HAL
open science

Sparsity and morphological diversity for multivalued data analysis

Jérome Bobin, Yassir Moudden, Jean-Luc Starck, Jalal M. Fadili

To cite this version:

Jérome Bobin, Yassir Moudden, Jean-Luc Starck, Jalal M. Fadili. Sparsity and morphological diversity for multivalued data analysis. SPIE Wavelets XIII, Aug 2009, San Diego, United States. pp.7446-48. hal-00813862

HAL Id: hal-00813862

https://hal.science/hal-00813862v1

Submitted on 16 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sparsity constraints for hyperspectral data analysis : linear mixture model and beyond

J. Bobin[a], Y. Moudden [b], J.-L. Starck [b] and J. Fadili [c],

[a] Department of Applied & Computational Mathematics
California Institute of Technology, Pasadena, California 91125, USA
[b] DAPNIA/SEDI-SAP, Service d'Astrophysique,
CEA/Saclay, 91191 Gif sur Yvette, France
[c] GREYC CNRS UMR 6072, Image Processing Group, ENSICAEN 14050, Caen Cedex, France

Keywords: Morphological diversity, sparsity, overcomplete representation, curvelets, wavelets, multichannel data, blind source separation, denoising, inpainting, multichannel convolution

ABSTRACT

The recent development of multi-channel sensors has motivated interest in devising new methods for the coherent processing of multivariate data. An extensive work has already been dedicated to multivariate data processing ranging from blind source separation (BSS) to multi/hyper-spectral data restoration. Previous work[1] has emphasized on the fundamental role played by sparsity and morphological diversity to enhance multichannel signal processing.

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. Inspired by GMCA, a recently introduced algorithm coined *Hyp*GMCA is described for BSS applications in *hyperspectral* data processing. It assumes the collected data is a linear instantaneous 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.

v

Introduction

Generalized Morphological Component Analysis (GMCA) is a recent algorithm for multivariate data analysis introduced in a previous paper.[2] It was applied successfully in a variety of multichannel data processing applications, including BSS, color image restoration and inpainting.[2,3] Given a data matrix $X \in \mathbb{R}^{m,t}$ and a number of components n, GMCA is used to decompose X into a sum of n rank one contributions X_k with different statistical and *spatio-spectral* properties. Each matrix 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 ν_k in a given dictionary of spatial waveforms $\mathbf{\Phi} \in \mathbb{R}^{t,t'}$, which for simplicity we take to be the same for all k. In matrix form, we write :

$$\mathbf{X} = \sum_k \mathbf{X}_k + \mathbf{N} = \sum_k a^k s_k + \mathbf{N} \quad (1)$$

$$= \mathbf{AS} + \mathbf{N} = \sum_k a^k \nu_k \mathbf{\Phi} + \mathbf{N} \quad (2)$$

where the kth line of $\mathbf{S} \in \mathbb{R}^{n,t}$ is s_k and the kth 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 σ^2. In the case of multichannel image data, the image from

Further author information: Send correspondence to bobin@acm.caltech.edu

the pth channel is formally represented here as the pth line of $\mathbf{X}$, x_p. The importance of sparsity in blind source separation was recently recognized.[4] 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 the latter marginal property of the sparse representations ν_k, GMCA requires morphological diversity to achieve its decomposition which is a property of their joint distribution. Let ν_k be the kth line of matrix $\boldsymbol{\nu} \in \mathbb{R}^{n,t'}$. The latter property expresses the assumption that there is little probability that a column of $\boldsymbol{\nu}$ 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.

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* that the mixed underlying objects of interest $\mathbf{X}_k = a^k s_k$ exhibit both sparse spectral signatures and sparse spatial morphologies in known dictionaries of spectral and spatial waveforms. The proposed algorithm, referred to as *hyp*GMCA was devised to account for the additional *a priori* sparsity constraint on the mixing matrix *i.e.* to enforce that columns a^k have a sparse representation in $\boldsymbol{\Psi}$, a given dictionary of spectral waveforms.

In the next section, we discuss and build a modified MAP objective function which formalizes the desired spatio-spectral sparsity constraint. The resulting *hyp*GMCA algorithm is given in section 1.1. In section 1.2, numerical experiments with synthetic and real hyperspectral data illustrate the efficiency of the proposed algorithm.

In Section 2, we extend *Hyp*GMCA to handle a wider range of mixture models beyond the celebrated instantaneous mixture model. Based upon *Hyp*GMCA we introduce a new algortihm that can solve BSS problems account for missing data or multichannel convolution models. Promising preliminary results are shown in Section 2.3.

1. BLIND SOURCE SEPARATION FOR HYPERSPECTRAL DATA

With the above assumptions, equation (1) is rewritten as follows :

$$\mathbf{X} = \sum_k \mathbf{X}_k + \mathbf{N} = \sum_k \boldsymbol{\Psi} \gamma^k \nu_k \boldsymbol{\Phi} + \mathbf{N} \tag{3}$$

where $\mathbf{X}_k = a^k s_k$ are rank one matrices sparse in $\boldsymbol{\Omega} = \boldsymbol{\Psi} \otimes \boldsymbol{\Phi}$ * such that a^k has a sparse representation γ^k in $\boldsymbol{\Psi}$ while s_k has a sparse representation ν_k in $\boldsymbol{\Phi}$. Denote $\alpha_k = \gamma^k \nu_k$ the rank one matrix of coefficients representing $\mathbf{X}_k$ in $\boldsymbol{\Omega}$.

