## Morphological Diversity and Sparsity for Multichannel Data Restoration

J. Bobin · Y. Moudden · J. Fadili · J.-L. Starck

© Springer Science+Business Media, LLC 2008

Abstract Over the last decade, overcomplete dictionaries and the very sparse signal representations they make possible, have raised an intense interest from signal processing theory. In a wide range of signal processing problems, sparsity has been a crucial property leading to high performance. As multichannel data are of growing interest, it seems essential to devise sparsity-based tools accounting for such specific multichannel data. Sparsity has proved its efficiency in a wide range of inverse problems. Hereafter, we address some multichannel inverse problems issues such as multichannel morphological component separation and inpainting from the perspective of sparse representation. In this paper, we introduce a new sparsity-based multichannel analysis tool coined multichannel Morphological Component Analysis (mMCA). This new framework focuses on *multichannel* morphological diversity to better represent multichannel data. This paper presents conditions under which the mMCA converges and recovers the sparse multichannel representation. Several experiments are presented to demonstrate the applicability of our approach on

J. Bobin (🖾) · Y. Moudden · J.-L. Starck CEA-DAPNIA/SEDI, Service d'Astrophysique, CEA/Saclay, 91191 Gif-sur-Yvette, France e-mail: jerome.bobin@cea.fr

Y. Moudden e-mail: ymoudden@cea.fr

J.-L. Starck e-mail: jstarck@cea.fr

#### J. Fadili

GREYC CNRS UMR 6072, Image Processing Group, ENSICAEN 14050, Caen Cedex, France e-mail: jalal.fadili@greyc.ensicaen.fr a set of multichannel inverse problems such as morphological component decomposition and inpainting.

**Keywords** Sparsity · Overcomplete representations · Multichannel data · Restoration

## 1 Introduction

This paper addresses several multichannel data recovery problems such as multichannel morphological component decomposition and inpainting. We first need to define with care what multichannel data are. Such data are often physically composed of *m* observations (a color layer in color images, an observation at a fixed frequency for multispectral data and so on). One classical example of such multichannel data are the hyperspectral data provided by satellite observations; a fixed geographic area is observed at *m* different frequencies. More formally, we assume that each observation is made of t samples. We will write each observation as a  $1 \times t$  row vector  $\{x_i\}_{i=1,\dots,m}$ . For convenience, those m vectors are stacked in a  $m \times t$  matrix  $\mathbf{X} = [x_1^T \cdots x_m^T]^T$ . In a wide range of applications, the data X are often degraded by the acquisition system (convolution, missing data to quote a few) and contaminated by additive noise of finite variance. The various restoration problems we address in this paper are modeled as follows:

$$\mathbf{Y} = \mathcal{F}(\mathbf{X}) + \mathbf{N} \tag{1}$$

where  $\mathcal{F}$  is the *degradation* mapping, **X** is the original multichannel data to be recovered and **N** models noise or model imperfections. The mapping  $\mathcal{F}$  will depend on the recovery problem.

In the last decade sparsity has been one of the leading concepts in many areas of signal and image processing (restoration [1], feature extraction [2], source separation [3– 5], to name only a few). In a wide range of applications and viewpoints, researchers have advocated the use of overcomplete signal representations. Indeed, the attractiveness of redundant signal representations lies in their ability to sparsely represent a large class of signals. Furthermore, overcomplete representations allow more flexibility in signal representation and entail effectiveness at many signal processing tasks such as restoration, separation, compression, estimation etc. In neuroscience, the mammalian primary visual system has been shown to be probably in need of overcomplete representations [6].

In our representation, a *monochannel* row vector signal  $x \in \mathbb{R}^{1 \times t}$  is assumed to be the linear combination of T > t signal waveforms or *atoms*  $\{\phi_i\}_{i=1,...,T}$ :  $x = \sum_{i=1} \alpha_i \phi_i$ , where  $\alpha_i = \langle \phi_i, x \rangle$  are called the decomposition coefficients of x in the dictionary  $\mathbf{\Phi} = [\phi_1^T, \ldots, \phi_T^T]^T$  (the  $T \times t$  matrix whose lines are the atoms normalized to a unit  $\ell_2$ -norm). The signal x is said to be sparse in  $\mathbf{\Phi}$  if only a few of the entries of the row vector  $\alpha = [\alpha_1 \cdots \alpha_T]$  are non-zero. We define  $\Lambda_x$ , the *support* of x in  $\mathbf{\Phi}$ , as the set of indices of the non-zero entries in  $\alpha$ :

$$\Lambda_x = \{i \mid |\alpha_i| \neq 0\}. \tag{2}$$

From a heuristic viewpoint, the use of sparse representations are motivated by their ability to compactly represent the structure in the data. For instance, in image processing, a number of dictionaries have been designed that can capture very different features in an image: discrete cosine basis for globally oscillating patterns, wave atoms for local oscillatory textures [7], wavelets for pointwise singularities [8], curvelets for edges and contours [9, 10].

Returning to the problem in (1), a first and somewhat naive approach would consist in treating separately *m* single-channel restoration problems.<sup>1</sup> However, this is clearly suboptimal since, as we pointed out earlier, the msingle channel observations share some joint structure requiring a coherent processing of the m channels simultaneously. For instance, the color layers of color images have often similar patterns. Thus recovering each observation separately is far from being optimal. Such inter-observation structures have to be properly modeled in order to enhance multichannel data restoration. A first sparsity-based solution consists in assuming that each channel must share the same sparsity pattern *i.e.* the same support in  $\Phi$ . Such solutions have been proposed for several applications in previous work on multichannel sparse decomposition including [11, 12], in which all channels are constrained to have a common sparsity pattern, and [13-15] in which a specific sparsity measure is used. In this paper, we address a more general problem as we assume no constraint on the sparsity pattern of the different channels. We adopt a different point of view and model *inter-observation* structures as sparse patterns in a specific representation. The multichannel data are no longer modeled as a concatenation of observations that are individually sparse in a *spatial/temporal* representation but as a *single* signal that is sparse in a *multichannel* representation.

*Contributions* In this paper we propose solving multichannel data restoration problems within the paradigm of multichannel sparse representations. Section 2 introduces a new sparsity-based framework for analyzing multichannel signals coined multichannel Morphological Component Analysis (mMCA). In Sect. 2.2.1 we enlighten the connections between mMCA and other extensions of sparse decomposition problems to the multichannel case. Section 2.3 extends the MCA algorithm [16, 17] to the multichannel case and introduces convergence results proving the efficiency of mMCA in providing sparse decompositions. Two applications in data recovery issues of the proposed mMCA algorithm are described: multichannel morphological component decomposition in Sect. 3.1 and inpainting in Sect. 3.2. We also put forward a mMCA-based adaptive restoration algorithm to better match the sparse representation to the multichannel data at hand.

