

# Hyperspectral BSS using GMCA with *spatio-spectral* sparsity constraints

Y. Moudden\* and J. Bobin

**Abstract**—GMCA is a recent algorithm for multichannel data analysis which was used successfully in a variety of applications including multichannel sparse decomposition, blind source separation (BSS), color image restoration and inpainting. Building on GMCA, the purpose of this contribution is to describe a new algorithm for BSS applications in hyperspectral data processing. It assumes the collected data is a mixture of components exhibiting sparse spectral signatures as well as sparse spatial morphologies, each in specified dictionaries of spectral and spatial waveforms. We report on numerical experiments with synthetic data and application to real observations which demonstrate the validity of the proposed method.

**Index Terms**—GMCA, MCA, sparsity, morphological diversity, hyperspectral data, multichannel data, Blind Source Separation, wavelets, curvelets.

## I. INTRODUCTION

Over the last few years, the use of multi-channel sensors has spread widely in a variety of research fields ranging from astronomy to geophysics. This has raised interest in methods for the coherent processing of multivariate data, as well as more specific approaches for hyperspectral data. In this context, the data matrix  $\mathbf{X} \in \mathbb{R}^{m,t}$  is composed of images of size  $\sqrt{t} \times \sqrt{t}$  observed in  $m$  different wavelength bands. A widely used approach to model such data consists in assuming that each row  $x_p$  of  $\mathbf{X}$  is the linear combination of  $n$  so-called sources :  $\forall i = 1, \dots, m; x_p = \sum_k a_{pk} s_k + n_p$  where  $s_k$  is known as a source and  $a_{ik}$  models for the contribution of the  $k$ -th source in the  $p$ -th channel. The term  $n_i$  stands for noise or source imperfections. By defining the so-called mixing matrix  $\mathbf{A}$  the entries of which are  $\mathbf{A}[p, k] = a_{pk}$  and the source matrix  $\mathbf{S}$  the rows of which are the sources  $\{s_i\}_{i=1, \dots, n}$ , the data  $\mathbf{X}$  are more concisely modeled as follows :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}$$

where  $\mathbf{N}$  models some additive noise contribution. Blind source separation methods then aim at estimating both  $\mathbf{A}$  and  $\mathbf{S}$  from the data  $\mathbf{X}$ . Several statistical approaches have been applied to solve this problem. In a nutshell, designing an effective blind source separation method reduces to finding a measure of diversity between the sources. In the last two decades, the mainstream approach has been independent component analysis (ICA - see [1], [2] and references therein). These statistical approaches aim at designing blind source separation methods that enforce the statistical independence of the sought after sources.

Inspired by recent advances in computational harmonic analysis, sparsity-based blind source separation methods have been introduced in [3], [4]. More specifically, Generalized Morphological Component Analysis (GMCA) is a recent algorithm designed in [4] which is used to decompose a given data matrix  $\mathbf{X} \in \mathbb{R}^{m,t}$  into a specified

number  $n$  of rank one contributions  $\mathbf{X}_k$  with different statistical and *spatio-spectral* properties. Each matrix  $\mathbf{X}_k$  is the product of a spectral signature  $a^k \in \mathbb{R}^{m,1}$  and a spatial density profile  $s_k \in \mathbb{R}^{1,t}$ . A major assumption of GMCA is that each  $s_k$  has a sparse representation  $\nu_k$  in a given dictionary of spatial waveforms  $\Phi \in \mathbb{R}^{t,t'}$ , which for simplicity we take to be the same for all  $k$ . In matrix form, we write :

$$\begin{aligned} \mathbf{X} &= \sum_k \mathbf{X}_k + \mathbf{N} = \sum_k a^k s_k + \mathbf{N} & (1) \\ &= \mathbf{AS} + \mathbf{N} = \sum_k a^k \nu_k \Phi + \mathbf{N} & (2) \end{aligned}$$

where the  $k^{\text{th}}$  line of  $\mathbf{S} \in \mathbb{R}^{n,t}$  is  $s_k$  and the  $k^{\text{th}}$  column of  $\mathbf{A} \in \mathbb{R}^{m,n}$  is  $a^k$ . The  $m \times t$  random matrix  $\mathbf{N}$  is included to account for modeling errors, or instrumental noise, assumed to be Gaussian, uncorrelated inter- and intra- channels, with variance  $\sigma^2$ . In the case of multichannel image data, the image from the  $p^{\text{th}}$  channel is formally represented here as the  $p^{\text{th}}$  line of  $\mathbf{X}$ ,  $x_p$ . The importance of sparsity in blind source separation was recently recognized in [3]. The sparse coefficient vector  $\nu_k \in \mathbb{R}^{1,t'}$  has most of its entries close to zero while only a few have significant amplitudes. In addition to this marginal property of the sparse representations  $\nu_k$ , GMCA also requires morphological diversity to achieve its decomposition which is a property of their joint distribution. Let  $\nu_k$  be the  $k^{\text{th}}$  line of matrix  $\nu \in \mathbb{R}^{n,t'}$ . The latter property expresses the assumption that there is little probability that a column of  $\nu$  will have more than one significant entry. This is true for instance of sparse independent random processes. It is also true of a random vector generated such that at most one entry is significant, in which case the entries are not independent variables. In that setting, sparsity and morphological diversity helps discriminating between the sought after sources. Furthermore, the use of sparse representations also makes GMCA more robust to noise than commonly used methods. In the context of BSS, GMCA has been shown to outperform standard state-of-the-art methods. Beyond source separation problems, GMCA was applied successfully in a variety of multichannel data processing applications, color image restoration and inpainting [4], [5].

