MULTILINEAR SPECTRAL UNMIXING OF HYPERSPECTRAL MULTIANGLE IMAGES

M. A. Veganzones[a], J. Cohen[a], R. Cabral Farias[a], R. Marrero[a,b], J. Chanussot[a,c] and P. Comon[a]

[a]GIPSA-Lab, St Martin d'Heres cedex, France
[b]Institut de Planetologie et d'Astrophysique de Grenoble (IPAG), St Martin d'Heres cedex, France
[c]Faculty of Electrical and Computer Engineering, University of Iceland, Iceland

ABSTRACT

Spectral unmixing is one of the most important and studied topics in hyperspectral image analysis. By means of spectral unmixing it is possible to decompose a hyperspectral image in its spectral components, the so-called endmembers, and their respective fractional spatial distributions, so-called abundance maps. New hyperspectral missions will allow to acquire hyperspectral images in new ways, for instance, in temporal series or in multi-angular acquisitions. Working with these incoming huge databases of multi-way hyperspectral images will raise new challenges to the hyperspectral community. Here, we propose the use of compression-based non-negative tensor canonical polyadic (CP) decompositions to analyze this kind of datasets. Furthermore, we show that the non-negative CP decomposition could be understood as a multi-linear spectral unmixing technique. We evaluate the proposed approach by means of Mars synthetic datasets built upon multi-angular in-lab hyperspectral acquisitions.

Index Terms— Multilinear spectral unmixing, hyperspectral multiangle images, multiway analysis, Canonical Polyadic, nonnegative tensor decomposition.

1. INTRODUCTION

Imaging spectroscopy [1] (*a.k.a.* hyperspectral imaging) is concerned with the measurement, analysis, and interpretation of spectra acquired from a given scene or object [2]. Hyperspectral images (HSI) are usually stored in a non-negative matrix form, $\boldsymbol{X} \in \mathbb{R}_+^{N \times D}$, where N denotes the number of pixels in the image and D denotes the number of spectral bands. Furthermore, new missions and sensor developments will collect time series of hyperspectral data, *e.g.* MODIS mission[1], and multiangle images, *e.g.* CRISM mission[2]. The huge amount of hyperspectral data that will be delivered in the near future will pose new challenges to hyperspectral image analysis.

Here, we propose to use tensor analysis [3] (*a.k.a.* multiway or multiarray analysis) to face this challenge. Time series or multi-angle hyperspectral big data could be understood as non-negative tensors, $\boldsymbol{\mathcal{X}} \in \mathbb{R}_+^{N \times D \times T}$, where N, D and T denote the dimensionality of the spatial, spectral and time/angle ways, respectively. One of the most successful techniques to decompose tensors in low-rank terms is the Canonical Polyadic decomposition (CP) [4], sometimes coined Candecomp/Parafac [5]. The CP decomposition could be understood as an extension of the linear unmixing of 2-way (spatial and spectral) hyperspectral data [6] to the multi-linear unmixing of multi-way (more than two) hyperspectral tensors.

Conventional spectral unmixing aims to decompose a hyperspectral image, $\boldsymbol{X}$, into the spectral signatures of the materials present in the image and their spatial distributions, known respectively as *endmembers* and *fractional abundances*. Following the widely used linear mixing model [7]:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{E}^\top + \boldsymbol{\Xi}, \qquad (1)$$

where $\boldsymbol{E} \in \mathbb{R}^{D \times P}$ denotes the matrix of endmembers, with each of the P columns representing the spectral signature of a given macroscopic material; $\boldsymbol{A} \in \mathbb{R}^{N \times P}$ denotes the matrix of fractional abundances corresponding to each of the P endmembers, and $\boldsymbol{\Xi} \in \mathbb{R}^{N \times D}$ denotes additive noise.

