ACCURACY AND PERFORMANCE OF OPTIMIZED BAYESIAN SOURCE SEPARATION FOR HYPERSPECTRAL UNMIXING

Frédéric Schmidt[1], Albrecht Schmidt[2], Erwan Tréguier[2], Maël Guiheneuf[2], Saïd Moussaoui[3] and Nicolas Dobigeon[4]*

[1] IDES, UMR CNRS 8148, Université Paris Sud, Orsay, France
[2] ESA, ESAC, Villafranca del Castillo, Spain
[3] IRCCyN, UMR CNRS 6597, Nantes, France
[4] University of Toulouse, IRIT/INP-ENSEEIHT, France

ABSTRACT

Bayesian Prior Source Separation (BPSS) with positivity constraint is a useful unsupervised approach for hyperspectral data unmixing. The main interest of this approach is to ensure the positivity of the unmixed component spectra and abundances. Moreover, a recent extension has been proposed to impose the sum-to-one (full additivity) constraint to the estimated abundances. Unfortunately, even if positivity and full additivity are two necessary properties to get physically interpretable results, the use of BPSS algorithms is limited by high computation time and large memory requirements since these Bayesian algorithms employ Markov Chain Monte Carlo methods. This article introduces an implementation strategy which allows one to apply such algorithms to a full hyperspectral image, as typical in Earth and Planetary Science, with reduced computation cost. We studied a technical optimization. We also study the effect of convex hull pixel selection as a preprocessing step and discuss the impact of such preprocessing on the relevance of the estimated component spectra and abundance maps as well as on the whole computation times. For that purpose, we use two different datasets: a synthetic one and a real hyperspectral image from Mars.

Index Terms— Hyperspectral imaging, source separation, Bayesian estimation, implementation strategy, computation time.

1. INTRODUCTION

In hyperspectral imaging, each image recorded by the sensor is the solar light reflected and diffused back from the observed planet surface and atmosphere at a particular spectral band. Under some assumptions related to surface and atmosphere properties - i.e.: lambertian surface, no intimate mixture, no diffusion terms in the atmosphere, homogeneous geometry in the scene - each measured spectrum (each pixel of the observed images for several spectral bands) can be modeled as a linear mixture of the scene component (*endmembers*) spectra [1]. In this model, the weight of each component spectrum can be linked to its abundance in the surface area corresponding to the underlying pixel.

By considering P pixels of an hyperspectral image acquired at L frequency bands, the observed spectra are gathered in a $P \times L$ data matrix $\mathbf{X}$. Each row of this matrix contains a measured spectrum at a pixel with spatial index $p = 1, \ldots, P$. According to the linear mixing model, the pth spectrum can be expressed as a linear combination of the R pure spectra of the surface components. Using matrix notations, this linear spectral mixing model can be written as

$$\boldsymbol{X} \approx \boldsymbol{A}\boldsymbol{S} \tag{1}$$

The rows of matrix $\boldsymbol{S}$ contain the surface pure spectra of the R components and each element A_{pr} of matrix $\boldsymbol{A}$ corresponds to the abundance of the rth component in pixel with spatial index p. The source separation problem consists of finding matrices $\boldsymbol{S}$ and $\boldsymbol{A}$ [2].

When solving this separation problem with hyperspectral data, several constraints can be considered to reduce the set of the admissible solutions. A first hard constraint is the positivity (or non-negativity) of the elements of both matrices $\boldsymbol{S}$ and $\boldsymbol{A}$ since they correspond to pure spectra and abundances of the surface components. A second constraint that may be imposed is the sum-to-one (additivity) constraint of the abundances. Indeed the abundance weights correspond to proportions and therefore should sum to unity. This constrained separation problem can be conveniently addressed in a Bayesian framework. Two algorithms that perform unsupervised separation under positivity and sum-to-unity constraints have been recently proposed [3, 4]. The complexity of the inference from the posterior distribution of the parameters of interest is tackled using Markov Chain Monte Carlo methods. The computation time drastically increases with the image size and these algorithms have been not applied for large image processing in spite of their high effectiveness.