## 1.1 Notations and Definitions

We here define some useful quantities and notations. In the multichannel case, the data **X** live in the tensor product space of  $\mathbb{R}^m$  and  $\mathbb{R}^t$ :  $\mathbf{X} \in \mathbb{R}^m \otimes \mathbb{R}^t$ , where *m* is the number of channels. The upcoming proofs and results can be extended easily to the case where  $\mathbf{X} \in \mathbb{R}^{t_1} \otimes \cdots \otimes \mathbb{R}^{t_p}$  with  $p \ge 2$ . Multichannel data often consist of *m* "observations"

 $p \ge 2$ . Multichannel data often consist of *m* "observations"  $\{x_i\}_{i=1,...,m}$  from *m* channels, each lying in  $\mathbb{R}^t$ . For convenience, we use the following matrix notation:

$$\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$
(3)

where each channel  $\{x_i\}_{i=1,...,m}$  is a  $1 \times t$  row vector. A multichannel dictionary is no more than a set of vectors living in  $\mathbb{R}^m \otimes \mathbb{R}^t$ . For instance, a basis of  $\mathbb{R}^m \otimes \mathbb{R}^t$  is readily obtained as the tensor product of a basis of  $\mathbb{R}^{m,1}$  (say  $\Xi$ ) and a basis of  $\mathbb{R}^{1,t}$  (say  $\Phi$ ). Projecting **X** onto each atom of the multichannel basis  $\Psi = \Xi \otimes \Phi$  is done as follows:

$$\alpha = \mathbf{\Xi}^T \mathbf{X} \mathbf{\Phi}^T. \tag{4}$$

Let  $\psi_{\gamma=\{i,j\}} = \xi_i \otimes \phi_j$  be an atom (*i.e.* element) of the multichannel dictionary  $\Psi$ . This atom can be written in matrix

<sup>&</sup>lt;sup>1</sup>Note that this solution may only hold if the degradation mapping applies on each channel separately:  $\mathcal{F}(\mathbf{X}) = [\mathcal{F}(x_1)^T \cdots \mathcal{F}(x_m)^T]^T$ .

form as:  $\psi_{ij} = \xi_i \phi_j$  which is an  $m \times t$  rank-1 matrix. If  $\Xi$  or  $\Phi$  is orthonormal then  $\Psi$  is also orthonormal. In fact, the scalar product between two multichannel atoms is such that:<sup>2</sup>

$$\langle \psi_{ip}, \psi_{jq} \rangle = \langle \xi_i, \xi_j \rangle \langle \phi_p, \phi_q \rangle.$$
(5)

The mutual coherence (see [18] and references therein) of a dictionary  $\mathbf{\Phi}$ , defined as follows:

$$\mu_{\Phi} = \max_{p \neq q} |\langle \phi_p, \phi_q \rangle| \tag{6}$$

is a measure of how its atoms *look like each other*. Handling subsets of elements that belong to  $\Psi$  will also be needed. Let  $\Lambda = \{\{i_1, j_1\}, \dots, \{i_T, j_T\}\}$  be a set of index couples. The active subdictionary obtained by restricting the dictionary  $\Psi$  to the atoms whose indices are the elements of  $\Lambda$  is written  $\Psi_A$ . This notation will be useful to define the support of a signal **X** in  $\Psi$ . Assume that **X** is *K*-sparse in  $\Psi$ then  $\mathbf{X} = \sum_{\{i,j\} \in A_x} \alpha_{i,j} \psi_{ij} = \sum_{\{i,j\} \in A_x} \alpha_{i,j} \xi_i \phi_j$  where  $\Lambda_x$ is the support of **X**, and  $\Psi_{\Lambda_x}$  the corresponding active subdictionary.

According to the definition of the scalar product between two multichannel atoms given in (5), the mutual coherence for multichannel dictionaries  $\Psi = \Xi \otimes \Phi$  is as follows:

$$0 \le \mu_{\Psi} = \max\{\mu_{\Xi}, \mu_{\Phi}\} < 1 \tag{7}$$

for orthonormal  $\Xi$  and  $\Phi$ . In the next, the Frobenius norm of a matrix **X** is  $\|\mathbf{X}\|^2 = \text{Trace}(\mathbf{X}^T \mathbf{X})$ . The  $\ell_1$  norm of **X** is defined as the sum of the absolute values of the entries of matrix **X**.

## 2 Morphological Component Analysis for Multichannel Data

2.1 Morphological Diversity and Morphological Component Analysis

#### 2.1.1 An Introduction to Morphological Diversity

A monochannel signal x is said to be sparse in a waveform dictionary  $\Phi$  if it can be well represented from a few dictionary elements. More precisely, let us define  $\alpha$  such that:

$$x = \alpha \Phi. \tag{8}$$

The entries of  $\alpha$  are commonly called "coefficients" of x in  $\Phi$ . In that setting, x is said to be sparse in  $\Phi$  if most entries of  $\alpha$  are nearly zero and only a few have "significant"

amplitudes. Particular  $\ell_0$ -sparse signals are generated from a few non-zero dictionary elements. Note that this notion of sparsity is strongly dependent on the dictionary  $\Phi$ ; see e.g. [19, 20] among others. As discussed in [2], a single basis is often not well-adapted to large classes of highly structured data such as "natural images". Furthermore, over the past ten years, new tools have emerged from modern computational harmonic analysis: wavelets, ridgelets [21], curvelets [9, 10, 22], bandlets [23], contourlets [24], to name a few. It is quite tempting to combine several representations to build a larger dictionary of waveforms that will enable the sparse representation of larger classes of signals. Nevertheless, when  $\Phi$ is overcomplete (*i.e.* T > t), the solution of (8) is generally not unique. In that case, the authors of [19] were the first to seek the sparsest  $\alpha$ , in terms of  $\ell_0$ -pseudo-norm, such that  $x = \alpha \Phi$ . This approach leads to the following minimization problem:

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t. } x = \alpha \Phi. \tag{9}$$

Unfortunately, this is an NP-hard optimization problem which is combinatorial and computationally unfeasible for most applications. The authors of [25] also proposed to convexify the objective functional by substituting the convex  $\ell_1$ norm for the  $\ell_0$  pseudo-norm leading to the following linear program:

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t. } x = \alpha \Phi. \tag{10}$$

This problem can be solved for instance using interiorpoint methods. It is known as Basis Pursuit [25] in the signal processing community. Nevertheless, problems (9) and (10) are seldom equivalent. Important research concentrated on finding equivalence conditions between the two problems [19, 20, 26–28]. See also [29] for an extensive review.