The non-negative CP decomposition allows to decompose the data tensor, $\boldsymbol{\mathcal{X}}$, into a multi-linear composition of R non-negative factor matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$, of size $N \times R$, $D \times R$ and $T \times R$, respectively, and a diagonal tensor of non-negative scaling factors, $\boldsymbol{\mathcal{L}}$, of size $R \times R \times R$:

$$\mathcal{X}_{ijk} = \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr} \lambda_r, \qquad (2)$$

where λ_r denotes the diagonal scaling factors of $\boldsymbol{\mathcal{L}}$. The *rank* of tensor $\boldsymbol{\mathcal{X}}$ is defined as the minimal number R of terms necessary for the equality above to hold exactly. In practice, the data tensor is subject to modelling errors or measurement noise, and it is convenient to find its best rank-R approxima-

tion by minimizing the following objective function

$$\Upsilon(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{\mathcal{L}}) = \|\boldsymbol{\mathcal{X}} - (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \cdot \boldsymbol{\mathcal{L}}\|, \qquad (3)$$

for some well chosen norm, instead of attempting to compute the exact CP decomposition (2). It is now known that tensors of order 3 or larger do not always admit a rank-R approximate, when $R > 1$, especially in $\mathbb{R}$ or $\mathbb{C}$. But fortunately, it has been shown in [8] that this obstacle does no longer holds for nonnegative tensors, and that the problem is well-posed in $\mathbb{R}^+$: best lower nonnegative rank approximates always exist.

The non-negative CP decomposition could be understood as an extension of the linear unmixing of 2-way (spatial and spectral) hyperspectral data [6] to the multi-linear un-mixing of multi-way (more than two) hyperspectral tensors, where rank-one factors are expected to be related to spatial abundances, spectral signatures and changes in time/angle. Furthermore, the CP decomposition is a blind technique in the sense that no *a priori* information is needed, *i.e.*, when the spectral signatures of the materials in the image are unknown. Recently, authors in [9] have proposed a compression-based nonnegative CP decomposition of large tensors. We propose to make use of this compression-based technique to perform the nonnegative CP decomposition of the multi-angular hyperspectral images and interpret the results in terms of a multi-linear spectral unmixing.

The remainder of the paper is as follows. Sec. 2 describes the ProCo-ALS compression-based nonegative CP decomposition introduced in [9]. In Sec. 3 we provide the case study with multi-angle in-lab acquisition of Martian alike hyperspectral synthetic images. Finally, we provide some conclusions and discuss on further work in Sec. 4.

2. COMPRESSION-BASED NONNEGATIVE CP DECOMPOSITION

2.1. Non-negative CP approximation

It is well known in the optimization community that computing the nonnegative CP decomposition of a positive tensor is a difficult problem. Given a multiway data set $\boldsymbol{\mathcal{X}}$, we want to solve the following minimization problem:

$$\begin{aligned} \operatorname{argmin} & \|\boldsymbol{\mathcal{X}} - (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \cdot \boldsymbol{\mathcal{L}}\|_F^2 \\ \text{w.r.t. } & \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \\ \text{s.t. } & \boldsymbol{A} \succeq 0, \boldsymbol{B} \succeq 0, \boldsymbol{C} \succeq 0 \end{aligned} \qquad (4)$$

which is highly non-convex. Yet many algorithms provide rather precise but costly computation, and these algorithms can be divided into two main classes:

- *All-at-once gradient-based descent, e.g.* [10]: all CP parameters are updated at the same time using a gradient scheme (standard or conjugate gradient) and nonnegativity constraints are assimilated through barriers or soft penalizations.

- *Alternating minimization*: the cost function is minimized in an alternating way for each factor ($\boldsymbol{A}$, $\boldsymbol{B}$ or $\boldsymbol{C}$) while

the others are fixed. The most commonly used method for nonnegative CP decomposition is alternating nonnegative least squares (ANLS), e.g. [11].

