this term might lead us to irregular points, i.e. to coordinate-wise minima that may not be local minima of the cost function, [6]. In an effort to circumvent those two impediments a) we use a smooth approximation of the low-rank promoting term by adding a small constant η^2 and b) we replace the smooth terms of (5) with quadratic approximate functions, [6].

Let us first denote the smooth part of (5) as,

$$f(\mathbf{\Phi}, \mathbf{W}) = \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{W}^T\|_F^2 + \delta \sum_{i=1}^{r} \sqrt{\|\phi_i\|_2^2 + \|w_i\|_2^2 + \eta^2}$$

(6)

and with $g(\mathbf{W}) = \lambda_1\|\mathbf{W}\|_1$ its non-smooth separable part. Considering the matrices $\mathbf{\Phi}$ and $\mathbf{W}$ as blocks in our problem, each one of them may be updated by alternatingly minimizing the following approximate cost functions,

$$\mathbf{W}^k = \underset{\mathbf{W} \geq 0}{\operatorname{argmin}} \langle \hat{\mathbf{G}}_{\mathbf{W}}^k, \mathbf{W} - \hat{\mathbf{W}}^{k-1} \rangle$$
$$+ \frac{1}{2\alpha^k} \operatorname{vec}\left(\mathbf{W} - \hat{\mathbf{W}}^{k-1}\right)^T \hat{\mathbf{H}}_{\mathbf{W}}^k \operatorname{vec}\left(\mathbf{W} - \hat{\mathbf{W}}^{k-1}\right)$$
$$+ g(\mathbf{W})$$

(7)

and

$$\mathbf{\Phi}^k = \underset{\mathbf{\Phi} \geq 0}{\operatorname{argmin}} \langle \hat{\mathbf{G}}_{\mathbf{\Phi}}^k, \mathbf{\Phi} - \hat{\mathbf{\Phi}}^{k-1} \rangle$$
$$+ \frac{1}{2\alpha^k} \operatorname{vec}\left(\mathbf{\Phi} - \hat{\mathbf{\Phi}}^{k-1}\right)^T \hat{\mathbf{H}}_{\mathbf{\Phi}}^k \operatorname{vec}\left(\mathbf{\Phi} - \hat{\mathbf{\Phi}}^{k-1}\right)$$

(8)

where $\hat{\mathbf{G}}_{\mathbf{W}}^k = \nabla_{\mathbf{W}} f(\hat{\mathbf{\Phi}}^{k-1}, \hat{\mathbf{W}}^{k-1})$, $\hat{\mathbf{G}}_{\mathbf{\Phi}}^k = \nabla_{\mathbf{\Phi}} f(\hat{\mathbf{\Phi}}^{k-1}, \hat{\mathbf{W}}^k)$, $\hat{\mathbf{H}}_{\mathbf{W}}^k = \hat{\mathbf{\Phi}}^{k-1,T} \hat{\mathbf{\Phi}}^{k-1} + \hat{\mathbf{D}}^{k-1}$, $\hat{\mathbf{H}}_{\mathbf{\Phi}}^k = \hat{\mathbf{W}}^{k,T} \hat{\mathbf{W}}^k + \hat{\mathbf{D}}^{k-1}$ and $\hat{\mathbf{D}}^k$ is a $r \times r$ diagonal matrix with elements

$$\hat{d}_{ii}^k = \frac{\delta}{\sqrt{\|\hat{\phi}_i^k\|_2^2 + \|\hat{w}_i^k\|_2^2 + \eta^2}}.$$

(9)

Note that the small constant η added for smoothing purposes in the introduced low-rank regularization term, averts zeros values on the denominator of $\hat{d}_{ii}^k$. In addition, $\langle \cdot, \cdot \rangle$ denotes inner matrix product, $\operatorname{vec}(\cdot)$ is the vectorization operation and k is the iteration index. Minimization of (7) as such gives rise to a scaled proximal operator, [7], in the form

$$\mathbf{W}^k = \operatorname{prox}_{\|\cdot\|_1}^{\hat{\mathbf{H}}_{\mathbf{W}}^k}\left(\mathbf{W}^{k-1} - (\hat{\mathbf{H}}_{\mathbf{W}}^k)^{-1}\hat{\mathbf{G}}_{\mathbf{W}}^k\right).$$