The aim of this article is to discuss some implementation

*Corresponding author: frederic.schmidt@u-psud.fr

strategies which allow to apply these algorithms even if image sizes are large. Previous work about blind source separation of hyperspectral images have been done [5, 6, 7] but only few using positivity/sum-to-unity constraints [8]. A previous proposal [8] is to combine independent component analysis (ICA) and Bayesian positive source separation (BPSS). This strategy presents a limitation related to the difficulty to determine the number of pixels to retain from each independent component class. In this paper, we investigate another pixel selection strategy based on the computation of the convex hull of the hyperspectral data and discuss its influence on the separation performances. The discussion about the estimation of the number of sources, or "intrinsic dimension" [9], will not be addressed in this article.

2. OPTIMIZATION

The optimization consists of two independent parts: (i) Technical Optimization (TO) to reduce the memory footprint, the average cost of algorithmic operations, and make smarter reuse of memory (ii) Convex Hull Optimization (CHO) to reduce the number of spectra to consider.

Both parts enabled us to analyze images that so far where not open to analysis. The authors stress that the techniques applied in (i) do not alter the results of the original algorithm. On the other hand, the optimization strategy (ii) only selects a subset of the original input and therefore has the potential to significantly change the results. We therefore need to evaluate the kind of impact of the strategy (ii).

2.1. Technical Optimization (TO)

So far, the algorithms introduced in [3, 4] and referred to as BPSS and BPSS2, respectively, could be successfully launched on an image of a restricted size, typically of a few thousand pixels. The main goal of this work is to optimize these algorithms to process a whole hyperspectral image of 100 000 spectra as it typically occurs in Earth and Planetary Science. Since the time requirements of the computation increase drastically with a larger number of of pixels in an image, another challenging objective is to reduce as much as possible the computing time, using all possible ways. This algorithm has been implemented in MATLAB©.

Memory: Fragmentation of memory may occur when variables are resized after the allocation. In this case, the memory management might not be able to allocate a chunk of memory that is large enough to hold the new variable. In our case, to reduce the impact of garbage collection, we found it useful to pre-allocate the matrices.

Precision: MATLAB© by default computes on double precision but the computation with single saves a lot of computation time while providing sufficient arithmetic precision. Furthermore, most datasets come as single precision.

OS Architecture: It is interesting to know that MATLAB© is limited in terms of using of the memory use depending on the Operating System (OS) and the MATLAB© version.

Parallelization: MATLAB© contains libraries to automatically parallelize parts of the algorithms on a single computer. We choose to run BPSS on a 4-core machine. The underlying matrix libraries already provide a certain level of parallelism depending on the number of available cores. However, in the future, parts of the code could be parallelized and the jobs could be submitted to a grid in order to speed up the process.

2.2. Convex Hull Optimization (CHO)

The proposed pixel selection strategy is based on the convex hull of the data matrix projection into the subspace defined by the seven first principal components. The convex hull of a point set is the smallest convex set that includes all the points [10]. It can be used as an concise description of these point feature. Consequently, the corresponding pixels in the hyperspectral images contain all the component spectra while some of them contain a high abundance of the components.

3. PERFORMANCE AND ACCURACY

3.1. Performance

The overall performance of TO are better compared to the original version of the BPSS algorithm.We estimated to win up to 60 % computation time on a x86 processor architecture. BPSS memory consumption is also significantly reduced in order to be performed on a typical hyperspectral image of 100 000 spectra and 128 wavelength. We estimate that the CHO reduce the computation time by a factor of 50 times.

3.2. Accuracy

The results of TO is found to be equivalent to the original code within numerical precision. One have to better choose the TO optimized version in order to save resources.

The effect of the CHO can lead a significant bias in accuracy. We studied both synthetic and real data in order to estimated the bias.

3.2.1. Synthetic datasets

Several synthetic datasets have been generated by mixing 3, 5 or 10 endmembers, with randomly distributed abundances with uniform distribution. The generated datasets are of size 200x500 pixels similar to a typical hyperspectral image.The following spectra have been used as endmembers: H_2O ice and CO_2 ice spectra [11, 12] and mineral spectra from the USGS Digital Spectral Library splib06a [13], resampled to the 128 wavelengths of OMEGA C Channel [14]. To ensure the sum-to-one constraint on the *n* endmember abundances, a uniform distribution on the *n*-simplex has been used.

Based on this method, datasets for which the maximum abundance of each single endmember was limited to a certain value (100%, 80% and 60%) have also been considered. This latter data, that are called 'cutoff', allows one to test the

method efficiency in front of various conditions in terms of purity of the samples (in cases where pure components occur in the dataset or not).