2.2. Compression-based CP decomposition

Large tensors decomposition is actually a hot topic in the tensor decomposition area, especially when constraints are included in the optimization problem. An approach to handle large tensor decomposition is through the use of compression. The general idea is that the original data array $\boldsymbol{\mathcal{X}}$ can be equivalently represented by one or a few arrays $\boldsymbol{\mathcal{X}}_c$ with reduced dimensions $N_c \times D_c \times T_c$. The compressed tensor is then decomposed by solving

$$\begin{aligned} \operatorname{argmin} \Upsilon &= \|\boldsymbol{\mathcal{X}}_c - (\boldsymbol{A}_c, \boldsymbol{B}_c, \boldsymbol{C}_c)\|_F^2 \\ \text{w.r.t. } & \boldsymbol{A}_c, \boldsymbol{B}_c, \boldsymbol{C}_c, \end{aligned} \qquad (5)$$

where $\boldsymbol{A}_c$, $\boldsymbol{B}_c$, $\boldsymbol{C}_c$ are compressed versions of the original factor matrices, with reduced number of rows N_c, D_c, T_c, but the same number of columns R. For simplification purposes the diagonal matrix of scalings is absorbed in $\boldsymbol{C}_c$. Note that, after the compressed factors are obtained, a decompression operation is carried out to recover the factors in the original dimensions.

A common approach [12, p. 92] is through an approximation of the High Order Singular Value Decomposition (HOSVD). The HOSVD [13] approximates the original data in the following way

$$\boldsymbol{\mathcal{X}}_{ijk} \approx \sum_{lmn}^{N_c, D_c, T_c} \boldsymbol{U}_{il} \boldsymbol{V}_{jm} \boldsymbol{W}_{kn} \left[\boldsymbol{\mathcal{X}}_c\right]_{ijk}, \qquad (6)$$

or using the same notation as in the CP model

$$\boldsymbol{\mathcal{X}} \approx (\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}) \boldsymbol{\mathcal{X}}_c, \qquad (7)$$

where $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$ are matrices with orthogonal unit-norm columns. In practice, these matrices are obtained by truncating the first N_c, D_c and T_c left singular vectors of the 3 unfoldings of $\boldsymbol{\mathcal{X}}$ (unfoldings are different concatenations of matrix slices of the tensor, see *e.g* [3] and references therein). Note that, in this case, the compressed and uncompressed factors are related in the following way:

$$\boldsymbol{A} \approx \boldsymbol{U} \boldsymbol{A}_c, \quad \boldsymbol{B} \approx \boldsymbol{V} \boldsymbol{B}_c, \quad \boldsymbol{C} \approx \boldsymbol{W} \boldsymbol{C}_c, \qquad (8)$$

which shows that $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$ can be seen as decompression operators.

2.3. ProCo-ALS compression-based non-negative CP decomposition

Adapted instances of the two main classes of algorithms for nonnegative CP decomposition are proposed in [9] to include the compression step. In the all-at-once gradient descent setting, the Compressed Conjugate Gradient (CCG) algorithm is

proposed, while a modification of ANLS, Projected and Compressed ALS (ProCo-ALS), is presented in the alternating setting. In this work, we make use of the later only.

The idea of ProCo-ALS is to carry out an approximate projection of the unconstrained least squares (LS) solution on the set of feasible solutions. This approximate projection is given in three steps: first the factor is decompressed (D), then the decompressed factor is projected onto the nonnegative orthant (P), and, finally, the result of the projection is recompressed (R). The four steps of each ProCo ALS iteration to update the factor $\boldsymbol{A}_c$, denoted here $\hat{\mathbf{A}}_{pc}^{k+1}$, are detailed below:

$$\text{LS:} \qquad \hat{\mathbf{A}}_c^{k+1} = \boldsymbol{X}_c^{(1)} \left(\hat{\mathbf{C}}_{pc}^k \odot \hat{\mathbf{B}}_{pc}^k \right)^{\dagger}$$

$$\text{Approx. proj.} \begin{cases} \text{D:} & \hat{\mathbf{A}}^{k+1} = \boldsymbol{U} \hat{\mathbf{A}}_c^{k+1} \\ \text{P:} & \left[\hat{\mathbf{A}}^{k+1} \right]_+ := \max \left(\mathbf{0}, \hat{\mathbf{A}}^{k+1} \right) \\ \text{R:} & \hat{\mathbf{A}}_{pc}^{k+1} = \boldsymbol{U}^{\top} \left[\hat{\mathbf{A}}^{k+1} \right]_+ \end{cases}$$