(10)

The computation of (10) has given way to disparate approximate or iterative schemes (proximal Newton methods), [7], since there exists no closed form solution for non diagonal[1] $\hat{\mathbf{H}}_{\mathbf{W}}^k$ as in our case. Herein, we propose to apply an incremental strategy, [8], for approximately solving (7). More specifically, a gradient step is first applied on the smooth part of

---

[1]Note that for $\hat{\mathbf{H}}_{\mathbf{W}}^k$ diagonal, the scaled proximal operator reduces to the known proximity operator of the ℓ_1 norm i.e. the soft-thresholding operator.

(7). The outcome of the gradient step is then provided as input to the proximal operator of the non-smooth term i.e., the ℓ_1 norm, whose output is finally projected to the feasible set. Regarding the update of $\mathbf{\Phi}$, it is a much simpler task due to the absence of non-smooth terms. In general for the parameter α^k in eqs. (7), (8) it holds $\alpha^k \in (0, 1]$. Overall, $\mathbf{\Phi}$ and $\mathbf{W}$ are computed by using the following expressions,

$$\mathbf{W}^k = \mathcal{P}_{\mathcal{R}_+^{K \times r}}\left(ST\left(\left(\hat{\mathbf{\Phi}}^{k-1,T}\hat{\mathbf{\Phi}}^{k-1} + \hat{\mathbf{D}}^{k-1}\right)^{-1}\hat{\mathbf{\Phi}}^{k-1}\mathbf{Y}, \lambda_1\right)\right)$$

(11)

$$\mathbf{\Phi}^k = \mathcal{P}_{\mathcal{R}_+^{N \times r}}\left(\left(\hat{\mathbf{W}}^{k,T}\hat{\mathbf{W}}^k + \hat{\mathbf{D}}^{k-1}\right)^{-1}\hat{\mathbf{W}}^k\mathbf{Y}\right)$$

(12)

where $ST(x, \lambda) = \operatorname{sign}(x)\max(|x|-\lambda)$ is the soft-thresholding operator and $\mathcal{P}_{\mathcal{R}_+^{K \times r}}$ is the projection operator onto $\mathcal{R}_+^{K \times r}$. Moreover, it can be easily shown that both $\hat{\mathbf{H}}_{\mathbf{W}}^k$ and $\hat{\mathbf{H}}_{\mathbf{\Phi}}^k$ bound the true Hessians i.e., $\nabla_{\mathbf{W}}^2 f(\hat{\mathbf{\Phi}}^{k-1}, \hat{\mathbf{W}}^{k-1})$ and $\nabla_{\mathbf{\Phi}}^2 f(\hat{\mathbf{\Phi}}^{k-1}, \hat{\mathbf{W}}^k)$, respectively, from above. Thus the quadratic approximate functions are upper-bounds of the original cost function. That said, the above-described scheme resembles the block successive upper-bound minimization framework of [6]. As stated earlier, and since we are dealing with a constrained minimization problem, the updates for $\mathbf{\Phi}$ and $\mathbf{W}$ are projected to the feasible set thus accounting for the nonnegativity constraint. In addition, problem (7) is solved inexactly by the incremental strategy described earlier. In view of these, for ensuring that the cost function decreases at each step, an extrapolation step is needed and the final estimates of $\mathbf{\Phi}$ and $\mathbf{W}$ at the kth iteration are obtained as

$$\hat{\mathbf{W}}^k = \hat{\mathbf{W}}^{k-1} + \beta_{\mathbf{W}}^k\left(\mathbf{W}^k - \hat{\mathbf{W}}^{k-1}\right),$$

(13)

$$\hat{\mathbf{\Phi}}^k = \hat{\mathbf{\Phi}}^{k-1} + \beta_{\mathbf{\Phi}}^k\left(\mathbf{\Phi}^k - \hat{\mathbf{\Phi}}^{k-1}\right).$$

(14)

Typically, $\beta_{\mathbf{W}}^k$ and $\beta_{\mathbf{\Phi}}^k$ are adjusted dynamically so that the cost function's sufficient decrease is guaranteed at each step. Along this line, numerous schemes, known as line search methods have come into play e.g. backtracking. Those schemes affect the convergence and rate of convergence of the algorithms to stationary points. The resulting algorithm is given in Algorithm 1.