*where $\otimes$ stands for the tensor product of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$

Initially, the objective of the GMCA algorithm is as follows :

$$\min_{\mathbf{A},\mathbf{S}} \sum_k \lambda_k \|\nu_k\|_1 + \frac{1}{2\sigma^2} \left\| \mathbf{X} - \sum_k a^k s_k \right\|_2^2 \quad \text{with} \;\; s_k = \nu_k \mathbf{\Phi} \tag{4}$$

which is derived as a MAP estimation of the model parameters $\mathbf{A}$ and $\mathbf{S}$ where the ℓ_1 penalty terms imposing sparsity come from Laplacian priors on the sparse representation ν_k of s_k in $\mathbf{\Phi}$. Interestingly, the treatment of $\mathbf{A}$ and $\mathbf{S}$ in the above is asymmetric. This is a common feature of the great majority of BSS methods which invoke a uniform *improper* prior distribution for the spectral parameters $\mathbf{A}$. Truly, $\mathbf{A}$ and $\mathbf{S}$ often have different roles in the model and very different sizes. However, dealing with so-called hyperspectral data, assuming that the spectral signatures a^k also have sparse representations γ^k in spectral dictionary $\mathbf{\Psi}$, this asymmetry is no longer so obvious. Also, a well known property of the linear mixture model (1) is its *scale and permutation invariance* : without additional prior information, the indexing of the $\mathbf{X}_k$ in the decomposition of data $\mathbf{X}$ is not meaningful and a^k, s_k can trade a scale factor in full impunity. A consequence is that unless *a priori* specified otherwise, information on the separate scales of a^k and s_k is lost due to the multiplicative mixing, and only a joint scale parameter for a^k, s_k can be estimated. This loss of information needs to be translated into a *practical* prior on $\mathbf{X}_k = a^k s_k = \mathbf{\Psi}\gamma^k \nu_k \mathbf{\Phi}$. Unfortunately, deriving the distribution of the product of two independent random variables γ^k and ν_k based on their marginal densities can be cumbersome. We propose instead that the following p_π is a good and practical candidate *joint sparse prior* for γ^k and ν_k after the loss of information induced by multiplication :

$$p_\pi(\gamma^k, \nu_k) \propto \exp(-\lambda_k \|\gamma^k \nu_k\|_1) \propto \exp(-\lambda_k \sum_{i,j} |\gamma_i^k \nu_k^j|) \tag{5}$$

where γ_i^k is the ith entry in γ^k and ν_k^j is the jth entry in ν_k. Note that the proposed distribution has the nice property, for subsequent derivations, that the conditional distributions of γ^k given ν_k and of ν_k given γ^k are both Laplacian distributions which are commonly and conveniently used to model sparse distributions. Finally, inserting the latter prior distribution in a Bayesian MAP estimator leads to the following minimization problem :

$$\min_{\{\gamma^k, \nu_k\}} \frac{1}{2\sigma^2} \left\| \mathbf{X} - \sum_k \mathbf{\Psi}\gamma^k \nu_k \mathbf{\Phi} \right\|_2^2 + \sum_k \lambda_k \|\gamma^k \nu_k\|_1 \tag{6}$$

Let us first note that the above can be expressed slightly differently as follows :

$$\begin{gathered} \min_{\{\alpha_k\}} \tfrac{1}{2\sigma^2} \|\mathbf{X} - \textstyle\sum_k \mathbf{X}_k\|_2^2 + \textstyle\sum_k \lambda_k \|\alpha_k\|_1 \\ \text{with } \mathbf{X}_k = \mathbf{\Psi}\alpha_k \mathbf{\Phi} \text{ and } \forall k, \text{rank}(\mathbf{X}_k) \leq 1 \end{gathered} \tag{7}$$

which uncovers a nice interpretation of our problem as that of approximating the data $\mathbf{X}$ by a sum of rank one matrices $\mathbf{X}_k$ which are sparse in the specified dictionary of rank one matrices. This is the usual ℓ_1 minimization problem[5] but with the additional constraint that the $\mathbf{X}_k$ are all rank one at most. The latter constraint is enforced here mechanically through a proper parametric representation of $\mathbf{X}_k = a^k s_k$ or $\alpha_k = \gamma^k \nu_k$. A similar problem was previously investigated[6] with a very different approach.

There have been previous reports of a symmetric treatment of $\mathbf{A}$ and $\mathbf{S}$ for BSS[7–9] however in the noiseless case. We also note that very recently, the objective function (6) was proposed[10] for dictionary learning oriented applications. However, the algorithm derived by M.Elad[10] is very different from the method proposed here which benefits from all the good properties of GMCA, notably its speed and robustness which come along the iterative thresholding with a decreasing threshold.

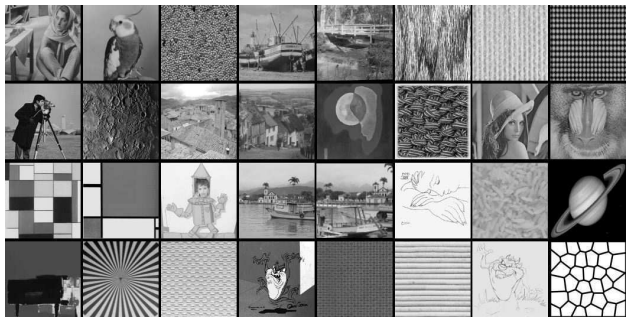

Figure 1. Image data set used in the experiments. Each image contains 128 by 128 pixels. They all have zero mean and are normalized to have unit variance.