In [2, 16], the authors proposed a practical algorithm coined Morphological Component Analysis (MCA) aiming at decomposing signals in overcomplete dictionaries made of a union of bases. In the MCA setting, x is the linear combination of D morphological components:

$$x = \sum_{i=1}^{D} \varphi_i = \sum_{i=1}^{D} \alpha_i \Phi_i$$
(11)

where  $\{\Phi_i\}_{i=1,...,D}$  are orthonormal bases of  $\mathbb{R}^t$ . Morphological diversity then relies on the sparsity of those morphological components in specific bases. In terms of  $\ell_0$  norm, this morphological diversity can be formulated as follows:

$$\forall \{i, j\} \in \{1, \dots, D\}; \quad j \neq i \Rightarrow \|\varphi_i \mathbf{\Phi}_i^T\|_0 < \|\varphi_i \mathbf{\Phi}_j^T\|_0.$$

$$(12)$$

In other words, MCA relies on the incoherence between the sub-dictionaries  $\{\Phi_i\}_{i=1,...,D}$  to estimate the morphological

<sup>&</sup>lt;sup>2</sup>In fact, by standard properties of the tensor product, one can easily show that the Gram matrix of a tensor product is the tensor product of the Gram matrices. That is,  $G_{\Psi} = G_{\Xi} \otimes G_{\Phi}$ .

components  $\{\varphi_i\}_{i=1,...,D}$  by solving the following convex minimization problem:

$$\{\varphi_i\}_{1 \le i \le D} = \underset{\{\varphi_i\}_{1 \le i \le D}}{\arg\min} \left\| x - \sum_{i=1}^{D} \varphi_i \right\|_2^2 + 2\lambda \sum_{i=1}^{D} \|\varphi_i \mathbf{\Phi}_i^T\|_1.$$
(13)

Note that the minimization problem in (13) is closely related to Basis Pursuit Denoising (BPDN—see [25]). In [30], we proposed a particular block-coordinate relaxation, iterative thresholding algorithm (MCA/MOM) to solve (13). Theoretical arguments as well as experiments were given showing that MCA provides at least as good results as Basis Pursuit for sparse overcomplete decompositions in a union of bases. Moreover, MCA turns out to be clearly much faster than Basis Pursuit. Then, MCA is a practical alternative to classical sparse overcomplete decomposition techniques.

## 2.1.2 Morphological Diversity in Multichannel Data

In the previous paragraph, we gave a brief description of morphological diversity in the monochannel case. In this paper, we extend morphological diversity to the multichannel case. In this particular setting, we assume that each observation or channel  $\{x_i\}_{i=1,...,m}$  is the linear combination of *D* morphological components:

$$\forall i \in \{1, \dots, m\}; \quad x_i = \sum_{j=1}^{D} \varphi_{ij} \tag{14}$$

where each morphological component  $\varphi_{ij}$  is sparse in a specific basis  $\Phi_j$ . Then each channel  $\{x_i\}_{i=1,...,m}$  is assumed to be sparse in the overcomplete dictionary  $\Phi$  made of the union of the *D* bases  $\{\Phi_i\}_{i=1,...,D}$ .

We further assume that each column of the data matrix **X** is sparse in the dictionary  $\Xi$  made of the union of D' bases  $\{\Xi_i\}_{i=1,...,D'}$  to account for inter-channel structures. The multichannel data **X** are then assumed to be sparse in the multichannel dictionary  $\Psi = [\Xi_1 \cdots \Xi_{D'}] \otimes [\Phi_1 \cdots \Phi_D]$ . The multichannel data are then modeled as the linear combination of  $D \times D'$  multichannel morphological components:

$$\mathbf{X} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \boldsymbol{\varpi}_{jk} \tag{15}$$

where  $\varpi_{jk}$  is sparse in  $\Xi_k \otimes \Phi_j$ . In this setting, separating two multichannel morphological components  $\varpi_{ip}$  and  $\varpi_{jq\neq ip}$  based on multichannel morphological diversity may put on different faces:

- Spatial or temporal (resp. spectral) morphologies: in this case  $i \neq j$  and p = q (resp. i = j and  $p \neq q$ ). The morphological components have the same spectral representation (resp. spatial basis) but one can discriminate between

them based on their spatial (resp. spectral) diversity. It is easily seen that the coherence between subdictionaries  $\Xi_p \otimes \Phi_i$  and  $\Xi_p \otimes \Phi_j$  (resp.  $\Xi_p \otimes \Phi_i$  and  $\Xi_q \otimes \Phi_i$ ) is upper-bounded by  $\mu_{\Phi}$  (resp.  $\mu_{\Xi}$ ).

- Both morphologies:  $i \neq j$  and  $p \neq q$ , the "separation" task seems easier as the morphological components share neither the same spectral basis nor the same spatial (or temporal) basis. Note that in this case, the coherence between  $\Xi_p \otimes \Phi_i$  and  $\Xi_q \otimes \Phi_j$  is lower than  $\mu_{\Xi} \mu_{\Phi} \leq \max\{\mu_{\Xi}, \mu_{\Phi}\}$ .

Analyzing multichannel signals requires accounting for their *spectral* and *spatial* morphological diversities. For that purpose, the proposed multichannel extension to MCA coined mMCA, which stands for multichannel Morphological Component Analysis, aims at solving the following minimization problem:

$$\min_{\{\varpi_{jk}\}} \left\| \mathbf{X} - \sum_{j=1}^{D} \sum_{k=1}^{D'} \varpi_{jk} \right\|_{2}^{2} + 2\lambda \sum_{j=1}^{D} \sum_{k=1}^{D'} \| \mathbf{\Xi}_{k}^{T} \varpi_{jk} \mathbf{\Phi}_{j}^{T} \|_{1}.$$
(16)

In Sect. 2.2.1 we enlighten the connections between the problem in (16) and the extension of BPDN to the multichannel case. We also provide straightforward multichannel extensions of well-known recovery results. In Sect. 2.3 we introduce an MCA-based block-coordinate relaxation, iterative thresholding algorithm to solve (16). We give new theoretical conditions under which the mMCA algorithm provides the solution to the problem in (16).

### 2.2 Multichannel Overcomplete Sparse Recovery

## 2.2.1 General Multichannel Overcomplete Sparse Representation

Solving the problem in (16) is a particular case of a more general extension of the problem in (10) to the multichannel case: decomposing data in an overcomplete multichannel dictionary  $\Psi = \Xi \otimes \Phi$  (recall that  $\Xi$  is a  $m \times M$  overcomplete dictionary with M > m,  $\Phi$  is a  $T \times t$  overcomplete dictionary with T > t). Similarly to (9), this requires solving the following problem:

$$\min_{\alpha} \|\alpha\|_{\ell_0} \quad \text{s.t. } \mathbf{X} = \mathbf{\Xi} \alpha \mathbf{\Phi} \tag{17}$$

where  $\alpha$  is an  $M \times T$  matrix and  $\|\alpha\|_{\ell_0}$  refers to the number of non-zero entries in  $\alpha$ . The convex  $\ell_1$  minimization problem (10) can also be rewritten in the multichannel setting:

$$\min_{\alpha} \|\alpha\|_{\ell_1} \quad \text{s.t. } \mathbf{X} = \mathbf{\Xi} \alpha \mathbf{\Phi} \tag{18}$$

where  $\|\alpha\|_{\ell_1} = \sum_{i,j} |\alpha_{ij}|$ . From the optimization viewpoint, monochannel and multichannel problems are similar. Recall

that in the case of a monochannel K-sparse signal to be decomposed in a dictionary with coherence  $\mu_{\Psi}$ , the two aforementioned problems share the same unique solution when the following condition holds [19]:

$$K < \frac{1}{2} \left( 1 + \frac{1}{\mu \Psi} \right).$$

The uniqueness and equivalence condition of the sparse multichannel decomposition problem in (17) is then similar to the monochannel case. Assume that X is K-sparse in the multichannel dictionary  $\Psi = \Xi \otimes \Phi$ . The  $\ell_0$  sparse decomposition problem in (17) has a unique solution and problems in (17) and (18) are equivalent when:

$$K < \frac{1}{2} \left( 1 + \frac{1}{\mu_{\Psi}} \right) \text{ where } \mu_{\Psi} = \max\{\mu_{\Xi}, \mu_{\Phi}\}.$$

In this framework, most results in the monochannel case [18,20, 26–28, 31] can be straightforwardly extended to the multichannel case. Previous work on multichannel sparse decomposition includes [32] which introduced the concept of multichannel dictionary.