All in all, the proposed algorithm requires an overestimate of the actual number of endmembers, which relaxes to a large degree the need for knowing in advance the exact number thereof. Then, as the algorithm proceeds, columns of the endmembers' and abundance matrices $\mathbf{\Phi}$ and $\mathbf{W}$ are jointly zeroed as a result of the ℓ_1/ℓ_2 norm based low-rank promoting term adopted. The remaining non-zero columns minimize the reconstruction error while at the same time sparsity is induced on the abundance matrix. Next we experimentally validate the favorable characteristics of the proposed algorithm i.e., its ability to uncover the true number of the endmembers existing in a hyperspectral scene, along with performing highly accurate blind unmixing.

Algorithm 1 The proposed sparse and low-rank NMF endmembers' number estimation and HU algorithm

---

Initialize $\hat{\mathbf{W}}^0, \hat{\mathbf{\Phi}}^0, \hat{\mathbf{D}}^0, \beta_{\mathbf{\Phi}}^0, \beta_{\mathbf{W}}^0$
Select δ, λ_1
for $k = 1, 2, \ldots, \text{Max\_iter}$
$\mathbf{W}^k = \mathcal{P}_{\mathcal{R}_+^{K \times r}} \left(ST \left(\left(\hat{\mathbf{\Phi}}^{k-1,T} \hat{\mathbf{\Phi}}^{k-1} + \hat{\mathbf{D}}^{k-1} \right)^{-1} \hat{\mathbf{\Phi}}^{k-1} \mathbf{Y} \right), \lambda_1 \right)$
$\hat{\mathbf{W}}^k = \hat{\mathbf{W}}^{k-1} + \beta_{\mathbf{W}}^{k-1} \left(\mathbf{W}^k - \hat{\mathbf{W}}^{k-1} \right)$
$\mathbf{\Phi}^k = \mathcal{P}_{\mathcal{R}_+^{L \times r}} \left(\left(\hat{\mathbf{W}}^{k,T} \hat{\mathbf{W}}^k + \hat{\mathbf{D}}^{k-1} \right)^{-1} \hat{\mathbf{W}}^k \mathbf{Y} \right)$
$\hat{\mathbf{\Phi}}^k = \hat{\mathbf{\Phi}}^{k-1} + \beta_{\mathbf{\Phi}}^{k-1} \left(\mathbf{\Phi}^k - \hat{\mathbf{\Phi}}^{k-1} \right)$
$\hat{d}_{ii}^k = \frac{\delta}{\sqrt{\|\hat{\phi}_i^k\|_2^2 + \|\hat{w}_i^k\|_2^2 + \eta^2}}, i = 1, 2, \ldots, r, \hat{\mathbf{D}}^k = \text{diag}(\hat{\mathbf{d}}^k)$
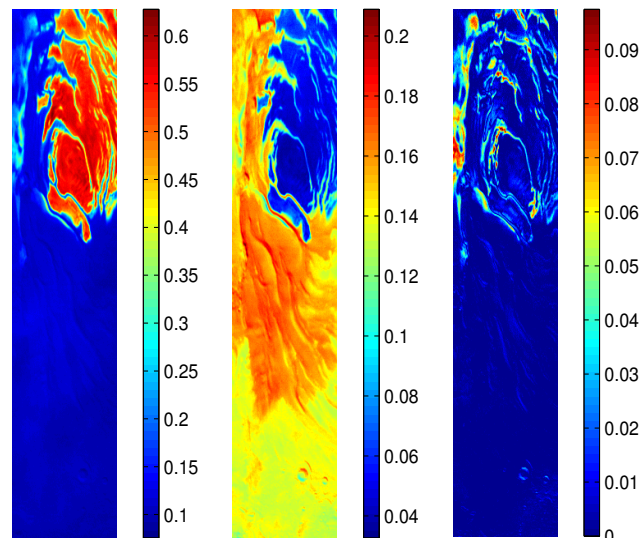Update $\beta_{\mathbf{W}}^k, \beta_{\mathbf{\Phi}}^k$
end
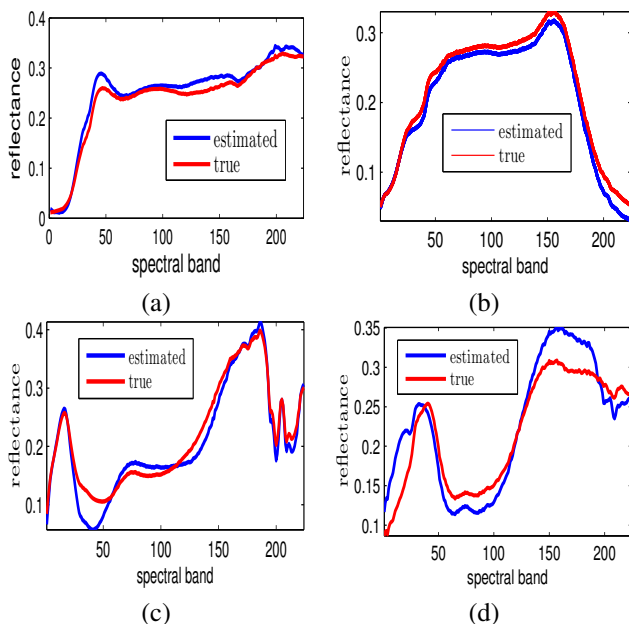
Fig. 1. Endmembers' spectral signatures obtained by the proposed algorithm on the simulated data experiment.