1.1. GMCA Algorithm for *hyperspectral* data

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

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

where the columns of γ are γ^k, the rows of ν are ν_k and $\alpha = \mathbf{\Psi}^T \mathbf{X} \mathbf{\Phi}^T$ is the coefficient matrix of data $\mathbf{X}$ in $\mathbf{\Omega}$. Thus, we are seeking a decomposition of matrix α 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 iterative algorithm, alternately minimizing with respect to γ and ν. Indeed, thanks to the chosen prior, for fixed γ (resp. ν), the *marginal* minimization problem over ν (resp. γ) is convex and is readily solved using a variety of methods. Inspired by the iterative thresholding methods,[11–13] akin to Projected Landweber algorithms, we obtain the following system of update rules :

$$\left\{ \begin{array}{rcl} \nu^{(+)} & = & \Delta_\eta \left(\nu^{(-)} + \boldsymbol{R}_\nu \left(\alpha - \gamma \nu^{(-)} \right) \right) \\ \gamma^{(+)} & = & \Delta_\zeta \left(\gamma^{(-)} + \left(\alpha - \gamma^{(-)} \nu \right) \boldsymbol{R}_\gamma \right) \end{array} \right. \tag{9}$$

where $\boldsymbol{R}_\nu$ and $\boldsymbol{R}_\gamma$ are appropriate relaxation matrices for the iterations to be non-expansive. Assume left invertibility of $\mathbf{A}$ and right invertibility of $\mathbf{S}$. Then, taking $\boldsymbol{R}_\nu = \left(\gamma^T \gamma\right)^{-1} \gamma^T$ and $\boldsymbol{R}_\gamma = \nu^T \left(\nu \nu^T\right)^{-1}$, the above are rewritten as follows :

$$\nu^{(+)} = \Delta_\eta \left(\left(\gamma^T \gamma\right)^{-1} \gamma^T \alpha \right) \tag{10}$$

$$\gamma^{(+)} = \Delta_\zeta \left(\alpha \nu^T \left(\nu \nu^T\right)^{-1} \right) \tag{11}$$

where vector η has length n and entries $\eta[k] = \lambda_k \|\gamma^k\|_1 / \|\gamma^k\|_2^2$, while ζ has length m and entries $\zeta[k] = \lambda_k \|\nu_k\|_1 / \|\nu_k\|_2^2$. The multichannel soft-thresholding operator Δ_η acts on each row k of ν with threshold $\eta[k]$ and Δ_ζ acts on each column k of γ with threshold $\zeta[k]$. Equations (10) and (11) are easily interpreted as thresholded alternate least squares solutions.

Finally, in the spirit of the fast GMCA algorithm,[2] 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 *hyp*GMCA :

1. Set the number of iterations $I_{\max}$ and initial thresholds $\lambda_k^{(0)}$
2. Transform the data $\mathbf{X}$ into α
3. While $\lambda_k^{(h)}$ are higher than a given lower bound $\lambda_{\min}$,
   - Update ν assuming γ is fixed using equation (10).
   - Update γ assuming ν is fixed using equation (11) .
   - Decrease the thresholds $\lambda_k^{(h)}$.
5. Transform back γ and ν to estimate $\mathbf{A}$ and $\mathbf{S}$.

With the threshold successively decaying towards zero along iterations, the current sparse approximations for γ and ν are progressively refined by including finer structures spatially and spectrally, alternatingly. This *salient to fine* estimation process is the core of *hyp*GMCA. This optimization technique has been used with success in GMCA[2]; decreasing the value of the threshold is likely to refrain the algorithm from falling into spurious local minima which makes it similar to *simulated annealing* techniques. 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 ℓ_1 sparsity measure, which comes as an approximation to the ℓ_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 a previous paper.[3,?] When non-unitary or redundant transforms are used, the above is no longer strictly valid. Nevertheless, simple shrinkage still gives satisfactory results in practice.[14] In the end, implementing the proposed update rules requires only a slight modification of the GMCA algorithm given in a previous paper.[3] 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 $\mathbf{\Omega}$ is the union of several orthonormal bases $\mathbf{\Omega}_k$ 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.

1.2. Numerical experiments

In this section, we compare the performance of *hyp*GMCA and GMCA in toy BSS experiments with 1D and 2D data. 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 128×128 images shown on Figure 1. These images provide us with 2D spatially structured processes which are sparse enough in the curvelet domain.[15] 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 σ^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 $\mathcal{C}_{\mathbf{A}} = \|\mathbf{I}_n - \mathbf{P}\tilde{\mathbf{A}}^{\dagger}\mathbf{A}\|_1$, 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 $\mathcal{C}_{\mathbf{A}}$ 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 *hyp*GMCA 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[15] with the same number of iterations.

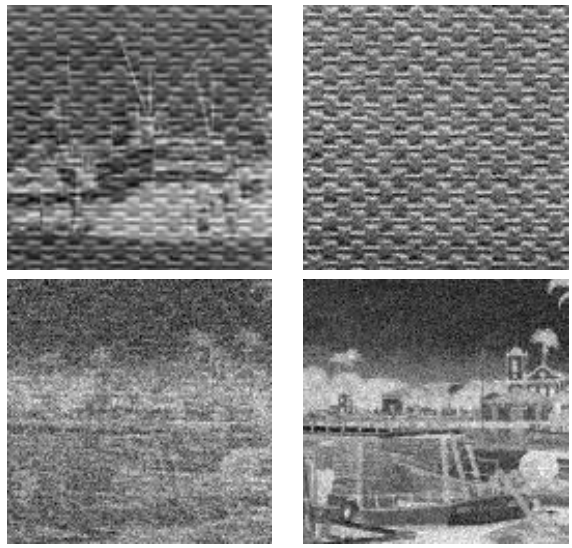

Figure 2. Four 128×128 mixtures out of the 128 channels. The SNR is equal to 20dB.