## Contribution of this paper

Building on GMCA, the purpose of this contribution is to describe a new algorithm for so-called *hyperspectral* data processing. In what follows, regardless of other definitions or models living in other scientific communities, the term *hyperspectral* will be used in reference to multichannel data with the following two specific properties : first that the number of channels is large and second that these achieve a *regular* if not *uniform* sampling of some additional and meaningful physical index (*e.g.* wavelength, space, time) which we refer to as the *spectral* dimension. Typically, hyperspectral imaging systems collect data in a large number (up to several hundreds) of contiguous intervals of the electromagnetic spectrum. For such data, in a BSS setting for instance, one may be urged by prior knowledge to set additional constraints on the estimated parameters  $A$  and  $S$  such as equality or positivity constraints but also regularity constraints not only in the spatial dimension but in the spectral dimension as well. For instance, it may be known a *priori*

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permission@ieee.org.

Y. Moudden is with DSM/IRFU/SEDI, CEA/Saclay, F-91191 Gif-sur-Yvette, France, e-mail: yassir.moudden@cea.fr

J. Bobin is with the Department of Mathematics, Stanford University, Stanford, California 94305, USA

Manuscript received ..... ; revised .....



### III. GMCA ALGORITHM FOR *hyperspectral* DATA

For the sake of simplicity, consider now that the multichannel dictionary  $\Omega = \Psi \otimes \Phi$  reduces to a single orthonormal basis, tensor product of orthonormal bases  $\Psi$  and  $\Phi$  of respectively spectral and spatial waveforms. In this case, the minimization problem (6) is best formulated in coefficient space as follows :

$$\min_{\{\gamma^k, \nu_k\}} \frac{1}{2\sigma^2} \|\alpha - \gamma\nu\|_2^2 + \sum_{k=1}^n \lambda_k \|\gamma^k \nu_k\|_1 \quad (8)$$

where the columns of  $\gamma$  are  $\gamma^k$ , the rows of  $\nu$  are  $\nu_k$  and  $\alpha = \Psi^T \mathbf{X} \Phi^T$  is the coefficient matrix of data  $\mathbf{X}$  in  $\Omega$ . Thus, we are seeking a decomposition of matrix  $\alpha$  into a sum of sparse rank one matrices  $\alpha_k = \gamma^k \nu_k$ .

Unfortunately, there is no obvious closed form solution to problem (8), which is also clearly non-convex. Similarly to the GMCA algorithm, we propose instead a numerical approach by means of a block-coordinate relaxation (BCR) [15] iterative algorithm, alternately minimizing with respect to  $\gamma$  and  $\nu$ . Indeed, thanks to the chosen prior, for fixed  $\gamma$  (resp.  $\nu$ ), the *marginal* minimization problem over  $\nu$  (resp.  $\gamma$ ) is convex and is readily solved using a variety of methods. Inspired by the iterative thresholding methods described in [16]–[18], akin to Projected Landweber algorithms, we obtain the following system of update rules :

$$\nu^{\text{new}} = \Delta_\eta \left( (\gamma^T \gamma)^{-1} \gamma^T \alpha \right) \quad (9)$$

$$\gamma^{\text{new}} = \Delta_\zeta \left( \alpha \nu^T (\nu \nu^T)^{-1} \right) \quad (10)$$

where vector  $\eta$  has length  $n$  and entries  $\eta[k] = \frac{\sigma^2 \lambda_k \|\gamma^k\|_1}{\|\gamma^k\|_2^2}$ , while  $\zeta$  has length  $m$  and entries  $\zeta[k] = \frac{\sigma^2 \lambda_k \|\nu_k\|_1}{\|\nu_k\|_2^2}$ . The multichannel soft-thresholding operator  $\Delta_\eta$  acts on each row  $k$  of  $\nu$  with threshold  $\eta[k]$  and  $\Delta_\zeta$  acts on each column  $k$  of  $\gamma$  with threshold  $\zeta[k]$ . Equations (9) and (10) are easily interpreted as thresholded alternate least squares solutions. A complete derivation of these update rules is given in Appendix A.

Finally, in the spirit of the fast GMCA algorithm [4], it is proposed that a solution to problem (8) can be approached efficiently using the following symmetric iterative thresholding scheme with a progressively decreasing threshold, which we refer to as *hypGMCA* :

1. Set the number of iterations  $I_{\max}$  and initial thresholds  $\lambda_k^{(0)}$
2. Transform the data  $\mathbf{X}$  into  $\alpha$
3. While  $\lambda_k^{(h)}$  are higher than a given lower bound  $\lambda_{\min}$ ,
  - Update  $\nu$  assuming  $\gamma$  is fixed using equation (9).
  - Update  $\gamma$  assuming  $\nu$  is fixed using equation (10).
  - Decrease the thresholds  $\lambda_k^{(h)}$ .
5. Transform back  $\gamma$  and  $\nu$  to estimate  $\mathbf{A}$  and  $\mathbf{S}$ .

With the threshold successively decaying towards zero along iterations, the current sparse approximations for  $\gamma$  and  $\nu$  are progressively refined by including finer structures spatially and spectrally, alternately. This *salient to fine* estimation process is the core of *hypGMCA*. The final threshold should vanish in the *noiseless* case or it may be set to a multiple of the noise standard deviation as in common detection or denoising methods. Soft thresholding results from the use of an  $\ell_1$  sparsity measure, which comes as an approximation to the  $\ell_0$  pseudo-norm. Applying a hard threshold instead towards the end of the iterative process, may lead to better results as was noted experimentally in [4], [5]. When non-unitary or redundant transforms are used, the above is no longer strictly valid. Nevertheless, simple

shrinkage still gives satisfactory results in practice as studied in [19]. In the end, implementing the proposed update rules requires only a slight modification of the GMCA algorithm given in [5]. Where a simple least squares linear regression was used in the GMCA update for  $a^k$ , the proposed update rule applies a thresholding operator to the least squares solution thus enforcing sparsity on the estimated spectral signatures as *a priori* desired. The case where the dictionary  $\Omega$  is the union of several orthonormal bases  $\Omega_k$  may also be handled with a BCR approach. Update rules are easily derived, leading however to a much slower algorithm requiring the different forward and reverse transformations to be applied at each iteration.