#### 2.3 Multichannel Morphological Component Analysis

The problem at stake in (16) can be solved by extending well-known sparse decomposition algorithms to the multichannel case (Basis Pursuit [25], Matching Pursuit [33, 34]), LARS/LASSO [35], Homotopy continuation [36, 37], Polytope Faces Pursuit (PFP) [38] to quote a few). Extending MP and OMP to the multichannel case has been proposed in [32]. Interestingly, most greedy sparse decomposition techniques are closely linked to variable selection. Indeed, LARS/LASSO [35] and Homotopy continuation [36, 37] were first introduced in statistics to solve variable selection problems with an  $\ell_1$  sparsity constraint. The aforementioned greedy methods iteratively select one dictionary atom at a time. Unfortunately, this stepwise selection of active atoms is burdensome and the process may be sped up as in [39] where a faster stagewise Orthogonal Matching Pursuit (StOMP) is introduced. It is shown to solve the  $\ell_0$  sparse recovery problem in (9) with random dictionaries under mild conditions.

Owing to the particular structure of the problem in (16), extending the MCA algorithm [17] to the multichannel case would lead to faster and still effective decomposition results. Recall that in the mMCA setting, the data X are assumed to be the linear combination of  $D \times D'$  morphological components  $\{\varpi_{jk}\}_{j=1,\dots,D;k=1,\dots,D'}$ . We define  $\Lambda_{jk}$  as the support (*i.e.* the indices of active atoms) of  $\overline{\sigma}_{jk}$  in the subdictionary  $\Psi_{ik} = \Xi_k \otimes \Phi_i$ . As **X** is *K*-sparse in the whole dictionary,

 $\sum_{j,k} \operatorname{Card}(\Lambda_{jk}) = K$ . The data can be decomposed as follows

$$\mathbf{X} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \varpi_{jk} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \sum_{i \in \Lambda_{jk}} \alpha_{jk}[i] \psi_{jk}[i].$$
(19)

Substituting (19) in (16), the mMCA algorithm approaches the solution to (16) by iteratively and alternately estimating each morphological component  $\varpi_{ik}$  in a Block-coordinate relaxed way (see [40]). Each matrix of coefficients  $\alpha_{ik}$  is then estimated as follows:

$$\alpha_{jk} = \arg\min_{\alpha_{jk}} \|\mathbf{R}_{jk} - \mathbf{\Xi}_k \,\alpha_{jk} \Phi_j\|^2 + 2\lambda \|\alpha_{jk}\|_{\ell_1}$$
(20)

where  $\mathbf{R}_{jk} = \mathbf{X} - \sum_{p,q \neq j,k} \Xi_q \alpha_{pq} \Phi_p$  is a residual term.

Since we are assuming that the subdictionaries  $\{\Phi_i\}_i$  and  $\{\Xi_k\}_k$  are orthonormal, the problem in (20) is equivalent to the following:

$$\alpha_{jk} = \arg\min_{\alpha_{jk}} \|\mathbf{\Xi}_k^T \mathbf{R}_k \mathbf{\Phi}_j^T - \alpha_{jk}\|^2 + 2\lambda \|\alpha_{jk}\|_{\ell_1}$$
(21)

which has a unique solution  $\alpha_{jk} = \Delta_{\lambda} (\boldsymbol{\Xi}_{k}^{T} \mathbf{R}_{k} \boldsymbol{\Phi}_{j}^{T})$  known as soft-thresholding with threshold  $\lambda$  as follows:

$$\Delta_{\lambda}(u[i]) = \begin{cases} 0 & \text{if } u[i] < \lambda, \\ u[i] - \lambda \text{ sign } (u[i]) & \text{if } u[i] \ge \lambda. \end{cases}$$
(22)

For a fixed  $\lambda$ , mMCA selects groups of atoms based on their scalar product with the residual  $\mathbf{R}_{ik}$ . Assuming that we select only the most coherent atom (with the highest scalar product) with the residual  $\mathbf{R}_{ik}$  then one mMCA iteration boils down to a stepwise multichannel Matching Pursuit (mMP) step. In contrast with mMP, the mMCA algorithm is allowed to select several atoms at each iteration. Thus, when hard-thresholding is used instead of soft-thresholding, mMCA is equivalent to a stagewise mMP algorithm. Allowing mMCA to select new atoms is made by decreasing the threshold  $\lambda$  at each iteration. The mMCA algorithm is summarized below:

1. Set the number of iterations  $I_{\text{max}}$  and threshold  $\lambda^{(0)}$ .

2. While  $\lambda^{(h)}$  is higher than a given lower bound  $\lambda_{min}$  (e.g. can depend on the noise variance, see Sect. 2.5),

For 
$$j = 1, ..., D$$
 and  $k = 1, ..., D$ 

• Compute the residual term  $\mathbf{R}_{jk}^{(h)}$  assuming the current

estimates of  $\overline{\omega}_{pq \neq jk}$ ,  $\tilde{\omega}_{pq \neq jk}^{(h-1)}$  are fixed:

$$\begin{split} \mathbf{R}_{jk}^{(h)} = \mathbf{X} - \sum_{pq \neq jk} \tilde{\boldsymbol{\omega}}_{pq \neq jk}^{(h-1)}. \end{split}$$
• Estimate the current coefficients of  $\tilde{\boldsymbol{\omega}}_{jk}^{(h)}$  by thresholding

with threshold  $\lambda^{(h)}$ :  $\tilde{\alpha}_{jk}^{(h)} = \Delta_{\lambda^{(h)}} (\Xi_k^T \mathbf{R}_{jk}^{(h)} \mathbf{\Phi}_j^T).$ • Get the new estimate of  $\varpi_{jk}$  by reconstructing from the selected coefficients  $\tilde{\alpha}_{ik}^{(h)}$ :

$$\tilde{\varpi}_{jk}^{(h)} = \Xi_k \tilde{\alpha}_k^{(h)} \Phi_j.$$

3. Decrease the threshold  $\lambda^{(h)}$  following a given strategy.

#### 2.3.1 The Thresholding Strategy

In [30] we proposed a thresholding strategy that is likely to provide the solution to the  $\ell_0$  sparse monochannel problem. The strategy which goes by the name of MOM (for "Mean of Max") can be extended to the multichannel case. At each iteration *h* the residual is projected onto each sub-dictionary and we define:

$$m_{jk}^{(h-1)} = \left\| \mathbf{\Xi}_{k}^{T} \left( \mathbf{X} - \sum_{p,q} \mathbf{\Xi}_{q} \tilde{\alpha}_{pq}^{(h-1)} \mathbf{\Phi}_{p} \right) \mathbf{\Phi}_{j}^{T} \right\|_{\ell_{\infty}}.$$
 (23)

The multichannel-MOM (mMOM) threshold is then computed as the mean of the two largest values in the set  $\{m_{jk}^{(h-1)}\}_{j=1,...,D'}$ 

$$\lambda^{(h)} = \frac{1}{2} \{ m_{j_0 k_0}^{(h-1)} + m_{j_1 k_1}^{(h-1)} \}.$$
<sup>(24)</sup>

In the next section, we show conditions under which mMCA/mMOM selects atoms without error and converges asymptotically to the solution of the multichannel  $\ell_0$  sparse recovery problem in (17).