1.3. Decomposition of Mars hyperspectral data

In this section, we illustrate the good behavior of *hyp*GMCA 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*), at wavelengths ranging from $0.93\mu m$ to 2.73 μm 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 5. Model 3 is clearly too simple to describe this hyperspectral reflectance data set. Non-linear instrumental and atmospheric effects are most likely to contribute to the *true* generative process. In any case, following a similar decision made in a previous study,[16] we use *hyp*GMCA 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 *hyp*GMCA algorithm, this constraint is enforced by projection of the estimated source maps $\mathbf{S}$ on the cone generated by the vectors with positive entries. The number of iterations is $I_{max} = 250$.
Amongst the $n = 10$ estimated sources only two are strongly correlated with the H_2O and CO_2 ice reference spectra. Figure 6 displays their spectral signatures compared to the reference spectra, and Figure 7 shows the corresponding spatial density (positive) maps. Owing to the non-linearity of the physical mixture process, the close fit between the estimated and reference CO_2 spectra is satisfactory. The H_2O ice spectrum is remarkably similar to the reference spectra for wavelengths higher than $1\mu m$. As also noted in a previous study,[16] the CO_2 ice appears located in large regions around the pole of planet Mars, while H_2O ice seems to lie be concentrated in some tight interstices of the Mars surface. In the end, despite the simple linear mixture model we used, *hyp*GMCA is able to extract components with spectral signatures that closely match reference spectra.

2. BEYOND THE LINEAR MIXTURE MODEL

2.1. More realistic mixture models

In the previous sections, we have assumed that the linear mixture model

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

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

holds. Nevertheless, when dealing with real-world data, this model is seldom valid. A more appropriate model of the data should take into account some observational process $\mathcal{H}$; beyond the standard linear mixture model $\mathbf{X} = \mathbf{AS}$, the data are rather modelled as a function of $\mathbf{A}$ and $\mathbf{S}$

$$\mathbf{X} = \mathcal{H}\left(\mathbf{A}, \mathbf{S}\right) + \mathbf{N} \tag{12}$$

Let us define $\mathbf{X^0} = \mathbf{AS}$. $\mathbf{X^0}$ are the data that would be measured if the model were an instantaneous linear mixture. The need for more sophisticated models is motivated by challenges in multi/hyperspectral analysis problems that include :

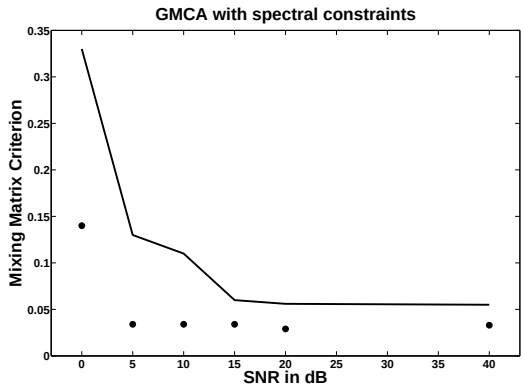

Figure 4. Evolution of the mixing matrix criterion $\mathcal{C}_{\mathbf{A}}$ as a function of the SNR in dB. *Solid line :* recovery results with GMCA. • *:* recovery results with hypGMCA.

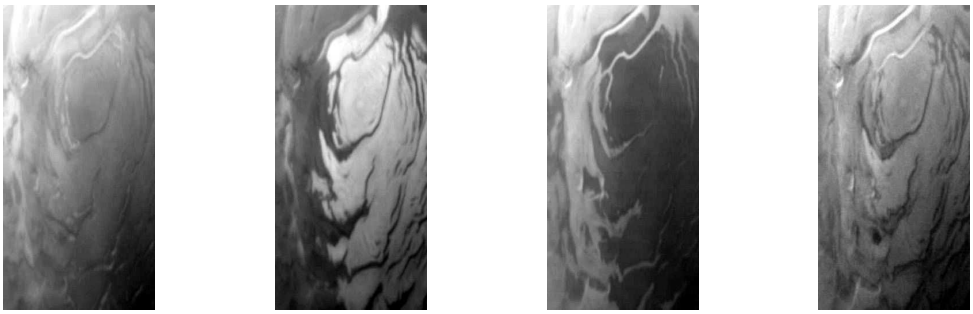

Figure 5. From left to right : Mars Express observations at wavelengths = 1.38 - 1.75 - 1.94 and 2.41 μm.

- **Missing data :** sensors are seldom perfect and some missing pixels may provide no information. In that case, the data $\mathbf{X}$ can be modeled as a linear mixture model in which pixels are missing

$$\begin{aligned} \mathbf{X} &= \mathcal{M} \odot \left(\mathbf{X^0}\right) + \mathbf{N} && (13) \\ &= \mathcal{M} \odot (\mathbf{AS}) + \mathbf{N} && (14) \end{aligned}$$

  where M is a $m \times t$ binary mask such that

$$\mathcal{M}[i,j] = \begin{cases} 1 \text{ if the pixel } (i,j) \text{ is observed} \\ 0 \text{ otherwise} \end{cases}$$

  The operation $\odot$ stands for the entrywise product of matrices of size $m \times t$.