### IV. NUMERICAL EXPERIMENTS

In this section, we compare the performance of *hypGMCA* and GMCA in toy BSS experiments with 1D and 2D. First we consider synthetic 2D data consisting of  $m = 128$  mixtures of  $n = 5$  image sources, generated according to the linear mixing model (1). The sources were drawn at random from a set of structured  $128 \times 128$  images shown on Figure 1. These images provide us with 2D spatially structured processes which are sparse enough in the curvelet domain [20]. The spectral signatures, *i.e.* the columns of the mixing matrix, were generated as sparse processes in some orthogonal wavelet domain given *a priori*. The wavelet coefficients of the spectra were sampled from a Laplacian probability density with scale parameter  $\mu = 1$ . Finally, white Gaussian noise with variance  $\sigma^2$  was added to the pixels of the synthetic mixture data in the different channels. Figure 2 displays four typical noisy simulated mixture data with SNR = 20dB.

The graph on figure 4 traces the evolution of  $\mathcal{C}_A = \|\mathbf{I}_n - \mathbf{P}\tilde{\mathbf{A}}^\dagger \mathbf{A}\|_1$ , which we use to assess the recovery of the mixing matrix  $\mathbf{A}$ , as a function of the SNR which was varied from 0 to 40dB. Matrix  $\mathbf{P}$  serves to reduce the scale and permutation indeterminacy inherent in model (3) and  $\tilde{\mathbf{A}}^\dagger$  is the pseudo-inverse of the estimated matrix of spectral signatures. In simulation, the true source and spectral matrices are known so that  $\mathbf{P}$  can be computed easily. Criterion  $\mathcal{C}_A$  is then strictly positive, and null only if matrix  $\mathbf{A}$  is correctly estimated up to scale and permutation. Finally, as we expected since it benefits from the added *a priori* spectral sparsity constraint it enforces, the proposed *hypGMCA* is clearly more robust to noise. A visual inspection of figure 3 allows a further qualitative assessment of the improved source recovery provided by correctly accounting for *a priori* spatial as well as spectral sparsity. The images on the right hand side were obtained with GMCA while the images on the left were obtained with *hypGMCA*. In all cases, both methods were run in the curvelet domain [20] with the same number of iterations.

#### *Behaviour in higher dimensions*

In a second experiment, GMCA and *hypGMCA* are compared as the number  $n$  of sources is increased while the numbers of samples  $t$  and channels  $m$  are kept constant. Then, increasing the number of sources in the mixture makes the separation task more difficult. We consider now 1D synthetic source processes  $\mathbf{S}$  generated from *i.i.d.* Laplacian probability density distributions with scale parameter  $\mu = 1$ . The Dirac basis was taken as the dictionary of spatial waveforms  $\Phi$ . The entries of the mixing matrix are also drawn from *i.i.d.* Laplacian distributions with scale parameter  $\mu = 1$  and the Dirac basis was also taken as dictionary of spectral waveforms  $\Psi$ . The data are not contaminated by noise. The number of samples is  $t = 2048$  and the number of channels is  $m = 128$ . Figure 5 depicts the comparisons between GMCA and its extension to the hyperspectral setting. Each point of this figure has been computed as the mean over 100

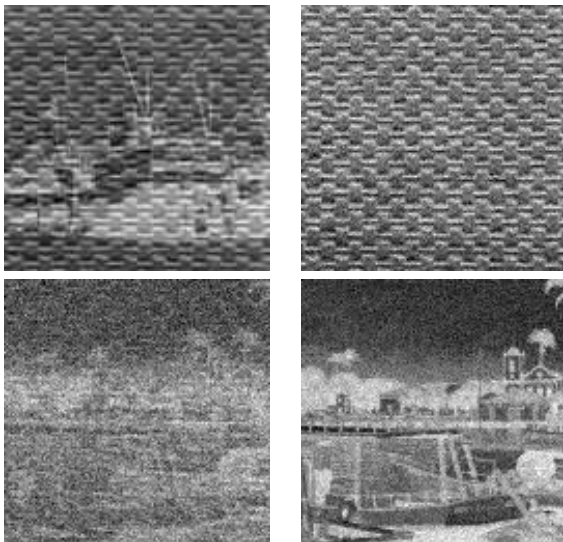


Figure 2. Four  $128 \times 128$  mixtures out of the 128 channels. The SNR is equal to 20dB.

trials. The top panel of Figure 5 features the evolution of the recovery SNR when the number of sources varies from 2 to 64. At lower  $n$ , the *spatiospectral* sparsity constraint only slightly enhances the source separation. However, as  $n$  becomes larger than 15 the spectral sparsity constraint clearly enhances the recovery results. For instance, when  $n = 64$ , *hypGMCA* outperforms the original GMCA by up to 12dB. The lower panel of Figure 5 shows the behavior of both algorithms in terms of  $\mathcal{C}_1 = \sum_{i=1}^n \|a^i s_i - \tilde{a}^i \tilde{s}_i\|_1 / \sum_{i=1}^n \|a^i s_i\|_1$ . As expected, accounting for spectral sparsity yields sparser results. Furthermore, as the number of sources increases, the deviation between the aforementioned methods becomes wider.

