Sparsity and morphological diversity for hyperspectral data analysis

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

SPARSITY AND MORPHOLOGICAL DIVERSITY FOR HYPERSPECTRAL DATA ANALYSIS

J. Bobin[1]*, Y. Moudden*[2]*, J.-L. Starck*[2] *and J. Fadili*[3]

[1]California Institute of Technology, Applied and Computational Mathematics,
217-50 Pasadena, CA 91125, USA.

[2]CEA, Saclay, IRFU, SEDI/Service d'Astrophysique
Orme des Merisiers 91191 Gif-sur-Yvette, France.

[3]GREYC CNRS UMR 6072, Image Processing Group
ENSICAEN 14050, Caen Cedex, France.

ABSTRACT

Recently morphological diversity and sparsity have emerged as new and effective sources of diversity for Blind Source Separation. Based on these new concepts, novel methods such as Generalized Morphological Component Analysis have been put forward. The latter takes advantage of the very sparse representation of structured data in large overcomplete dictionaries, to separate sources based on their morphology. Building on GMCA, the purpose of this contribution is to describe a new algorithm for *hyperspectral* data processing. Large-scale hyperspectral data refers to collected data that exhibit sparse spectral signatures in addition to sparse spatial morphologies, in specified dictionaries of spectral and spatial waveforms. Numerical experiments are reported which demonstrate the validity of the proposed extension for solving source separation problems involving hyperspectral data.

Index Terms— source separation, hyperspectral data, sparsity, remote sensing, wavelets, astronomy.

1. INTRODUCTION

Generalized Morphological Component Analysis (GMCA) is a recent algorithm for multichannel data analysis and blind source separation introduced in [2], which relies on sparsity and morphological diversity to disentangle the source signals from observed mixtures. Indeed, sparsity has been recognized recently as an effective source of diversity for BSS [6]. The successes of GMCA in many generic multichannel data processing applications, including BSS, color image restoration and inpainting [3, 2], strongly motivated research to extend its applicability. In particular, there are instances where one is urged by *prior* knowledge to set additional constraints on the estimated parameters (*e.g.* equality constraints, positivity). Building on GMCA, the purpose of this contribution is to describe an algorithm for so-called *hyperspectral* data processing. The modified GMCA algorithm this entails is then given in section 3. Finally, numerical experiments in section 4 demonstrate the efficiency of the proposed method to deal with Hyperspectral data in component separation problems.

2. MODEL AND OBJECTIVE FUNCTION

Hyperspectral imaging systems collect data in a large number (up to several hundreds) of contiguous spectral intervals so that it makes sense to consider *e.g.* the regularity of some measurement from one channel to its neighbors. Typically, the spectral signatures of the structures in the data may be known *a priori* to have a sparse representation in some specified possibly redundant dictionary of template *spectral* waveforms.
In what follows, regardless of other definitions or models living in other scientific communities, the term *hyperspectral* will be used generically to identify multichannel data with the above two specific properties *i.e.* that the number of channels is large and 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. We further assume that hyperspectral data is structured *a priori* according to the simple yet common instantaneous linear mixture model given as follows :

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} = \sum_{k=1}^{n} \mathbf{X}_k + \mathbf{N} \qquad (1)$$

where the measurements $\mathbf{X} \in \mathbb{R}^{m,t}$ are a mixture of contributions from various objects $\mathbf{X}_k$ with different statistical and *spatio-spectral* properties. These induce *inter-* and *intra-* sensor structures or signal coherence in space and across sensors which we classically assume to be well represented by rank one matrices, product of a spectral signature $a^k \in \mathbb{R}^{m,1}$ and a spatial density profile $s_k \in \mathbb{R}^{1,t}$. Here, s_k and a^k are respectively the k^{th} row and column of $\mathbf{S} \in \mathbb{R}^{n,t}$ and $\mathbf{A} \in \mathbb{R}^{m,n}$. The $m \times t$ random matrix $\mathbf{N}$ is included to account for Gaussian instrumental noise assumed, for simplicity, to be uncorrelated inter- and intra- channel with variance σ^2.
Initially, the GMCA algorithm aims at solving the following non-convex problem :

$$\min_{\mathbf{A},\mathbf{S}} \sum_k \lambda_k \|\nu_k\|_1 + \frac{1}{2\sigma^2} \left\| \mathbf{X} - \sum_k a^k s_k \right\|^2 \qquad (2)$$