## 2.4 Recovering Sparse Multichannel Decompositions Using mMCA

The mMOM rule defined in (23)–(24) is such that mMCA will select, at each iteration, atoms belonging to the same subdictionary  $\Psi_{jk} = \Xi_k \otimes \Phi_j$ . Although it seems more computationally demanding, the mMOM strategy has several nice properties. We show sufficient conditions under which (i) mMCA/mMOM selects atoms belonging to the active atom set of the solution of the  $\ell_0$  sparse recovery problem (Exact Selection Property), (ii) mMCA/mMOM converges exponentially to **X** and its sparsest representation in  $\Psi$ . Let's mention that the mMCA/mMOM exhibits an auto-stopping behavior, and requires only one parameter  $\lambda_{min}$  whose choice is easy and discussed in Sect. 2.5.

The next proposition states that mMCA/mMOM verifies the Exact Selection Property (ESP) at each iteration.

**Proposition 1** (Exact Selection Property) *Suppose that* **X** *is K*-sparse such that:

$$\mathbf{X} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \sum_{i \in \Lambda_{jk}} \alpha_{jk}[i] \psi_{jk}[i]$$

where  $K = \sum_{j,k} \operatorname{Card}(\Lambda_{jk})$  satisfying  $K < \frac{\mu_{\psi}^{-1}}{2}$ . At the *h*-th iteration, assume that the residual  $\mathbf{R}^{(h)}$  is *K*-sparse such that:

$$\mathbf{R}^{(h)} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \sum_{i \in \Lambda_{jk}} \beta_{jk}[i] \psi_{jk}[i].$$

🖄 Springer

Then mMCA/mMOM picks up coefficients belonging to the support of **X** at iteration (h).

The proof is deferred to the appendix. When the previous Exact Selection Property holds, the next proposition shows that mMCA/mMOM converges exponentially to **X** and its sparsest representation in  $\Psi = [\Xi_1 \cdots \Xi_{D'}] \otimes [\Phi_1 \cdots \Phi_D]$ .

**Proposition 2** (Convergence) *Suppose that* **X** *is K-sparse such that:* 

$$\mathbf{X} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \sum_{i \in \Lambda_{jk}} \alpha_{jk}[i] \psi_{jk}[i]$$

where  $K = \sum_{j,k} \operatorname{Card}(\Lambda_{jk})$ .

If  $K < \frac{\mu_{\Psi}}{2}$  then mMCA/mMOM converges exponentially to **X** and its sparsest representation in  $\Psi$ . More precisely, the residual converges to zero at an exponential rate.

The proof to this second proposition is also given in the appendix. Note that the above conditions are far from being sharp. Exact Selection and convergence may still be valid beyond the bounds retained in the latter two statements.

#### 2.5 Handling Bounded Noise with mMCA

When bounded noise perturbs the data, the data are modeled as follows:

$$\mathbf{X} = \sum_{j=1}^{D} \sum_{k=1}^{D'} \sum_{i \in \Lambda_{jk}} \alpha_{jk}[i] \psi_{jk}[i] + \mathbf{Z}$$
(25)

where **Z** is a bounded noise:  $||\mathbf{Z}|| < \epsilon$ . Sparse recovery then needs to solve the following problem:

$$\min_{\alpha_{jk}} \sum_{j=1}^{D} \sum_{k=1}^{D'} \|\alpha_{jk}\|_{\ell_0} \quad \text{s.t.} \left\| \mathbf{X} - \sum_{j=1}^{D} \sum_{k=1}^{D'} \mathbf{\Xi}_k \alpha_{jk} \mathbf{\Phi}_j \right\| < \epsilon.$$
(26)

Sparse recovery and stability conditions have been studied in [41–43] in the monochannel case. More particularly, conditions are proved in [41] under which OMP verifies an Exact Selection Property in the presence of bounded noise  $\|\mathbf{Z}\| < \epsilon$ . They also showed that the OMP solution lies in a  $\ell_2$  ball centered on the exact solution to the  $\ell_0$  sparse recovery problem with a radius on the order of  $\epsilon$ . Exhibiting similar stability results in the mMCA setting is challenging and will be addressed in future work. In the mMCA framework, assuming the noise level is known, the mMCA/mMOM algorithm stops when  $\lambda \leq \lambda_{\min}$  with  $\lambda_{\min} = 3\epsilon$ . **Fig. 1** Multichannel morphological component separation: the original data **X** 



#### 2.6 Choosing the Overcomplete Dictionary

The choice of the overcomplete dictionary is a key step as it determines where we will be looking for a sparse representation. It is the expression of some prior information we have available on the signal. Interestingly, the  $\ell_1$  sparse recovery problem can be seen in the light of a Bayesian framework. Solving the following problem

$$\min_{\{\alpha_{jk}\}} \left\| \mathbf{X} - \sum_{j=1}^{D} \sum_{k=1}^{D'} \mathbf{\Xi}_{k} \alpha_{jk} \Phi_{j} \right\|^{2} + 2\lambda \sum_{j=1}^{D} \sum_{k=1}^{D'} \|\alpha_{jk}\|_{\ell_{1}}$$
(27)

is equivalent, in a Bayesian framework, to making the assumption among others of an independent sparse Laplacian *prior* on the coefficients of each morphological component in the sparse representation domain. Choosing the set of subdictionaries is then equivalent to assuming some specific *prior* for each morphological component.

Furthermore, the attractiveness of mMCA lies in its ability to take advantage of sparse representations which have implicit fast analysis and synthesis operators without requiring the explicit manipulation of each atom: wavelets, curvelets [10], bandlets [23], contourlets [24], ridgelets [21], wave atoms [7] to name a few. As a consequence, mMCA is a fast non-linear sparse decomposition algorithm whose computational complexity is dominated by that of the transforms involved in the dictionary.

In the next image processing experiments, we will assume that a wide range of images can be decomposed into a piecewise smooth (contour) part and an oscillating texture part. We will assume *a priori* that the contour part is sparse in the curvelet tight frame, and the texture part is sparsely described by the local discrete cosine transform (DCT) [8].<sup>3</sup> However, all the results we previously proved were given assuming that each subdictionary was an orthonormal basis. When the selected subdictionaries are more generally tight frames, the solution to (21) is no longer a simple thresholding. Nevertheless, in [44] and [45], the authors showed that thresholding is the first step towards solving (21) when the subdictionary is redundant. Rigorously, proximal-type iterative shrinkage is shown to converge to a solution of (21). In practice, even when the subdictionary is a tight frame (for instance the curvelet frame) we will only use a single thresholding step to solve (21).