(9)

where $\odot$ denotes the Khatri-Rao product, $\dagger$ indicates the pseudoinverse, $\boldsymbol{X}_c^{(1)}$ is the unfolding in the first way of $\boldsymbol{\mathcal{X}}_c$ and $\max(\cdot, \cdot)$ denotes the element wise maximum. To obtain the updates $\hat{\mathbf{B}}_{pc}^{k+1}$ and $\hat{\mathbf{C}}_{pc}^{k+1}$, the same procedure is applied with the appropriate unfolding of $\boldsymbol{\mathcal{X}}_c$ and decompression operators.

3. STUDY CASE

Here, we provide experimental evidence of the validity of the proposed multilinear spectral unmixing, by means of compression-based nonnegative CP decomposition, using synthetic data generated to simulate multiangle acquisitions of the Martian surface. First, we introduce the methodology employed to generate the synthetic datasets. Then, we explain the experimental methodology and finally we provide the results.

3.1. Dataset

To synthesize the multiangle datasets we have made use of in-lab acquisitions of three different materials that present similar properties to the materials found in Martian surface (*coal in graphite composites*, *natural coal* and *basalt*) for 10 emergence angles: $-60°$, $-50°$, $-40°$, $-30°$, $-20°$, $20°$, $30°$, $40°$, $50°$ and $60°$. Objects reflectance changes according to light incidence and emergence angles. Introducing these aspects into the hyperspectral analysis could help to better identify the materials on the sensed scenes. Fig. 1 depicts the spectra of the three materials for each of the ten angular acquisitions. The spatial abundances have been simulated to mimic natural surfaces. For each pixel, $\mathbf{x}_j$, the abundances sum-up-to-one, that is, $\sum_{i=1}^{3} a_{ij} = 1$, $\forall j$. Also, a pixel purity constraint is imposed to the abundances, that is, the maximum value that an abundance can take is bounded: $a_{ij} \leq p$, $\forall i, j$, where $p \in (0.33, 1]$ denotes the purity value. We have considered five different pixel purity
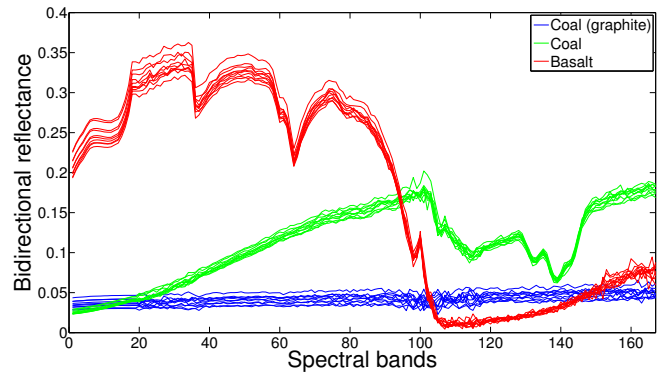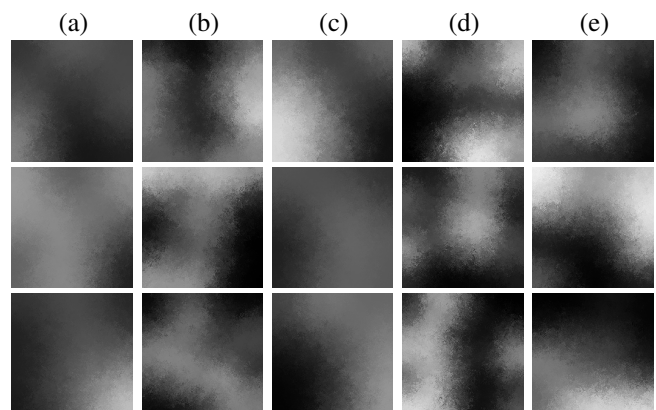

Fig. 1. Endmembers.