- **Multichannel convolution :** multi/hyperspectral data are often mesured by different sensors in an instrument having a fixed aperture. As stated by standard laws in optics, a given observation has a point spread function (PSF) that depends on the wavelength at which it is observed. Then, each datum $\{x_i\}_{i=1,\cdots,m}$ has a specific PSF $\{\mathbf{H}_i\}_{i=1,\cdots,m}$:

$$\forall i \in \{1, \cdots, m\}; \quad x_i = \mathbf{H}_i \star x_i^0 + n_i \qquad (15)$$

  This so-called *multichannel convolution* model would be very helpful when dealing with real-world data as provided by space telescopes such as the ESA space mission Planck[†].

[†]See *http://planck.esa.int*

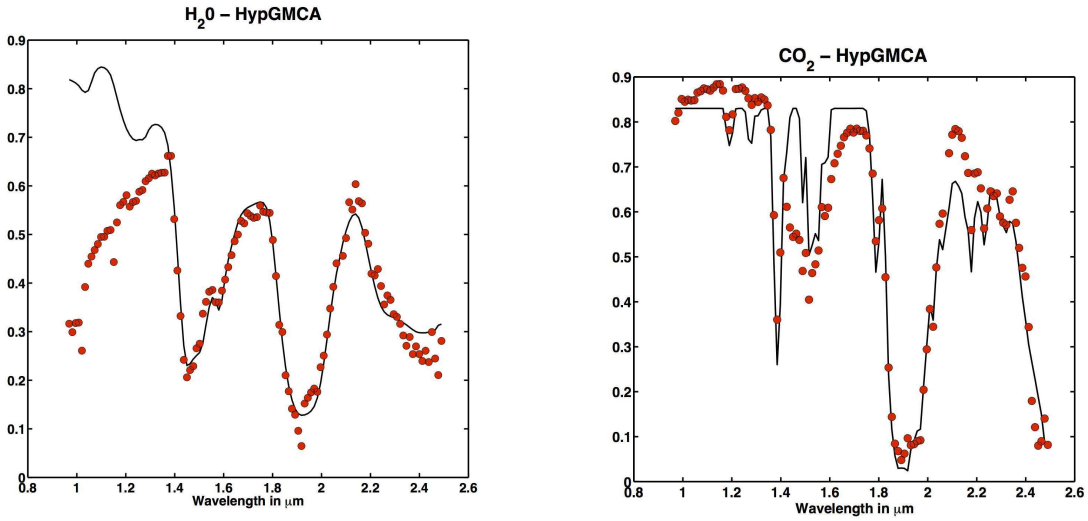

Figure 6. Left picture : Reference (solid line) and estimated (•) spectra for H_2O ice. **Right picture :** Reference (solid line) and estimated (•) spectra for CO_2 ice.

Figure 7. Estimated spatial concentration maps of H_2O ice (**left**) and CO_2 ice (**right**).

In the above multichannel models (the list is obviously not exhaustive), the observation mapping $\mathcal{H}$ does not preserve the structure of the instantaneous linear mixture model $\mathbf{X^0} = \mathbf{AS}$. Retrieving the mixing matrix $\mathbf{A}$ and the sources $\mathbf{S}$ can't be performed by simply applying GMCA or *Hyp*GMCA. When solving inverse problems, the direct problem must be properly taken into account to estimate the parameters $\mathbf{A}$, $\mathbf{S}$. We proposed[17] a first approach to extend GMCA to solve BSS problems involving a multichannel convolution. Based upon GMCA, this preliminary approach was designed to account for the sparsity of the sources. In this paper, we introduce a new algorithm that enables to account for realistic physical prior such as

- spatio-spectral sparsity constraints such as in *Hyp*GMCA
- positivity of the mixing matrix and the sources

In the remaining of this section we will put more emphasis on the multichannel convolutive model[‡]. To illustrate how strenuous this problem is, let's consider the following synthetic example:

- The data $\mathbf{X}$ are composed of $m = 32$ channels each of which being an image of size 256×256.
- The number of sources is $n = 3$ and their entries are positive as displayed in Figure 10.
- The mixing matrix of size 32×3 and its entries are positive as shown in Figure 11.
- the noise is white and Gaussian with standard deviation 2 (which corresponds, in the experiment, to a SNR in each channel ranging from 17 to 42 dB).
- the convolution kernels $\{\mathbf{H}_i\}_{i=1,\cdots,m}$ are Gaussian .

The complexity of this *non-linear*[§] blind source is twofold : i) the noise makes it hard to deconvolve each channel before applying *Hyp*GMCA and ii) the recoverability of each source will highly depend on its weight (mixing matrix) in each observation. To give an illustration of this point, let us consider Figure 8.