#### Additional positivity constraint

In hyperspectral data models, the sources often have an interpretable physical meaning (temperature, reflectance, etc). Hence, the entries of the sources to be estimated ought to be positive. The *hypGMCA* algorithm is adapted to account for the positivity of the sources. The new update rules are derived in Appendix B.

In the present experiment, the sources are drawn randomly from the set of normalized  $128 \times 128$  images shown on Figure 1. The number of sources in the synthetic mixtures is  $n = 5$ . The associated spectral signatures are generated from a Laplacian probability density with scale parameter  $\mu = 1$  in a given orthogonal wavelet basis. Both the columns of the mixing matrix and the sources are constrained to be positive. The number of channels is  $m = 128$ . White Gaussian noise with variance matrix  $\sigma^2$  is added to the mixed sources in each channel.

The separation algorithms, GMCA, *hypGMCA* and *hypGMCA* with positivity constraints, were used in the curvelet domain with 100 iterations. Figure 6 pictures the evolution of the mixing matrix criterion  $\mathcal{C}_A$  when the SNR varies from 0 to 40dB. To further assess the efficiency of our algorithms, figure 6 also displays the results obtained using two state of the art methods developed for hyperspectral data unmixing in geoscience and remote sensing applications namely the Vertex Component Analysis method (VCA) [7] and the Minimum Volume Enclosing Simplex Algorithm (MVES) [8]. Matlab implementations available online for both methods were used in this comparison. These two unmixing algorithms aim at estimating the spectral signatures of the endmembers (mixing matrix) as well as their fractional abundance maps (sources) based on the observed mixture



Figure 3. **Left column** : Estimated sources using the original GMCA algorithm. **Right column** : Estimated sources using the new *hypGMCA*.

data. VCA and MVES both assume positive sources, a property which is true of the synthetic data set used here. However, VCA and MVES make some additional assumptions which are not made by *hypGMCA*. For instance, VCA and MVES assume the fractional abundances of all endmembers sum to one in all pixels. The data was scaled accordingly before applying these methods. VCA also assumes noiseless data and the existence of pure pixels. Low purity is shown in [8] to impact heavily on the performance of VCA and may explain the bad results we obtained with this algorithm even at high SNR. On the other hand, MVES does not assume purity and proves to be a very efficient unmixing algorithm at high SNR, the performance of *hypGMCA+positivity* being only slightly better.

Both VCA and MVES operate in pixel space and make no use of the spatial structures in the abundance maps so that both have rapidly degrading performance as the noise level increases. By design, *hypGMCA+positivity* takes advantage of the *spatiospectral* sparsity of the data. This feature obviously provides our method with greater robustness to noise as shown on figure 6. As expected, accounting for the additional positivity prior results in greater efficiency of *hypGMCA+positivity* over *hypGMCA* and GMCA and provides significant separation enhancement at all noise levels. This is partly explained by the lower dimension of parameter space to be explored

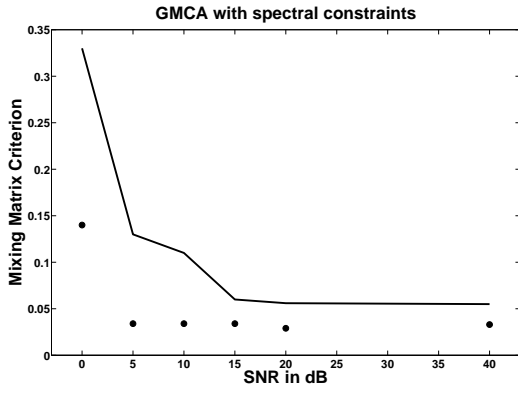


Figure 4. Evolution of the mixing matrix criterion  $C_A$  as a function of the SNR in dB. Solid line : recovery results with GMCA. • : recovery results with hypGMCA.

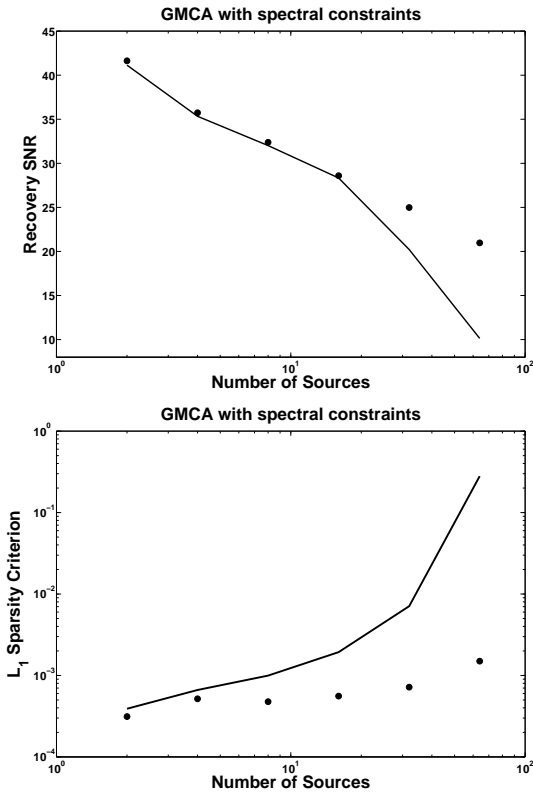


Figure 5. Abscissa : Number of sources. Ordinate - top : Recovery SNR. Ordinate - bottom : sparsity-based criterion  $C_1$ . Solid line : recovery results with GMCA. • : recovery results with hypGMCA.

for source estimation *i.e.*  $nt/2^t$  instead of  $nt$ .