Fig. 2. Abundances.

values: $p \in \{0.95, 0.9, 0.85, 0.8, 0.75\}$. Lower values of p indicate a highly mixed data. Fig. 2 shows examples of abundances generated for three endmembers and the five selected pixel purity values. For each pixel purity value we have generated 10 independent sets of abundances for a total of 50 multiangle images. Each multiangle image is built using the linear mixing model in (1) where the corresponding endmembers are used according to the emergence angle. Finally, we also composed a second dataset by incorporating additive noise of 40dB to the synthetic images.

3.2. Experimental methodology

For each of the 50 data images of the two, noise-free and noisy, datasets, we performed 10 independent Monte Carlo runs of the ProCo-ALS algorithm. Each of the ProCo-ALS runs is initialized to random nonnegative factors. The original tensors have dimensions $[40000 \times 167 \times 10]$ corresponding to the spatial, spectral and multiangular ways respectively. The compressed tensor has dimensions $[100 \times 25 \times 5]$. The rank is fixed to $R = 3$. The results were compared in terms of average normalized root mean squared error ($\widehat{\text{nRMSE}}$), between the original tensor, $\boldsymbol{\mathcal{X}}$, and the tensor reconstructed
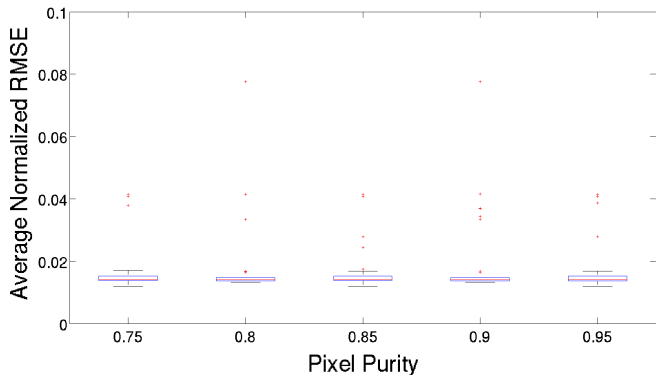
Fig. 3. Reconstruction error results data without noise.

from the CP factors, $\hat{\boldsymbol{\mathcal{X}}}$:

$$\widehat{\mathrm{nRMSE}}\left(\boldsymbol{\mathcal{X}}, \hat{\boldsymbol{\mathcal{X}}}\right) = \sqrt{\frac{1}{NDT} \sum_{i=1}^{N} \sum_{j=1}^{D} \sum_{k=1}^{T} \left(\frac{\mathcal{X}_{ijk} - \hat{\mathcal{X}}_{ijk}}{\|\boldsymbol{\mathcal{X}}\|_F}\right)^2}. \tag{10}$$

Each of the estimated spectral factors, $\hat{\mathbf{b}} \in \hat{\boldsymbol{B}}$, were compared to each of the actual spectra, $\mathbf{b} \in \boldsymbol{B}$, using the angular error, $\hat{\theta}$:

$$\hat{\theta}\left(\mathbf{b}, \hat{\mathbf{b}}\right) = \arccos\left(\frac{\mathbf{b}^\top \hat{\mathbf{b}}}{\|\mathbf{b}\|\|\hat{\mathbf{b}}\|}\right). \tag{11}$$