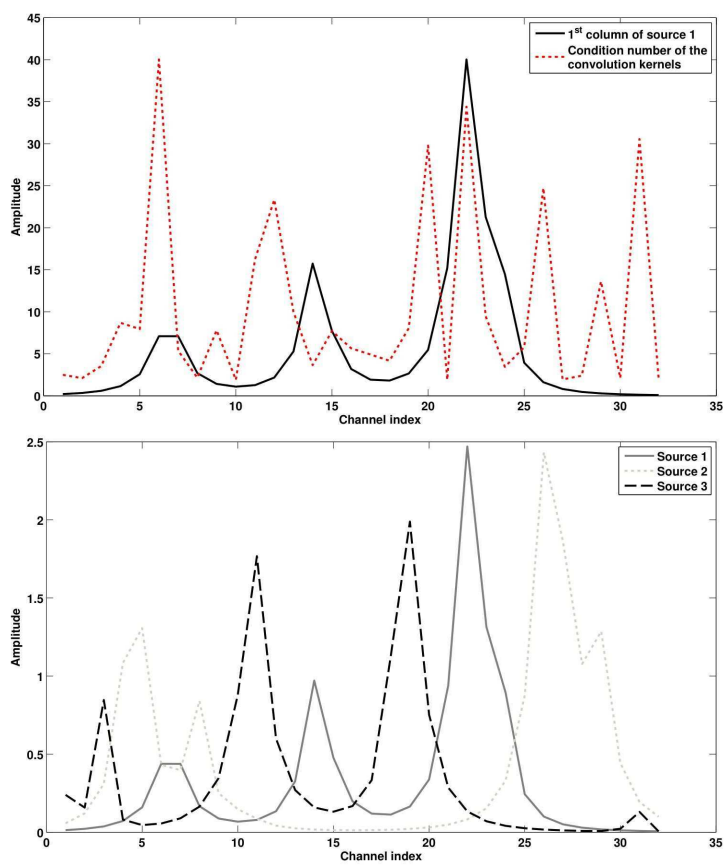

Figure 8. Top picture : Spectrum related to source 1 superimposed with the condition number of each convolution kernel $\{H_i\}_{i=1,\cdots,m}$. **Bottom picture :** Spectra $\{a^i\}_{1,\cdots,m}$.

[‡]Which is different form the convolutive mixing model which is popular in the audio separation community.

[§]Here *non-linearity* means that the linear observation mapping $\mathcal{H}$ does not preserve the linearity of the mixture model

The top picture of Figure 8 shows the spectrum of source 1 (normalized for a better readibility) and the condition number of each convolution kernel $\{\mathbf{H}_i\}_{i=1,\cdots,m}$ which gives a measure of the deterioration caused by convolution on each channel. Clearly, the recoverability of a given source will be higher when i) the sources have different contributions in each channel (which is reasonable if one assumes that the sources have different spectral morphologies) and ii) if the channels in which the source has high contributions are only slightly deteriorated by the convolution (*i.e.* low condition number). We can remark in the picture at the bottom of Figure 8 that the sources have morphologically different spectra. Furthermore, source 1 is mainly active in channel 22; the condition number $\mathbf{H}_{22}$ is equal to 34.4 which tends to make the recovery of source 1 difficult. For the sake of concision, a thorough analysis of the multichannel convolutive model will be presented in a forthcoming paper.
We would like to point out that, to our knowledge, there exists no other methods solving joint blind source separation and deconvolution.

2.2. Extending *Hyp*GMCA

In the next, we will make the following assumptions

- the sources $\mathbf{S}$ are sparse in $\mathbf{\Phi}$ and have non negative entries.
- the mixing matrix $\mathbf{A}$ has sparse columns in $\mathbf{\Psi}$ and non negative entries.

Following the framework of *Hyp*GMCA , the mixing matrix $\mathbf{A}$ and the sources $\mathbf{S}$ are estimated according to a MAP estimation as follows

$$\min_{\{a^i\geq 0, s_i\geq 0\}_{i=1,\cdots,n}} \sum_i \lambda_i \|\mathbf{\Psi}^T a^i s_i \mathbf{\Phi}^T\|_{\ell_1} + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2 \tag{16}$$

where $u \geq 0$ means that the entries of the vector u are non negative. Based upon *Hyp*GMCA , the parameters $\mathbf{A}$ and $\mathbf{S}$ are estimated iteratively and alternately via a *Sparse coding / Dictionary learning* scheme[18] such that

- **Sparse coding :** the sources are estimated assuming that the mixing matrix $\mathbf{A}$ is known

$$\min_{\{s_i\geq 0\}_{i=1,\cdots,n}} \sum_i \lambda_i \|\mathbf{\Psi}^T a^i s_i \mathbf{\Phi}^T\|_{\ell_1} + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2 \tag{17}$$

  Assuming that $\mathcal{H}$ is bilinear (which is the case for the aforementioned examples), the model is linear and the problem above is convex. Let us define $\mathbf{\Pi_A}$ the diagonal matrix whose diagonal is the vector $[\lambda_1\|\mathbf{\Psi}^T a^1\|_{\ell_1}, \cdots, \lambda_m\|\mathbf{\Psi}^T a^m\|_{\ell_1}]$. Then the problem in Equation (17) can be recast as follows

$$\min_{\mathbf{S}\geq\mathbf{0}} \|\mathbf{\Pi_A}\mathbf{S}\mathbf{\Phi}^T\|_1 + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2 \tag{18}$$

  where $\|\,.\,\|_1$ is the matrix $(1,1)$-norm. It can be solved efficiently using state-of-the-art fast algorithm for ℓ_1 minimizations.[19]

- **Dictionary learning :** the mixing matrix is estimated while fixing the sources $\mathbf{S}$

$$\min_{\{a^i\geq 0\}_{i=1,\cdots,n}} \sum_i \lambda_i \|\mathbf{\Psi}^T a^i s_i \mathbf{\Phi}^T\|_{\ell_1} + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2 \tag{19}$$

  Let $\mathbf{\Pi_S}$ stand for the diagonal matrix whose diagonal is the vector $[\lambda_1\|s_1\mathbf{\Phi}^T\|_{\ell_1}, \cdots, \lambda_m\|s_m\mathbf{\Phi}^T\|_{\ell_1}]$. Then, updating $\mathbf{A}$ is equivalently made as follows

$$\min_{\mathbf{A}\geq\mathbf{0}} \|\mathbf{\Psi}^T\mathbf{A}\mathbf{\Pi_S}\|_1 + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2 \tag{20}$$

Similarly to the problem in Equation (17), this problem can be solved efficiently using recent fast algorithm such as NESTA.[19] Let us remark the symmetry of the proposed algorithms : $\mathbf{A}$ and $\mathbf{S}$ play similar roles in the hyperspectral data case.