The *spectral* dictionary  $\Xi$  is chosen based on a spectral sparsity assumption. The choice of the dictionary  $\Xi$  relies on sparsity prior information.

## **3** Applications to Some Sparse Multichannel Image Inverse Problems

#### 3.1 Multichannel Morphological Component Separation

In this section, we illustrate the ability of mMCA algorithm at extracting the so-called morphological components. For the sake of simplicity, the multichannel dictionary  $\Psi = \Xi \otimes$  $\Phi$  is such that  $\Xi$  and  $\Phi$  are both the union of the DCT-based basis ( $\Xi_2$  and  $\Phi_2$ ) and an orthogonal wavelet transform ( $\Xi_1$ and  $\Phi_1$ ). Hereafter, the data **X** are assumed to be the linear combination of 4 multichannel morphological components corresponding to the following multichannel bases:  $\Xi_1 \otimes$  $\Phi_1$ ,  $\Xi_2 \otimes \Phi_1$ ,  $\Xi_2 \otimes \Phi_1$  and  $\Xi_2 \otimes \Phi_2$ .

Without loss of generality, the experiment above involves mono-dimensional signals. Each multichannel morphological component is made of 256 channels. Each channel has 256 entries. Each multichannel morphological component  $\{\varpi_{ij}\}_{i=1,2; j=1,2}$  is the linear combination of multichannel atoms belonging to  $\{\Xi_i \otimes \Phi_j\}_{i=1,2; j=1,2}$  respectively. The related coefficients  $\{\alpha_{jk}\}_{j=1,2; k=1,2}$  have been drawn according to a Bernoulli-Gaussian distribution: the atoms are active (*i.e.* non zero) with probability  $p = 5.10^{-3}$  with random zero mean and unit variance Gaussian values. The left panels of Figs. 2 to 5 show the 5th channel (rows of  $\varpi$ ) and 200th column (columns of  $\varpi$ ) of these synthetic morphological components.

The multichannel MCA algorithm is then used to estimate the 4 aforementioned morphological components. The

<sup>&</sup>lt;sup>3</sup>An alternative choice would be the wave atoms [7].



Fig. 2 Multichannel morphological component separation component related to  $\Xi_1 \otimes \Phi_1$ : (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morpho-

 Table 1
 Recovery signal-to-noise ratio

	$\overline{\varpi}_{11}$	$\varpi_{12}$	$\varpi_{21}$	$\varpi_{22}$
Recovery SNR in dB	47.3	61.9	47.9	58.2

recovered multichannel morphological components are then depicted on the right panels of Figs. 2 to 5. At first sight, the pictured samples of the multichannel morphological components are visually well estimated. More quantitatively, Table 1 provides the recovery signal-to-noise ratio (dB).

Following the first visual impression, the mMCA algorithm performs well at extracting multichannel morphological components.

## 3.2 Multichannel Inpainting

In this section, we address the problem of multichannel image inpainting. In this context, a set of samples are missing



Estimated Morphological Component  $\Xi_1 \otimes \Phi_1$  – Multichannel Sample 200



logical component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the estimated morphological component

(set to zero) in the data. According to the general inverse problem framework in (1), the observed data are modeled as follows:

$$\mathbf{Y} = \mathcal{M} \odot \mathbf{X} + \mathbf{N} \tag{28}$$

where  $\mathcal{M}$  is a multichannel binary mask that multiplies the data matrix **X** entrywise. The entries of the multichannel mask  $\mathcal{M}$  take the value zero when the corresponding data pixels are missing and one otherwise. In this setting, data restoration is about *recovering* the missing pixels. Note that a multichannel mask applies on the whole data **X**; the missing pixels may not be the same in each channel. In the monochannel case, image inpainting is an old "interpolation" problem for which a wide range of techniques have been devised: variational approaches [46–50], sparsity-based methods [17, 51, 52]. Interestingly, the MCA based inpainting method described in [17] can be interpreted within the Expectation Maximization framework as shown in [53].



**Fig. 3** Multichannel morphological component separation component related to  $\Xi_1 \otimes \Phi_2$ : (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morpho-

We first propose to extend this algorithm to the multichannel case. We assume that the multichannel data are the linear combination of  $D \times D'$  multichannel morphological components as described in (19). The sparsity driven inpainting objective can be written as follows:

$$\min_{\{\alpha_{jk}\}} \left\| \mathbf{Y} - \mathcal{M} \odot \left[ \sum_{j=1}^{D} \sum_{k=1}^{D'} \mathbf{\Xi}_{k} \alpha_{jk} \mathbf{\Phi}_{j} \right] \right\|^{2} + 2\lambda \sum_{j=1}^{D} \sum_{k=1}^{D'} \|\alpha_{jk}\|_{\ell_{1}}.$$
(29)

Let  $\mathcal{M}^c$  denote the logical opposite of  $\mathcal{M}$  such that  $\mathcal{M}^c$  is also a binary mask where entries equal to one indicate invalid or missing pixels while zeros indicate ones that are present. Extending this inpainting algorithm to the multichannel setting boils down to a two-step iterative algorithm:





logical component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the estimated morphological component

- Update the data estimate E-step:  $\mathbf{Y}^{(h)} = \mathbf{Y} + \mathcal{M}^c \odot \mathbf{X}^{(h-1)}$
- Sparse decomposition **M-step**: a mMCA decomposition step of the current data estimate with the threshold  $\lambda^{(h)}$ .

The mMCA-based inpainting algorithm is summarized as follows:

1. Set the number of iterations  $I_{\text{max}}$  and threshold  $\lambda^{(0)}$ .

2. While  $\lambda^{(h)}$  is higher than a given lower bound  $\lambda_{min}$  (e.g. can depend on the noise variance as in Sect. 2.5),

a. Compute the hypercube estimate:  $\mathbf{Y}^{(h)} = \mathbf{Y} + \mathcal{M}^c \odot \tilde{\mathbf{X}}^{(h-1)}$ . b. Initialize to zero each residual morphological components  $\{\tilde{\boldsymbol{\varpi}}_{jk}\}^{(h-1)}$ .