Fig. 2. Abundance maps of South Polar Cap obtained by the proposed algorithm.

4. EXPERIMENTS

In this section we test the performance of the proposed algorithm to a simulated and a real data experiment.

4.1. Simulated data experiment

In this experiment we aim at corroborating the competence of the proposed algorithm in uncovering the true number of endmembers along with estimating the spectral signatures of the endmembers. To this end, we generate a 500×4 abundance matrix whose elements follow a uniform distribution in the interval [0,1]. This matrix is then sparsified by randomly keeping only 30% of its elements. From the USGS spectral library, we select randomly 4 endmembers' spectral signatures

measured at $L = 224$ distinct spectral bands. Then, we linearly produce $K = 500$ simulated pixels' spectral signatures under the LMM framework. The pixel spectral signatures are then contaminated with additive i.i.d. Gaussian noise with standard deviation $\sigma = 10^{-3}$.

Since the main premise of our approach is the development of a blind unmixing method that exhibits robustness in the absence of knowledge of the true number of endmembers, we initialize the proposed algorithm with an overestimate $r = 10$ of the actual number of endmembers. Both endmembers' and abundance matrices are randomly initialized according to the uniform distribution. Interestingly, the proposed algorithm converges to abundance and endmembers' matrices consisting of 4 non-zero columns, which is the same as the actual number of endmembers that produced the data. Moreover, it can be easily observed from Fig. 1, that the estimated endmembers' spectral signatures present high degree of similarity to the real ones. Hence, we can conclude that the proposed algorithm is, in principle, capable of carrying out the challenging task of simultaneously estimating the number of endmembers and performing blind unmixing.

4.2. Real data Experiment

Herein our goal is to test and validate the proposed algorithm on a real hyperspectral dataset. To this end we utilized the OMEGA ORB0041 image (known as South Polar Cap) which covers a large part of the south polar of Mars. Previous studies, e.g. [9], reported three principal chemical species on the surface: H_2O ice, CO_2 ice, and mineral dust.