The two-step algorithm is then organized as follows

1. Set the initials values $\mathbf{S}^{(0)}$ and $\mathbf{A}^{(0)}$, the number of iterations $I_{\max}$ and initial thresholds $\lambda_k^{(0)}$
2. While $\lambda_k^{(h)}$ are higher than a given lower bound $\lambda_{\min}$,
   - Update the sources $\mathbf{S}$ by solving
     $\min_{\mathbf{S}\geq\mathbf{0}} \sum_i \lambda_i \|\mathbf{\Psi}^T a^i s_i \mathbf{\Phi}^T\|_{\ell_1} + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2$
   - Update the sources $\mathbf{A}$ by solving.
     $\min_{\mathbf{A}\geq\mathbf{0}} \sum_i \lambda_i \|\mathbf{\Psi}^T a^i s_i \mathbf{\Phi}^T\|_{\ell_1} + \frac{1}{2}\|\mathbf{X} - \mathcal{H}(\mathbf{A},\mathbf{S})\|_F^2$
   - Decrease the thresholds $\lambda_k^{(h)}$.

Similarly to GMCA and *Hyp*GMCA , the decreasing thresholds λ_k plays an important role : the parameters $\mathbf{A}$ and $\mathbf{S}$ are first estimated from the most significant features of each sources; it has been shown[2] that this process makes GMCA and *Hyp*GMCA more robust to spurious local minima. Moreover, in the spirit of continuation techniques,[12] solving ℓ_1-regularized least-square problems is computationally faster when the multiplier of the ℓ_1 term is high. Starting from high values of λ_k tends to accelerate the convergence of the algorithm as each iteration (h) provides good starting points $\mathbf{S}^{(h)}$ and $\mathbf{A}^{(h)}$ for the next step where $\lambda_k^{(h)} > \lambda_k^{(h+1)}$.

2.3. Preliminary recovery results

In this section, we present some preliminary numerical results obtained with synthetic data. The problem set up has been described in Section 2.1; the data are composed of $m = 32$ channels; each observation is an image of size 256×256 (see Figure 9); choosing $\mathbf{\Phi}$ as an undecimated wavelet frame is appropriate. The spectra are chosen to be spiky positive signals (bottom picture of Figure 8); assuming that the columns of the matrix $\mathbf{A}$ are sparse in a 1D orthogonal wavelet basis $\mathbf{\Psi}$ is then reasonable. The convolution kernels $\mathbf{H}_i$ are Gaussian-shaped with random width. White Gaussian noise with standard deviation $\sigma = 2$ is added (which corresponds, in the experiment, to a SNR in each channel that ranges from 17 to 42 dB).

Figure 9. Left picture : Observation x_7. **Right picture :** Observation x_{22}.

We used NESTA[19] to solve each subproblem (namely sparse coding and dictionary learning). Assuming that the norm of each source is approximately the same, the thresholds $\{\lambda_i\}_{i=1,\cdots,n}$ are chosen to be the same and $\lambda_{min} = 3\sigma$. In more realistic applications, these parameters should be estimated; this point will be studied in a future work. The number of iterations needed for convergence is $I_{\max} = 50$. The original and estimated sources are shown in Figure 10.

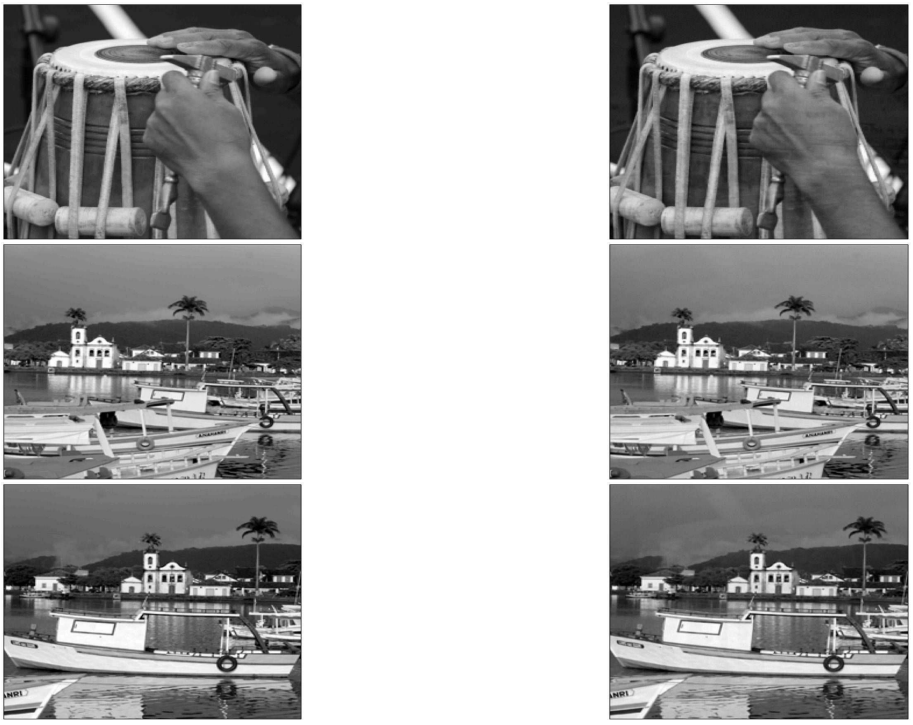

Figure 10. Left pictures : Original sources. **Right pictures :** Estimated sources.