## V. DECOMPOSITION OF MARS HYPERSPECTRAL DATA

In this section, we illustrate the good behavior of *hypGMCA* for real-world hyperspectral data analysis. We applied the proposed algorithm to hyperspectral data from the 128 channels of spectrometer OMEGA on Mars Express ([www.esa.int/marsexpress](http://www.esa.int/marsexpress)), at wavelengths ranging from  $0.93 \mu\text{m}$  to  $2.73 \mu\text{m}$  with a spectral resolution of 13 nm. The data are calibrated so that each pixel measures a reflectance. Example maps collected in four different channels are shown on figure 7. Model (3) is clearly too simple to describe this hyperspectral reflectance data set. Non-linear instrumental and

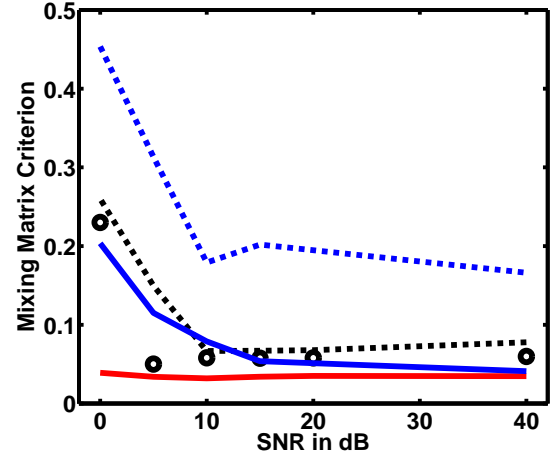


Figure 6. Evolution of the mixing matrix criterion  $C_A$  as a function of the SNR in dB. Black dotted line : recovery results with GMCA. • : recovery results with *hypGMCA*. Red solid line : recovery results with *hypGMCA* and the additional positivity constraints on the sources and the mixing matrix. Blue dotted line : recovery results VCA. Blue solid line : recovery results with MVES.

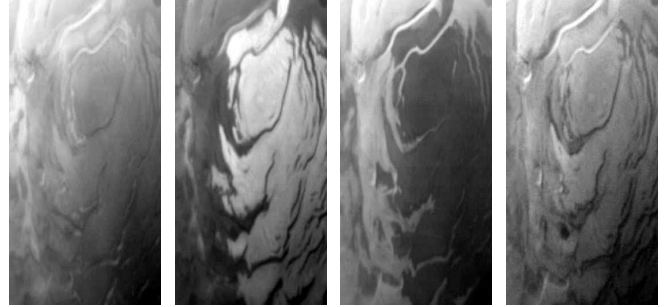


Figure 7. From left to right : Mars Express observations at wavelengths = 1.38 - 1.75 - 1.94 and 2.41  $\mu\text{m}$ .

atmospheric effects are most likely to contribute to the *true* generative process. In any case, following a similar decision made in [6], we use *hypGMCA* to fit the linear mixing model 3 to the data. Obviously, this is making the physically plausible assumption that the contributing components we seek to separate have sparse spectral signatures as well as sparse spatial concentration maps in *a priori* specified orthogonal wavelet bases. We also assume that the sources are positive. In the *hypGMCA* algorithm, this constraint is enforced by projection of the estimated source maps  $S$  on the cone generated by the vectors with positive entries. The number of iterations is  $I_{max} = 250$ .

VCA, MVES and *hypGMCA* have been set up so that the number components to be estimated is  $n = 10$ . Among the  $n = 10$  estimated components, for each of the three methods applied, we focus on the two components with corresponding spectra which are most correlated with the  $\text{H}_2\text{O}$  and  $\text{CO}_2$  ice reference spectra. Figure 8 displays their spectral signatures compared to the reference spectra, and Figure 9 shows the corresponding spatial density (positive) maps. Given the non-linearity of the physical mixture process, the close fit between the estimated and reference  $\text{CO}_2$  spectra is satisfactory. This figure also shows the spectra estimated with VCA and MVES. With these methods the  $\text{CO}_2$  spectrum is also very well estimated. The  $\text{H}_2\text{O}$  ice spectrum estimated with *hypGMCA* is remarkably similar to the reference spectra for wavelengths higher than  $1.4 \mu\text{m}$ . In that case, VCA and MVES seems to be better in the range  $0.93 - 1.4 \mu\text{m}$ . Let us

notice that the spectral behavior of the  $\text{CO}_2$  and  $\text{H}_2\text{O}$  are similarly flat in this range. The range  $> 1.4 \mu\text{m}$ , where these components exhibit several modes, is slightly more interesting. In this spectral band, VCA performs rather well with the exception that the first mode around  $1.5 \mu\text{m}$  is not accurately estimated. MVES provides good results in the band  $1.4 - 2.3 \mu\text{m}$  but the estimated  $\text{H}_2\text{O}$  spectrum diverges from the reference spectrum in the band  $> 2.3 \mu\text{m}$ .

As also noted in [6], the  $\text{CO}_2$  ice appears located in large regions around the pole of planet Mars, while  $\text{H}_2\text{O}$  ice seems to be concentrated in some tight interstices of the Mars surface. Let us notice that this component is clearly more challenging as it is less preponderant in the data than  $\text{CO}_2$ . In that case, the three methods perform quite well in estimating the  $\text{H}_2\text{O}$  spectrum. In the end, despite the simple linear mixture model we used, *hypGMCA* is able to extract components with spectral signatures that closely match reference spectra with performances at least similar to the state-of-the-art methods.