For 
$$j = 1, ..., D$$
 and  $k = 1, ..., D$ 

• Compute the residual term  $\mathbf{R}_{jk}^{(h)}$  assuming the current

estimates of  $\varpi_{pq\neq jk}$ ,  $\tilde{\varpi}_{pq\neq jk}^{(h-1)}$  are fixed:

$$\mathbf{R}_{jk}^{(h)} = \mathbf{X} - \sum_{pq \neq jk} \tilde{\varpi}_{pq}^{(h-1)}.$$

• Estimate the current coefficients of  $\tilde{\varpi}_{jk}^{(h)}$  by thresholding with threshold  $\lambda^{(h)}$ :



Fig. 4 Multichannel morphological component separation component related to  $\Xi_2 \otimes \Phi_1$ : (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morpho-

$$\begin{split} \tilde{\alpha}_{jk}^{(h)} &= \Delta_{\lambda^{(h)}} (\Xi_k^T \mathbf{R}_{jk}^{(h)} \mathbf{\Phi}_j^T). \\ \bullet \text{ Get the new estimate of } \overline{\varpi}_{jk} \text{ by reconstructing from the selected coefficients } \tilde{\alpha}_{jk}^{(h)}: \end{split}$$

$$\tilde{\boldsymbol{\sigma}}_{\mu}^{(h)} = \boldsymbol{\Xi}_{\mu} \tilde{\boldsymbol{\alpha}}_{\mu}^{(h)} \boldsymbol{\Phi}_{\mu}$$

c. Update the hypercube 
$$\tilde{\mathbf{X}}^{(h)} = \sum_{j=1}^{D} \sum_{j=1}^{D'} \tilde{\boldsymbol{\varpi}}_{jk}^{(h)}$$
  
3. Decrease the threshold  $\lambda^{(h)}$  following mMOM strategy.

In the next subsection, we apply the multichannel inpainting algorithm to Hyperspectral data.

## 3.2.1 Hyperspectral Data Inpainting

We deal with a hyperspectral data cube **X**. Unfortunately, as usual when dealing with real data, multichannel pixels are missing. In the experiment we carried out, the data **X** is a *Mars Orbiter*<sup>4</sup> hyperspectral cube composed of a  $128 \times 128$ 



logical component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the estimated morphological component

spatial observations measured at 64 different frequencies (channels). **X** is then a  $128 \times 128 \times 64$  hyperspectral data cube. Although the data are real hyperspectral data, the so called missing pixels were synthetically picked out. We generated a hyperspectral mask such that a random proportion of randomly selected pixels are missing. We used the mMCA algorithm assuming the data are sparse in the dictionary  $\Psi = \Xi \otimes \Phi$ . Each spectrum of the data is assumed to be sparse in the orthogonal one-dimensional wavelet basis  $\Xi$ . Each spatial observation is nearly sparse in the orthogonal bidimensional wavelet basis  $\Phi$ . Figure 6(a) displays the 10th original channel. Figure 6(b) depicts the masked channel with 50% missing pixels. Figure 6(c) shows the recovered image using mMCA. The SNR between the inpainted image and the original is 19.2 dB. Visually, mMCA does a good job at recovering the spatial features of the data cube.

Figure 7(a) depicts the original spectrum at pixel  $\{10, 10\}$  of the data cube. The plot of Fig. 7(b) shows the masked

<sup>&</sup>lt;sup>4</sup>See the Mars Orbiter website at http://mars.jpl.nasa.gov/mro/.



Fig. 5 Multichannel morphological component separation component related to  $\Xi_2 \otimes \Phi_2$ : (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morpho-

spectrum. From Fig. 7(c), it can be seen that the recovered spectrum is well estimated by mMCA.

We repeated the same experiment with different fractions of missing pixels. Figure 8 shows the evolution of the SNR in dB between the original data and the recovered hyperspectral data as the percentage of missing pixels is increased from 5 to 75%. Even when 75% of the pixels are missing, mMCA is able to recover the data with a SNR of 10 dB. Furthermore, we applied the monochannel MCA-based inpainting algorithm [54] on each channel separately. The dashed line in Fig. 8 displays the behavior of this MCA-based recovery method. Clearly, mMCA performs far better than MCA in this experiment. Accounting for interchannel structures then leads to tremendous enhancement in hyperspectral image inpainting. The transition from the monochannel to the multichannel setting relies on the ability to account for spatial and spectral information. This experiment clearly demonstrates that the mMCA algorithm is well suited to handle such particular



logical component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the estimated morphological component

data as it performs well in terms of spectral and spatial feature recovery.

Relation to Compressed Sensing In signal processing, every student learns that, owing to the Nyquist-Shannon sampling theorem, the number of samples needed to recover any signal is dictated by its bandwidth. That is, a bandlimited signal whose bandwidth is  $F_x$  can be perfectly reconstructed from  $F_x$  equispaced samples.

Very recently, an alternative sampling theory has emerged which shows that signals can be recovered from far fewer samples (measurements) than what the Nyquist-Shannon sampling theorem states. This new theory, which goes by the name of *compressed/ive sensing/sampling* was introduced in the seminal paper [55]. It relies on the compressibility of signals or more precisely on the property for some signals to be sparsely represented. From the compressed sensing (CS) viewpoint, sparse signals could be acquired "economically" (from a few samples) without loss of information. **Fig. 6** Restoring Mars-Orbiter hyperspectral data: spatial feature recovery. (**a**) 10th original channel. (**b**) Masked channel—50% of the pixels are missing. (**c**) Recovered channel using mMCA





This has many important implications as it suggests new ways of designing data acquisition and sampling protocols and systems.

In the few next lines, we give a very brief introduction of compressed sensing in the monochannel case. Assume  $x \in \mathbb{R}^t$  such that we "observe" or "measure" only M < t samples  $\{y_k\}_{k=1,...,M}$ :  $y_k = \langle x, \theta_k \rangle$ . These measurements are more conveniently represented in a matrix formulation:

$$y = x \Theta \tag{30}$$

where  $\Theta$  is the  $t \times M$  measurement or *sensing* matrix. In [56], the authors showed that assuming x has a K-sparse representation in  $\Phi$  (in [56], the sparse representation is the Fourier domain) then x can be exactly recovered by solving the following linear problem:

$$\min_{\mathbf{y}} \| \mathbf{x} \mathbf{\Phi}^T \|_{\ell_1} \quad \text{s.t. } \mathbf{y} = \mathbf{x} \mathbf{\Theta}.$$
(31)

More precisely, Candès *et al.* showed in [56] that if  $M > CK \log(t)$  then the previous  $\ell_1$  minimization problem provides the exact signal x with  $C \simeq 22(\delta + 1)$  and probability of success  $1 - O(t^{-\delta})$ . In other words, this result defines a new *non-linear* sampling theorem as pointed out in the review paper [57].

Since the seminal work of Donoho *et al.* and Candès *et al.*, we have witnessed a flurry of research activity addressing theoretical and practical issues arising in CS, see [58–62] to name a few.<sup>5</sup>

The observed data y defined in (30) can be considered as subspace "projection" of the original data x. When the sensing matrix  $\Theta$  is a submatrix of the identity matrix I, the entries of the observed data y are then a subset of the entries of x. The sensing step is then equivalent to "masking" some entries and keeping the others: the decoding step (estimating x from y) is then equivalent to the inpainting problem. Inpainting can be viewed as a particular case of compressed sensing.

#### 4 Steps Ahead—Learning the Sparse Representation

# 4.1 Adaptive Multichannel Morphological Component Separation

Throughout this paper, we focused on accounting for both *spectral* and *spatial* coherences/structures to better solve in-

<sup>&</sup>lt;sup>5</sup>A website at http://www.dsp.ece.rice.edu/cs/ is dedicated to Compressed Sensing Resources.