3.3. Results

Fig. 3 shows the reconstruction results obtained for the noise-free dataset. The ProCo-ALS nonnegative CP decomposition achieves very small reconstruction errors for all the cases. Similar results are obtained for the noisy dataset, see Fig. 4. This is not surprising since the compression step will act as a denoising filter making ProCo-ALS very robust to noise.
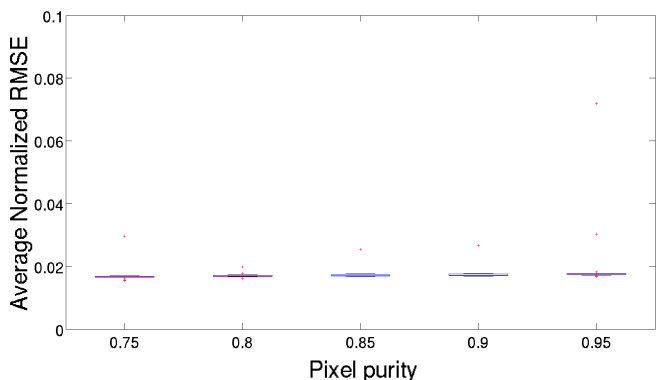

Fig. 4. Reconstruction error results data with noise.

Fig. 5 shows the average angular distances between the actual endmembers and the estimated spectral factors in the noisy dataset. For each run of the ProCo-ALS algorithm we
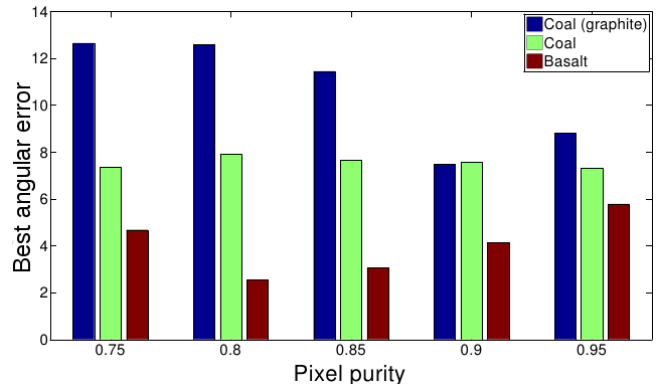

Fig. 5. Best angular errors.

compared each normalized actual endmember to each spectral factor, and kept the best pair matches of the actual endmembers. The results show that ProCo-ALS is able to estimate spectral factors similar to the actual endmembers (angular distances lower than 10 degrees) in most cases, although as the mixture contains less purity it is more difficult to retrieve an accurate spectra of the *coal (graphite)*. This is probably due to the shape of the material. Being so plane, the role of the *graphite* in the mixture with the other materials works mostly as a scaling factor, which in the CP decomposition is incorporated in the diagonal matrix, $\boldsymbol{\mathcal{L}}$.

Fig. 6 depicts the spectral factors obtained by the best (first row) and worst (second row) runs of the ProCo-ALS in terms of reconstruction error for the noisy dataset. Since the spatial factors are normalized, the scale does not correspond to the scale of the actual endmembers. Even so, the shapes of the spectral factors are very similar to the actual endmembers, so we can claim that they are physically meaningful. Finally, in Fig. 7, we show the average Pearson correlation between the actual abundances and the estimated spatial factors for the noisy dataset. In all cases the correlation is very high (greater than 0.8) for the three materials, showing that the spatial factors are linearly correlated to the physical fractional abundances, and thus, they retain their physical meaning.