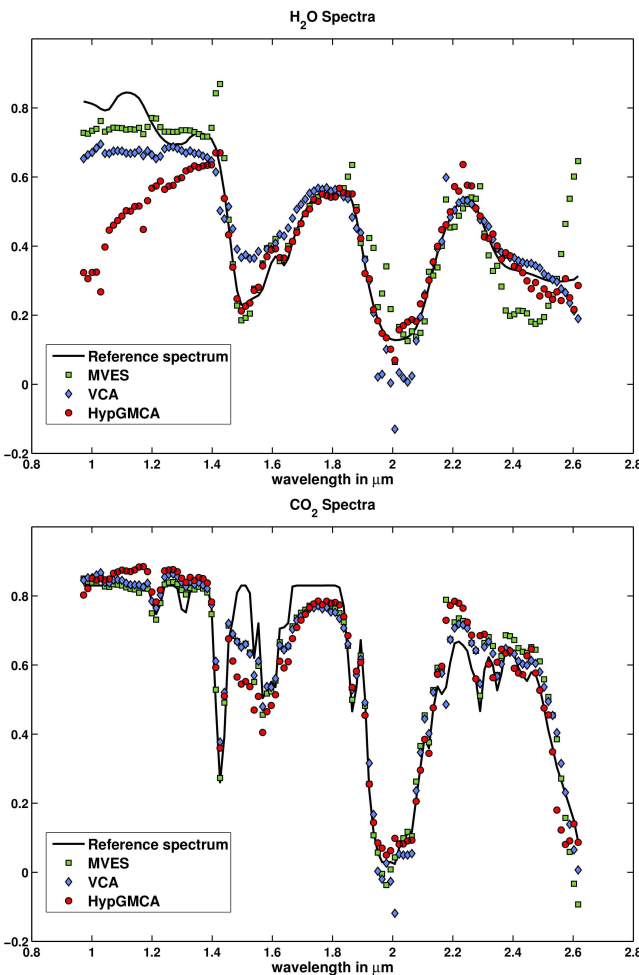


Figure 8. **Top picture** : Reference (solid line) and estimated spectra for  $\text{H}_2\text{O}$  ice. **Bottom picture** : Reference (solid line),  $\square$  and estimated spectra for  $\text{CO}_2$  ice. Legend: ( $\square$ ): MVES, ( $\diamond$ ): VCA and ( $\bullet$ ): *hypGMCA*.

## VI. CONCLUSION

We described a new algorithm, *hypGMCA*, for blind source separation in the case where it is known *a priori* that the spatial and spectral features in the data have sparse representations in known dictionaries of template waveforms. The proposed method relies on an iterative thresholding procedure with a progressively

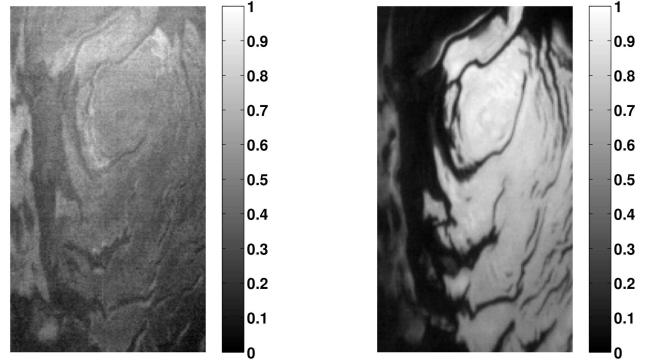


Figure 9. Estimated spatial concentration maps of  $\text{H}_2\text{O}$  ice (left) and  $\text{CO}_2$  ice (right).

decreasing threshold. As expected, and confirmed in numerical experiments, taking into account the additional prior knowledge of spectral sparsity enhances source separation. It also provides greater robustness to noise contamination as well as stability when the dimensionality of the problem increases. We also noted that accounting for the prior knowledge that the sources are positive requires only a slight modification of the algorithm. Finally, the proposed method was applied to real hyperspectral data from Omega on Mars Express. The close match between the learned spectra and the reference spectra is remarkable.

**Acknowledgments** : The authors are grateful to Olivier Forni for providing the hyperspectral data from Omega on Mars Express.

## APPENDIX A DERIVATION OF THE UPDATE RULES

Let us first point out that  $\nu_k$  is updated assuming that  $\gamma^k$  is fixed :

$$\min_{\nu_k} \frac{1}{2} \|\alpha - \gamma \nu\|_2^2 + \sigma^2 \lambda_k \|\gamma^k \nu_k\|_1 \quad (11)$$

We denote  $f_1(\nu_k) = \frac{1}{2} \|\alpha - \sum_{j \neq k} \gamma^j \nu_j - \gamma^k \nu_k\|_2^2$  and  $f_2(\nu_k) = \sigma^2 \lambda_k \|\gamma^k \nu_k\|_1$ . Then, the sources  $S$  are updated by solving the following minimization problem :

$$\min_{\nu_k} f_1(\nu_k) + f_2(\nu_k)$$

$f_1$  is a differentiable quadratic function; its gradient is Lipschitz with some constant  $L$ .  $f_2$  is a convex but nonsmooth function (it is not differentiable). Inspired by the iterative thresholding methods described in [16]–[18], akin to Projected Landweber algorithms,  $\nu_k$  is updated assuming that  $\gamma^k$  is fixed :

$$\nu_k^{new} = \text{prox}_{\rho f_2}(\nu_k - \rho \nabla f_1(\nu_k)) \quad (12)$$

where  $\nabla f_1$  is the derivative of  $f_1$ ,  $\rho$  is the gradient step length and  $\text{prox}_{f_2}$  is the so-called proximity operator associated with  $f_2$  [21]. In our setting, we have :

$$\nabla f_1(\nu_k) = -\gamma^k T (\alpha - \sum_{j \neq k} \gamma^j \nu_j - \gamma^k \nu_k) \quad (13)$$