In addition, a 3 components asymmetric dataset was investigated, with the abundances of one component (albite) being limited to a cutoff of 35% and the abundance of the two others (ices) not being limited. Note that for all the considered simulation scenarii, the number of sources to be estimated has been tuned to the actual number of endmembers used to produce the artificial dataset.

Due to curse of dimensionality, the more endmembers to be estimated with the fixed number of wavelength, the more difficult is the estimation. Still, BPSS2 gives excellent results for 10 sources, as all spectra are estimated with a correlation coefficient higher than 99% (see fig. 1).

In most of the tested cases, the quality of the estimation is unambiguously better with BPSS2 than with BPSS.

The cutoff affects the quality of the estimation, which is clearly better, for BPSS and BPSS2, when pure components occur in the dataset.

Our results clearly show that the method is very robust to noise. BPSS2 (without pixel selection) even manages to successfully overcome the addition of a 100-times amplified OMEGA-like noise.

For a cutoff from 80% to 60%, the source estimation of BPSS and BPSS2, after the convex hull selection, is significantly lower than without convex hull.

In case of an asymmetry in the cutoff, the results are better with pixel selection rather than without. BPSS2 with pixel selection is the only run (performed on this synthetic dataset) that allows to successfully estimate the 3 endmembers used to produce the dataset, including albite, whose abundances were limited to a cutoff of 35% and whose spectral signature is weaker than the one of the two other endmembers (ices). This result can be explained by the fact that pixel selection is able to extract the pixels with the strongest available albite signature, and consequently overcome the blinding effect of the ices occurring in the whole dataset, that affected the results when no pixel selection was performed.

3.2.2. OMEGA image

OMEGA [14] is a imaging spectrometer onboard Mars Express. We propose to study a reference image of the south polar cap of Mars [1].

When the convex hull selection is used as a pre-processing step to BPSS, the estimation is significantly better (see fig. 2 and 3). This results show that pixel selection is a way to reponderate the occurrence of rare endmember and thus is a interesting method to provide better results.

[1] http://sites.google.com/site/fredericschmidtplanets/Home/hyperspectral-reference

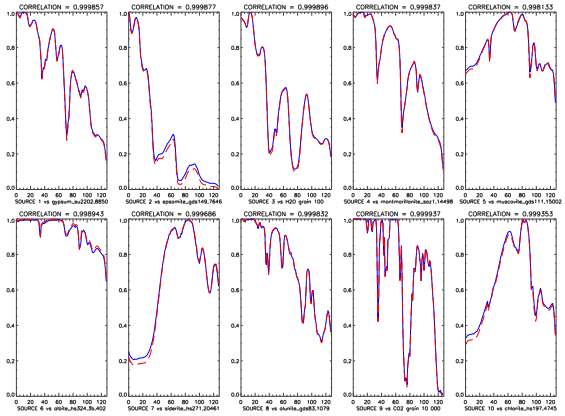

Fig. 1. Sources estimated by BPSS2 (blue lines) and their spectral matches (red dotted lines), for an artificial dataset with 10 endmembers (no cutoff, no noise).

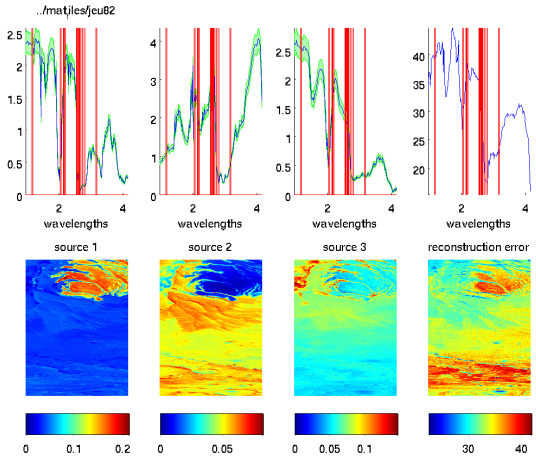

Fig. 2. Estimation of 3 sources of the entire OMEGA image 41_1 with BPSS using a preprocessing step of pixel selection using the convex hull method. The first and third source are clearly identified to CO_2 and H_2O ices with a correlation coefficient of 0.953 and 0.940. The spatial abundances are well estimated regarding the WAVANGLET classification method [12, 8]. The second source is identified to dust with a lower correlation coefficient (0.372).