where $s_k = \nu_k \mathbf{\Phi}$. We describe next the proposed modified GMCA algorithm for hyperspectral data processing when it is known *a priori* that the underlying objects of interest $\mathbf{X}_k = a^k s_k$ exhibit sparse spectral signatures and sparse spatial morphologies in known dictionaries of spectral and spatial waveforms. Accounting for this prior is made by assuming that each component $\mathbf{X}_k = a^k s_k$ has a sparse representation in a given *multichannel* signal waveform dictionnary $\{\mathbf{\Phi}, \mathbf{\Psi}\}$. The derived optimization problem for hyperspectral data is then the following :

$$\min_{\mathbf{A},\mathbf{S}} \sum_k \lambda_k \|\gamma_k \nu_k\|_1 + \frac{1}{2\sigma^2} \left\| \mathbf{X} - \sum_{k=1}^{n} a^k s_k \right\|^2 \qquad (3)$$

where $\mathbf{X}_k = a^k s_k = \mathbf{\Phi} \gamma_k \nu_k \mathbf{\Psi}$.

3. GMCA ALGORITHM FOR *HYPERSPECTRAL* DATA

For the sake of simplicity, consider now that the $\mathbf{\Psi}$ and $\mathbf{\Phi}$ reduce to respectively spectral and spatial waveforms orthogonal bases. In this case, the minimization problem (3) is best formulated in coefficient space as follows :

$$\min_{\{\gamma_k,\nu_k\}} \frac{1}{2\sigma^2} \|\alpha - \gamma\nu\|^2 + \sum_{k=1}^{n} \lambda_k \|\gamma^k \nu_k\|_1 \qquad (4)$$

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 solutions to problem (4). 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. Inspired by [4], we obtain the following system of update rules, akin to a Projected Landweber algorithm [1] and after some algebra (see [3]) :

$$\nu^{(+)} = \Delta_\eta \left(\left(\gamma^T \gamma\right)^{-1} \gamma^T \alpha \right) \qquad (5)$$

$$\gamma^{(+)} = \Delta_\zeta \left(\alpha \nu^T \left(\nu \nu^T\right)^{-1} \right) \qquad (6)$$

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 (5) and (6) rules are easily interpreted as thresholded alternate least squares solutions. Finally, in the spirit of the GMCA algorithm [3, 2], it is proposed that a solution to problem (4) 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 α
3. While $\lambda_k^{(h)}$ are higher than a given lower bound $\lambda_{\min}$,
   – Update ν assuming γ is fixed using equation (5).
   – Update γ assuming ν is fixed using equation (6).
   – Decrease the thresholds $\lambda_k^{(h)}$.
5. Transform back γ and ν to estimate $\mathbf{A}$ and $\mathbf{S}$.

The *salient to fine* process is also the core of hypGMCA. With the threshold successively decreasing towards zero along iterations, the current sparse approximations for γ and ν are progressively refined by including finer structures spatially and spectrally, alternatingly. 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. When non-unitary or redundant transforms are used, the above is no longer strictly valid. Nevertheless, simple shrinkage still gives satisfactory results in practice.

4. NUMERICAL EXPERIMENTS

In this section, we report on two toy BSS experiments in 1D and 2D to compare GMCA to its extension

hypGMCA. First we consider synthetic 2D data consisting of $m = 128$ mixtures of $n = 5$ image sources. The sources are a set of 128×128 structured images. The spectra were generated as sparse process 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 1 displays four typical noisy simulated mixture data with SNR $= 20$dB. A visual ins-

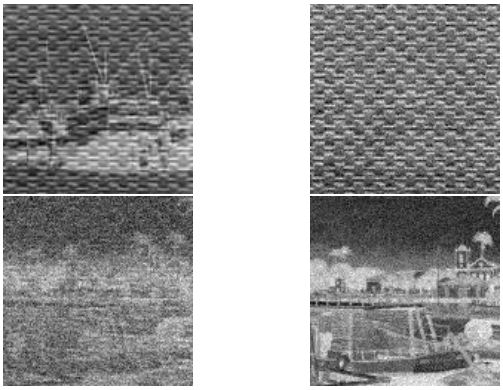

Fig. 1. Four 128×128 mixtures out of the 128 channels. The SNR is equal to 20dB.

pection of figure 2 allows a first qualitative assessment of the improved source recovery provided by correctly accounting for *a priori* spatial as well as spectral sparsity. The top images were obtained with GMCA while the bottom images, were obtained with hypGMCA. In all cases, both methods were run in the curvelet domain [5] with the same number of iterations. The graph on figure 3 gives more quantitative results. It traces the evolution of the mixing matrix criterion $\mathcal{C}_{\mathbf{A}} = \|\mathbf{I}_n - \mathbf{P}\tilde{\mathbf{A}}^\dagger \mathbf{A}\|_1$ as a function of the SNR which was varied from 0 to 40dB, where $\mathbf{P}$ serves to reduce the scale and permutation indeterminacy of the mixing model and $\tilde{\mathbf{A}}^\dagger$ is the pseudo-inverse of the estimated mixing matrix. In simulation, the true source and spectral matrices are known and so that $\mathbf{P}$ can be computed easily. The mixing matrix criterion is then strictly positive unless the mixing matrix 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.