$$\text{prox}_{\rho f_2}(\nu_k) = \Delta_{\eta_k}(\nu_k) \quad (14)$$

where  $\Delta_{\eta_k} \|\cdot\|_1$  is the soft-thresholding operator with threshold  $\eta_k = \rho \sigma^2 \lambda_k \|\gamma^k\|_1$ . In the context of hyperspectral data and assuming that the sought after components are morphologically different, their spectra are not too far from being mutually orthogonal. This means that the terms involving  $\gamma^k T \gamma^j$  with  $j \neq k$  can be neglected. The derivative of  $f_1$  can thus be approximated by :  $\nabla f_1(\nu_k) =$



$$-\gamma^{kT}(\alpha - \gamma^k \nu_k).$$

The last step consists in choosing the gradient step length  $\rho$ . We would like to recall that the gradient of  $f_1$  is Lipschitz with constant  $L$ . From the expression of its gradient in Equation (13), we get  $L = \|\gamma^k\|_2^2$ . From [21], convergence is guaranteed whenever  $\rho L \leq 1$ . We thus choose the largest gradient step length that guarantees convergence :  $\rho = \frac{1}{\|\gamma^k\|_2^2}$ . With this choice, the gradient descent step  $\nu_k - \rho \nabla f_1(\nu_k)$  can be rewritten as follows :

$$\nu_k - \rho \nabla f_1(\nu_k) = \frac{1}{\|\gamma^k\|_2^2} \gamma^{kT} \alpha$$

This leads to the following update for  $\nu_k$  :

$$\nu_k^{new} = \Delta_{\eta_k} \left( \frac{1}{\|\gamma^k\|_2^2} \gamma^{kT} \alpha \right) \quad (15)$$

where  $\eta_k = \frac{\sigma^2 \lambda_k \|\gamma^k\|_1}{\|\gamma^k\|_2^2}$ . We now define the vector  $\eta$  of size  $n$  with entries  $\eta[k] = \eta_k$ . Hence, the source matrix  $\nu$  is updated as follows :

$$\nu^{new} = \Delta_{\eta} \left( \text{Diag}(1/\|\gamma^k\|_2^2) \gamma^T \alpha \right)$$

where the multichannel soft-thresholding operator  $\Delta_{\eta}$  acts on each row  $k$  of  $\nu$  with threshold  $\eta[k]$ . Let us recall that we have made the assumption that the spectra are orthogonal, the diagonal matrix  $\text{Diag}(1/\|\gamma^k\|_2^2)$  is then an approximation for  $(\gamma^T \gamma)^{-1}$ . In practice, the spectra are not exactly orthogonal. In that case, better computational results are obtained by updating the source matrix  $\nu$  as follows :

$$\nu^{new} = \Delta_{\eta} \left( (\gamma^T \gamma)^{-1} \gamma^T \alpha \right) \quad (16)$$

Symmetrically, the parameter  $\gamma$  is updated as follows :

$$\gamma^{new} = \Delta_{\zeta} \left( \alpha \nu^T (\nu \nu^T)^{-1} \right)$$

where  $\zeta$  has length  $m$  and entries  $\zeta[k] = \frac{\sigma^2 \lambda_k \|\nu_k\|_1}{\|\nu_k\|_2^2}$ . The multichannel soft-thresholding operator  $\Delta_{\zeta}$  acts on each column  $k$  of  $\gamma$  with threshold  $\zeta[k]$ . Equations (9) and (10) are easily interpreted as thresholded alternate least squares solutions.

#### APPENDIX B POSITIVITY CONSTRAINT

Similarly to Appendix A, we rewrite Equation (8) in a simpler way. Assuming  $\gamma$  is fixed, we denote  $f_1(\nu) = \frac{1}{2} \|\alpha - \gamma \nu\|_2^2$  and  $f_2(\nu) = \sigma^2 \sum_{k=1}^n \lambda_k \|\gamma^k \nu_k\|_1$ . Then,  $\nu$  is updated by solving the following minimization problem :

$$\min_{\nu} f_1(\nu) + f_2(\nu)$$

where  $f_1$  is a differentiable quadratic function and  $f_2$  is a convex but nonsmooth function (it is not differentiable). In the framework of proximal calculus (see [21] and references therein), this optimization problem can be solved using the following fixed-point algorithm :

$$\nu^{new} = \text{prox}_{\rho f_2}(\nu - \rho \nabla f_1(\nu))$$

where  $\nabla f_1$  is the derivative of  $f_1$ ,  $\rho$  is the gradient steplength and  $\text{prox}_{f_2}$  is the so-called proximity operator associated with  $f_2$  [21]. In our setting, we have seen in Appendix A that  $f_2$  is the  $\ell_1$  norm and its proximity operator is the soft-thresholding operator.

In order to enforce the positivity of the sources  $\mathbf{S} = \nu \Phi$ ,  $\nu$  is updated by solving the following minimization problem :

$$\min_{\nu} f_1(\nu) + f_2(\nu) + i_C(\nu)$$

where  $i_C$  is the indicator function of the convex set  $C$  of all matrices  $Y$  such that  $Y \Phi$  has non-negative entries. We recall that the indicator

function  $i_C(Y)$  of convex set  $C$  is defined by [22] :

$$i_C(Y) = \begin{cases} 0 & \text{if } Y \in C \\ +\infty & \text{otherwise} \end{cases}$$