4. CONCLUSION

We propose an optimization of bayesian algorithm to study actual hyperspecral image and show that the technical optimization is equivalent and faster than the original algorithms. We also show that pixel selection, using convex hull, decrease the accuracy of the estimation for cutoff from 80% to 60%. Nevertheless, pixel selection increase the accuracy of the es-

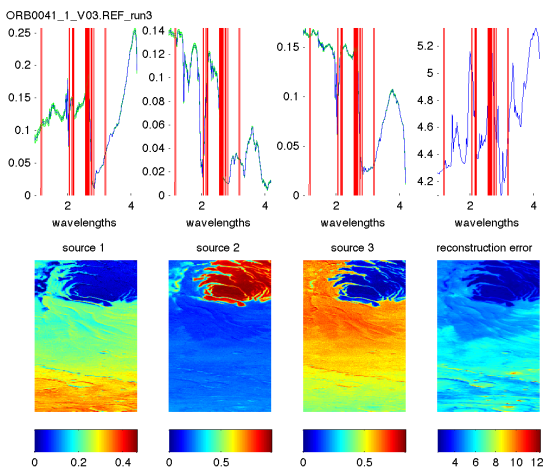

Fig. 3. Estimation of 3 sources of the entire OMEGA image 41_1 with BPSS without pixel selection. The second source is clearly identified to CO_2 ice with a correlation coefficient of 0.957. The first and third sources are identified to dust and water ice with lower correlation coefficients of 0.555 and 0.773. The spatial abundances of water ice is not well estimated regarding the WAVANGLET classification method [12].

timation when the sources are distributed with asymmetry, as suggested on results of both synthetic and real hyperspectral image.

5. ACKNOWLEDGEMENT

We would like to thank the OMEGA Team for the data management. We are also grateful to S. Douté and B. Schmitt for their ice spectral library. We acknowledge support from the Faculty of the European Space Astronomy Centre (ESAC).

6. REFERENCES

[1] D. Tanre, M. Herman, P. Y. Deschamps, and A. de Leffe, "Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties," *Applied Optics*, vol. 18, pp. 3587–3594, Nov. 1979.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, ser. Adaptive and Learning Systems for Signal Processing, Communications, and Control. New York: John Wiley, 2001.

[3] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, "Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling," *IEEE Trans. on Signal Processing*, vol. 54, no. 11, pp. 4133–4145, Nov. 2006.

[4] N. Dobigeon, S. Moussaoui, J.-Y. Tourneret, and C. Carteret, "Bayesian separation of spectral sources under non-negativity and full additivity constraints," *Signal Processing*, vol. 89, no. 12, pp. 2657–2669, Dec. 2009.

[5] M. Naceur, M. Loghmari, and M. Boussema, "The contribution of the sources separation method in the decomposition of mixed pixels," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 11, pp. 2642–2653, 2004.

[6] J. Nascimento and J. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 1, pp. 175–187, 2005.

[7] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 44, no. 9, pp. 2601–2616, 2006.

[8] S. Moussaoui, H. Hauksdottir, 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," *Neurocomputing*, vol. 71, no. 10-12, pp. 2194–2208, 2008.

[9] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 42, no. 3, pp. 608–619, 2004.

[10] C. B. Barber, D. Dobkin, and H. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, Dec. 1996.

[11] S. Douté and B. Schmitt, "A multilayer bidirectional reflectance model for the analysis of planetary surface hyperspectral images at visible and near-infrared wavelengths," *J. Geophysical Research*, vol. 103, no. 12, pp. 31 367–31 390, Dec. 1998.

[12] F. Schmidt, S. Douté, and B. Schmitt, "Wavanglet: An efficient supervised classifier for hyperspectral images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 45, no. 5, pp. 1374–1385, 2007.

[13] R. N. Clark *et al.*, "Usgs digital spectral library splib06a," U.S. Geological Survey, 2007. [Online]. Available: http://speclab.cr.usgs.gov

[14] J.-P. Bibring et *al.*, "Perennial water ice identified in the south polar cap of Mars," *Nature*, vol. 428, pp. 627–630, april 2004.