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. Increasing the number of sources makes the separation task more difficult. We consider now 1D syn-

Fig. 2. **Left column :** Estimated sources using the original GMCA algorithm. **Right column :** Estimated sources using the new hypGMCA.

thetic 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 $\mathbf{\Phi}$. 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 $\mathbf{\Psi}$. The data are not contaminated by noise. The number of samples is $t = 2048$ and the number of channels is $m = 128$. Figure 4 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 trials. The panel on the left of Figure 4 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 $n = 64$, the GMCA algorithm with the spectral sparsity constraint outperforms the original GMCA by up to 12dB. The right of Figure 4 shows the behavior of both algorithms in terms of $\mathcal{C}_{\ell_1} = \sum_{i=1}^n \left\|a^i s_i - \tilde{a}^i \tilde{s}_i\right\|_1 / \sum_{i=1}^n \left\|a^i s_i\right\|_1$. As expected, accounting for spectral sparsity yields sparser

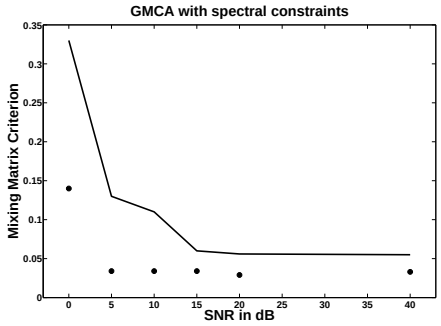

Fig. 3. **Evolution of the mixing matrix criterion $\mathcal{C}_{\mathbf{A}}$ as a function of the SNR in dB.** *Solid line :* recovery results with GMCA. • *:* recovery results with hypGMCA.

results. Furthermore, as the number of sources increases, the deviation between the aforementioned methods becomes wider.

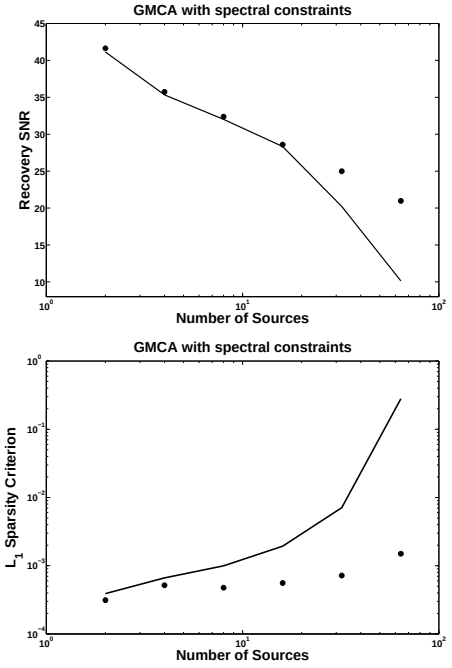

Fig. 4. **Abscissa :** Number of sources. **Ordinate - left :** Recovery SNR. **Right :** sparsity-based criterion $\mathcal{C}_{\ell_1}$. *Solid line :* recovery results with GMCA. • *:* recovery results with hypGMCA.

5. 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 decreasing threshold. This alone gives the method true robustness to noise. As expected, taking into account the additional prior knowledge of spectral sparsity leads to enhanced performance. It was illustrated by numerical experiments that spatiospectral sparsity yields robustness to noise contamination, as well as statbility when the dimensionality of the problem increases. Current work is on enforcing other prior constraints such as positivity and on applications of the proposed method.

6. REFERENCES

[1] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications - submitted*, 2007.

[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(11) :2662 – 2674, November 2007.

[3] J. Bobin, J-L. Starck, Y. Moudden, and J. Fadili. *Blind Source Separation : the Sparsity Revolution*. Advances in Imaging and Electron Physics. 2008.

[4] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4) :1168–1200, 2005.

[5] Jean-Luc Starck, Emmanuel J. Candès, and David L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6) :670–684, 2002.

[6] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13 :863–882, 2001.