This function  $i_C$  is convex and admits a proximity operator. Hence the above minimization problem can be solved by using the same fixed-point algorithm [21] :

$$\nu^{new} = \text{prox}_{\rho(f_2+i_C)}(\nu - \rho \nabla f_1(\nu))$$

This projected gradient algorithm requires evaluating the proximity operator associated with  $f_2 + i_C$  which has in general no closed-form expression. This could be performed exactly by using an extension of the well-known alternate projections algorithm to the proximity operators [23]. This kind of algorithm requires alternating the application of  $\text{prox}_{f_2}$  and  $\text{prox}_{i_C}$ , which can be computationally expensive. Furthermore, in practice, it turns out that a single application of  $\text{prox}_{f_2}$  and then  $\text{prox}_{i_C}$  provides good enough numerical results. That is why, in *hypGMCA*, the components are estimated as follows :

$$\nu^{new} = \text{prox}_{\rho i_C} \left( \text{prox}_{\rho f_2}(\nu - \rho \nabla f_1(\nu)) \right)$$

By definition  $\text{prox}_{\rho i_C}(\nu) = \text{prox}_{i_C}(\nu) = \text{Argmin}_{z \in C} \|z - \nu\|_2$ . As  $\Phi$  is orthogonal,  $\text{prox}_{i_C}(\nu)$  can be equivalently written also follows :

$$\text{prox}_{i_C}(\nu) = \left( \text{Argmin}_{z \in K} \|z - \nu \Phi\|_2 \right) \Phi^T.$$

where  $z$  belongs to  $K$ , the set of matrices with non-negative entries (also known as non-negative orthant). This expression can be equivalently rewritten as follows :  $\text{prox}_{i_C}(\nu) = P_K(\nu_{1/2} \Phi) \Phi^T$  where  $P_K$  is the orthogonal projection onto  $K$ . Recall that this projection is defined as follows :

$$\forall i, j; (P_K(Y))[i, j] = \begin{cases} Y[i, j] & \text{if } Y[i, j] \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

To conclude, when the positivity of the sources is enforced,  $\nu$  is updated as follows :

- 1) Update  $\nu$  with no positivity constraint as described in Appendix A to get an intermediate estimate  $\nu_{1/2}$  of  $\nu$ .
- 2) Enforce the positivity of the sources :

$$\nu^{new} = P_K(\nu_{1/2} \Phi) \Phi^T$$

#### REFERENCES

- [1] J. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.
- [2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.
- [3] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation*, vol. 13, pp. 863–882, 2001.
- [4] J. Bobin, J.-L. Starck, M. J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662 – 2674, November 2007. [Online]. Available: <http://perso.orange.fr/jbobin/pubs2.html>
- [5] J. Bobin, Y. Moudden, M. J. Fadili, and J.-L. Starck, "Morphological diversity and sparsity for multichannel data restoration," *Journal of Mathematical Imaging and Vision*, vol. 33, no. 2, pp. 149–168, 2008.
- [6] 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, pp. 2194–2208, 2008.
- [7] J. Nascimento and J. 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.
- [8] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, pp. 1089–1092, 2009.

- [9] D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via  $\ell^1$  minimization," *Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, 2003.
- [10] Z. Zhang, H. Zha, and H. Simon, "Low-rank approximations with sparse factors I: Basic algorithms and error analysis," *SIAM Journal on Matrix Analysis and Applications*, vol. 23, no. 3, pp. 706–727, 2002. [Online]. Available: [citeseer.ist.psu.edu/624345.html](http://citeseer.ist.psu.edu/624345.html)
- [11] J. Stone, J. Porrill, N. Porter, and I. Wilkinson, "Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions," *NeuroImage*, vol. 15, no. 2, pp. 407–421, 2002.
- [12] A. Hyvarinen and R. Karthikes, "Imposing sparsity on the mixing matrix in independent component analysis," *Neurocomputing*, vol. 49, pp. 151–162, 2002.
- [13] F. Theis, P. Gruber, I. Keck, A. Meyer-Base, and E. Lang, "Spatiotemporal blind source separation using double-sided approximate joint diagonalization," in *In Proc. EUSIPCO 2005*, 2005.
- [14] R. Rubinstein, M. Zibulevsky, and M. Elad, "Learning sparse dictionaries for sparse signal representations," *IEEE Transactions on Signal Processing*, 2008.
- [15] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361–379, 2000.
- [16] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, Aug 2004.
- [17] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for  $\ell_1$ -regularized minimization with applications to compressed sensing," Rice University, Tech. Rep., July 2007.
- [18] M. A. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [19] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5559–5569, 2006.
- [20] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.
- [21] P. L. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [22] R. T. Rockafellar, *Convex analysis*, ser. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1970.
- [23] H. Bauschke and P. L. Combettes, "A dykstra-like algorithm for two monotone operators," *Pacific Journal of Optimization*, vol. 4, no. 3, pp. 383–391, Sept. 2008.

PLACE  
PHOTO  
HERE

**Jerome Bobin** graduated from the Ecole Normale Supérieure (ENS) de Cachan, France, in 2005 and received the M.Sc. degree in signal and image processing from ENS Cachan and Paris-Sud University. He received the Agrégation de Physique in 2004 and a Ph.D. in Electrical Engineering in 2008 under the supervision of J.-L. Starck at the CEA. From 2008 to 2009, he has been a post-doctoral scholar in Applied Mathematics at Caltech. Since 2009, he is a post-doctoral scholar in the mathematics department at Stanford University in collaboration with Pr. Candes. His research interests include statistical signal processing, multiscale methods and sparse representations in signal and image processing and their applications in astrophysics.

PLACE  
PHOTO  
HERE

**Yassir Moudén** graduated in electrical engineering from SUPELEC, Gif-sur-Yvette, France and obtained the M.Sc. degree in physics from the Université de Paris VII, France, in 1997. He received a Ph. D. degree in signal processing from the Université de Paris-Sud, Orsay, France. He is currently with the CEA, Saclay, France, working on data acquisition and realtime data processing systems for large experiments in physics. His research interests include signal and image processing, data analysis as well as FPGAs and other realtime technology.