4. CONCLUSION

We have proposed the use of multiway analysis to analyze hyperspectral multiway (more than two) images. In particular, we propose to use compression-based nonnegative CP decompositions. We have tested the proposed approach, making use of the ProCo-ALS algorithm, to analyze multiangle synthetic images of the Martian surface. We have given solid evidence supporting the physical interpretability of the estimated spectral and spatial factors in terms of conventional unmixing. Thus, we argue that the nonnegative CP decomposition of hyperspectral tensors could be interpreted as a multi-linear
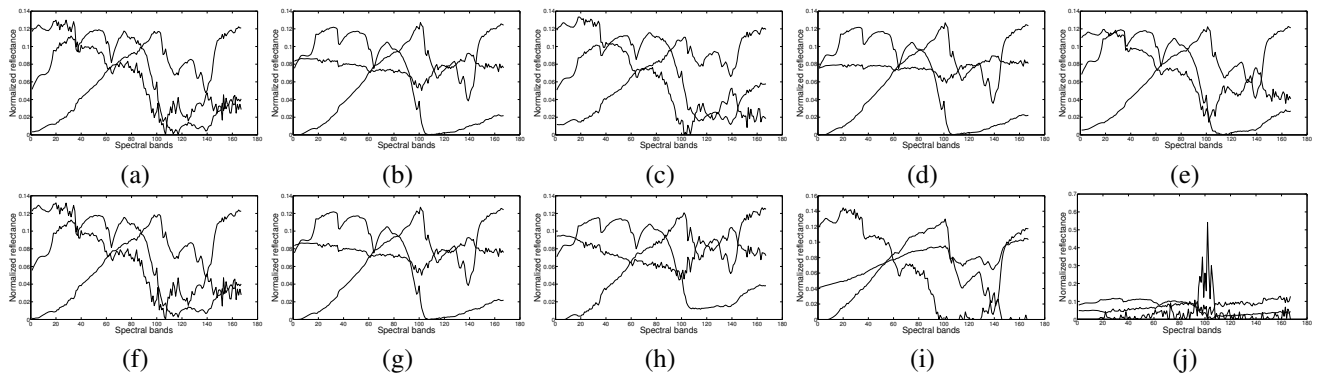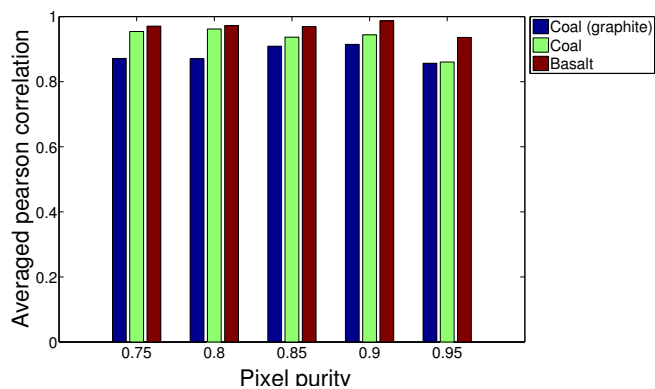
Fig. 6. Best and worst endmembers.

Fig. 7. Correlation abundances.

spectral unmixing process. Further research will extend this research avenue by incorporating physical constraints usually employed in spectral unmixing, such as the abundances sum-to-one constraint and spatial smoothness. Also, the spectral variability issue, which is a well know source of errors in spectral unmixing will be investigated, by incorporating topological and photometric information into the synthetic data. Also, further research will devote to the analysis of real Martian multiangle images taken by the CHRISM sensor.

REFERENCES

[1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.

[2] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Rem. Sens. of Env.*, vol. 113, Supplement 1, pp. 110–122, 2009.

[3] P. Comon, "Tensors : A brief introduction," *IEEE Sig. Proc. Mag.*, vol. 31, no. 3, pp. 44–53, May 2014.

[4] F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *J. Math. Physics*, vol. 6, no. 1, pp. 165–189, 1927.

[5] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, pp. 105–122, 2000.

[6] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, D. Qian, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[7] N. Keshava and J.F. Mustard, "Spectral unmixing," *IEEE Sig. Proc. Mag.*, vol. 19, no. 1, pp. 44–57, 2002.

[8] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *J. Chemometrics*, vol. 23, no. 7-8, pp. 432–441, 2009.

[9] J. E. Cohen, R. C. Farias, and P. Comon, "Fast decomposition of large nonnegative tensors," *IEEE Sig. Proc. Letters*, vol. 22, no. 7, pp. 862–866, July 2015, accepted in nov. 2014.

[10] P. Paatero, "A weighted non-negative least squares algorithm for three-way parafacfactor analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 223–242, 1997.

[11] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[12] R. Bro, *Multi-way analysis in the food industry: models, algorithms, and applications*, Ph.D. thesis, University of Amsterdam, The Netherlands, 1998.

[13] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Ana. and Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.