To evaluate the performance of the proposed algorithm on this real hyperspectral dataset, we suitably initialized the endmembers' matrix with the outcome of the VCA algorithm,
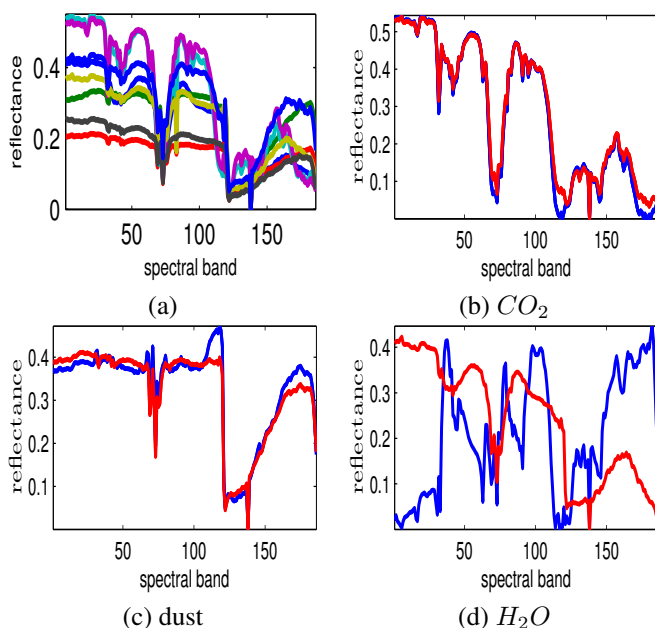
Fig. 3. (a) Eight endmembers' signatures obtained by VCA on South Polar Cap, (b)-(d) endmebers' signatures estimated by the proposed algorithm (blue lines) and VCA (red lines).

[10], which was run with an overestimate $r = 8$ of endmembers. As is shown in Fig. 3(a), the resulting by VCA endmembers' matrix contains correlated spectra, which is expected from our prior knowledge i.e, that there exist 3 (instead of 8) endmembers in this specific image. Since there exist no reference spectra of the actual endmembers' spectral signatures, for comparison purposes, we use the ones returned by VCA which, this time, was run for the correct number of endmembers. Notably, the proposed algorithm is proven to be capable of estimating the actual number of the endmembers. Moreover, the estimated endmembers' signatures (Fig. 3(b)-3(d)) for CO_2 and dust are quite close to those resulting by VCA, while the opposite holds for the respective spectral signature of the H_2O. This is probably due to the small abundance values of H_2O, as shown in Fig. 2(c). Additionally, as it can be observed in Fig. 2, the resulting abundance maps are quite close to those that have been published in literature, [9].

5. CONCLUSIONS

In this paper, the problems of a) estimating the number of endmembers and b) blind unmixing are formulated as a joint optimization task. The ambitious procedure is carried out by first formulating the problem via a novel ℓ_1/ℓ_2 norm based low-rank promoting term. Then, an alternating proximal Newton based algorithm is proposed for minimizing the emerging cost function. Promising results obtained on simulated and real data experiments corroborate the merits of this novel approach.

REFERENCES

[1] W. K. Ma, J. M. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Y. Chi, "A signal processing perspective on hyperspectral unmixing: Insights from remote sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 67–81, Jan 2014.

[2] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[3] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.

[4] X. Lu, H. Wu, and Y. Yuan, "Double constrained NMF for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2746–2758, May 2014.

[5] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4282–4297, Nov 2011.

[6] M. Hong, M. Razaviyayn, Z. Q. Luo, and J. S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 57–77, Jan 2016.

[7] J. D. Lee, Y. Sun, and M. A. Saunders, "Proximal Newton-type methods for minimizing composite functions," *SIAM Journal on Optimization*, vol. 24, no. 3, pp. 1420–1443, 2014.

[8] D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: A survey," *Optimization for Machine Learning*, vol. 2010, no. 1-38, p. 3, 2011.

[9] K. E. Themelis, F. Schmidt, O. Sykioti, A. A. Rontogiannis, K. D. Koutroumbas, and I. A. Daglis, "On the unmixing of MEx/OMEGA hyperspectral data," *Planetary and Space Science*, vol. 68, no. 1, pp. 34–41, 2012.

[10] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.