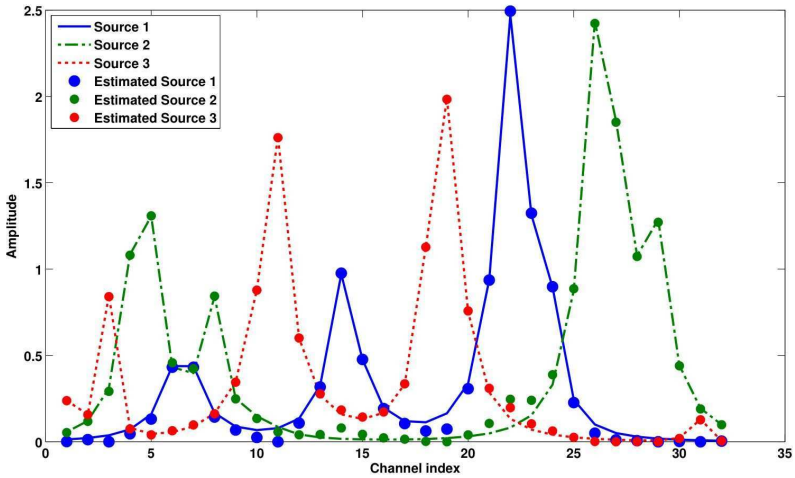

Figure 11. Original and estimated spectra.

Having in mind that solving such a problem is strenuous, the results are visually reasonable. The spectra, shown in Figure 11, are also well estimated. For this particular experiment, the mixing matrix criterion is equal to 0.0283. In the hyperspectral case, recovering each data cube $\{a^i s_i\}_{i=1,\cdots,n}$ with accuracy is important. Then measuring the recovery SNR of each of these cubes is a good measure of the quality of the separation. In that case, the mean value of the recovery SNR of these cubes is equal to 21.9 dB which confirms the good visual impression.

We previously[20] shedded on the GMCA's ability to solve multichannel restoration problems in an adaptive manner. Alongside the source separation problem, *Hyp*GMCA provides also a good way to restore

multichannel data as shown in Figure 12.B. This preliminary experiment provides encouraging results showing that the HGMCA framework may be an appropriate tool to analyze a potentially wide range of multi/hyperspectral data.

Figure 12. First column : Ideal observations x_{31}^0 (no convolution). **Column in the middle :** Noisy convolved observation x_{31}. **Left column :** Restored observation $\tilde{x}_{31}$.

3. CONCLUSION

We described a new algorithm, *hyp*GMCA, 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 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. Numerical experiments focus on a comparison between GMCA and *hyp*GMCA. GMCA is compared to state-of-the-art BSS algorithms.[2,3] We also present the first extension of *Hyp*GMCA to solve BSS problems involving observation mappings such as missing data or multichannel convolution models. Preliminary numerical results are given which show promising performances. Future work will focus on improving these results and to apply the new method to real-world data.

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

REFERENCES

1. J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Processing Letters* **13**(7), pp. 409–412, 2006.

2. J. Bobin, J.-L. Starck, M. J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing* **16**, pp. 2662 – 2674, November 2007.
3. 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* **33**(2), pp. 149–168, 2008.
4. M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computation* **13**, pp. 863–882, 2001.
5. D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Aca. Sci.* **100**, pp. 2197–2202, 2003.
6. 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* **23**(3), pp. 706–727, 2002.
7. J. Stone, J. Porrill, N. Porter, and I. Wilkinson, "patiotemporal independent component analysis of event-related fmri data using skewed probability density functions," *NeuroImage* **15**(2), pp. 407–421, 2002.
8. A. Hyvarinen and R. Karthikesh, "Imposing sparsity on the mixing matrix in independent component analysis," *Neurocomputing* **49**, pp. 151–162, 2002.
9. "Spatiotemporal blind source separation using double-sided approximate joint diagonalization," in *In Proc. EUSIPCO 2005*, 2005.
10. R. Rubinstein, M. Zibulevsky, and M. Elad, "Learning sparse dictionaries for sparse signal representations," 2008. Submitted.
11. 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* **57**, pp. 1413–1457, Aug 2004.
12. E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1 -regularized minimization with applications to compressed sensing," tech. rep., Rice University, July 2007.
13. 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 - To appear* , 2007.
14. M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Transactions on Information Theory* **52**(12), pp. 5559–5569, 2006.
15. J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising.," *IEEE Transactions on Image Processing* **11**(6), pp. 670–684, 2002.
16. 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* **in press**, 2008.
17. Y. Moudden, J. Bobin, and J.-L. Starck, "An iterative thresholding algorithm for joint deconvolution and separation of multichannel data," in *ADA V, Heraklion*, May 2008.
18. M. Aharon, M. Elad, and A. Bruckstein, "k-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing* **54**(11), pp. 4311–4322, 2006.
19. S. Becker, J. Bobin, and E. J. Candès, "Nesta: a fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Science - submitted* , 20099.
20. J. Bobin, J. Fadili, Y. Moudden, and J.-L. Starck, "Morphological diversity and sparsity: New insights into multivariate data analysis," in *Proceedings of the SPIE conference wavelets (SPIE 2007) - San Diego*, 2007.