Fig. 7 Restoring Mars-Orbiter hyperspectral data: recovering the spectra. (a) Original spectrum at pixel {10, 10}. (b) Masked spectrum—50% of the samples are missing. (c) Recovered spectrum using mMCA. *Abscissa*: channel number. *Ordinate*: spectrum amplitude



Fig. 8 Recovering the hyperspectral data: SNR in dB between the original data and the recovered hyperspectral data using mMCA (*solid line*) and MCA on each channel separately (*dashed line*)

verse problems such as inpainting or morphological component extraction issues. The choice of a particular multichannel representation relies on *prior* information. Recently, new sparsity driven approaches in signal recovery have focused on devising adaptive processes. Adapting the representation to the data has also been introduced in various fields. Most adaptive approaches are based on different concepts:

- Global: in various fields, adaptive schemes have been proposed to globally update the representation (see e.g. [63, 65]).
- Patch-based local: patch-based adaptive techniques have already been applied to restoration problems (see e.g. [64]).
- Adaptive search in tree-based bases: in the monochannel case, adaptive dictionary learning processes have been used (see e.g. [66, 67]) assuming that the sparse representation lies in a class of tree-based multiscale transforms (e.g. wavelet and cosine packets [8], bandlets [23] to cite only a few).

These sparsity-based adaptive techniques have provided astounding results in various fields. In this section, we introduced an adaptive version of the mMCA algorithm. In the multichannel case, such an adaptive recovery would have to be applied both on the spectral dictionary  $\Xi$  and the



Fig. 9 Adaptive multichannel morphological component separation: the  $256 \times 256$  original color image

spatial dictionary  $\Phi$ . Adapting the spatial dictionary to the data could be done by e.g choosing a decomposition tree assuming  $\Phi$  lies in a class of tree-based multiscale transforms. When applicable, the same adaptive process can be performed for  $\Xi$ .

The data X are made of 3 observed channels corresponding to each color layer (for instance red, green and blue) which cannot be strictly called spectra. Hopefully, the forthcoming results are still valid in a higher dimension problem. In the mMCA framework, D' = 1 and the data **X** are the linear combination of D multichannel morphological components. In this section, each channel of the data X is the linear combination of a texture part (assumed to be well sparsified by a Discrete Cosine Transform) and a contour part (sparsely represented by the curvelet tight frame). Figure 9 displays a toy-example based color image X. We then propose recovering the color morphological components using the proposed mMCA method which seeks to adapt the color space to the data **X**. In this context, we assume that  $\Xi$  is a 3  $\times$  3 invertible matrix. Adapting the spectral basis  $\Xi$  (*i.e.* the color space) to the data then amounts to estimate an "optimal" matrix  $\boldsymbol{\Xi}$ .

The mMCA algorithm is then adapted such that at each iteration h the matrix  $\Xi$  is updated by its least-squares estimate:

$$\mathbf{\Xi}^{(h+1)} = \arg\min_{\mathbf{\Xi}} \left\| \mathbf{X} - \mathbf{\Xi} \sum_{j=1}^{D} \varpi_{j}^{(h)} \mathbf{\Phi}_{j} \right\|^{2}.$$
 (32)

This problem has a unique minimizer defined as follows:

$$\mathbf{\Xi}^{(h+1)} = \mathbf{X} \left[ \sum_{j=1}^{D} \boldsymbol{\varpi}_{j}^{(h)} \mathbf{\Phi}_{j} \right]^{\dagger}$$
(33)

where  $\left[\sum_{j=1}^{D} \varpi_{j}^{(h)} \Phi_{j}\right]^{\dagger}$  is the pseudo-inverse of the matrix  $\sum_{j=1}^{D} \varpi_{j}^{(h)} \Phi_{j}$ .

The mMCA algorithm of Sect. 3.2 is then adapted as follows:

1. Set the number of iterations  $I_{\text{max}}$  and threshold  $\lambda^{(0)}$ .

2. While  $\lambda^{(h)}$  is higher than a given lower bound  $\lambda_{min}$  (e.g. can depend on the noise variance),

a. For j = 1, ..., D

• Compute the residual term  $\mathbf{R}_{j}^{(h)}$  assuming the current estimates of  $\varpi_{\{p\}\neq\{j\}}, \tilde{\varpi}_{p\neq i}^{(h-1)}$  are fixed:

$$\mathbf{R}_{j}^{(h)} = \mathbf{X} - \sum_{p \neq j} \tilde{\boldsymbol{\omega}}_{p}^{(h-1)}.$$

• Estimate the current coefficients of  $\tilde{\varpi}_{j}^{(h)}$  by thresholding with threshold  $\lambda^{(h)}$ :

$$\tilde{\alpha}_j^{(h)} = \Delta_{\lambda^{(h)}} (\boldsymbol{\Xi}^{(h)T} \mathbf{R}_j^{(h)} \boldsymbol{\Phi}_j^T).$$

• Get the new estimate of  $\omega_j$  by reconstructing from the selected coefficients  $\tilde{\alpha}_i^{(h)}$ :

$$\begin{split} \tilde{\boldsymbol{\varpi}}_{j}^{(h)} &= \boldsymbol{\Xi}^{(h)} \tilde{\boldsymbol{\alpha}}_{j}^{(h)} \boldsymbol{\Phi}_{j}. \\ \text{b. Update the spectral basis \boldsymbol{\Xi}:} \\ \boldsymbol{\Xi}^{(h+1)} &= \mathbf{Y}^{(h)} [\sum_{j=1}^{D} \boldsymbol{\varpi}_{j}^{(h)} \boldsymbol{\Phi}_{j}]^{\dagger}. \end{split}$$

 $\Xi^{(n+1)} = \mathbf{Y}^{(n)} [\sum_{j=1}^{j} \varpi_j^{(n)} \Phi_j]^{\dagger}.$ 3. Decrease the threshold  $\lambda^{(h)}$  following mMOM strategy.

The proposed **Adaptive** mMCA algorithm should be able to better account for inter-channel structure or correlations. In the next experiment, we compare a non-adaptive MCAbased algorithm with the new adaptive mMCA algorithm described below:

- *Non-adaptive approach*: a monochannel MCA is applied to each channel *separately*. It then amounts to applying the mMCA algorithm with the particular choice:  $\Xi = I$ .
- Adaptive approach: a global adaptive mMCA is applied to the whole data X.

The input toy-example image is displayed in Fig. 9. The texture part is composed of a globally oscillating pattern; it will be assumed to be well-sparsified by the DCT ( $\Phi_1$  with be the DCT-based basis).  $\Phi_2$  is the curvelet tight frame for the cartoon part. The decomposition results are illustrated in Fig. 10. Visually, both approaches (MCA and mMCA) seem to perform similarly providing good visual results. More quantitatively, Table 2 summarizes the respective recovery performances.

To conclude accounting for inter-channel structures greatly improves the decomposition task. In the next section, we apply a similar adaptive scheme to the multichannel inpainting problem.

#### J Math Imaging Vis

Fig. 10 Adaptive multichannel morphological component separation-components recovery: (a) Original color texture component. (b) Original color cartoon component. (c) Color texture component estimated with MCA. (d) Color cartoon component estimated with MCA. (e) Color texture component estimated with mMCA. (f) Color cartoon component estimated with mMCA



 Table 2
 Recovery signal-to-noise ratio

Recovery error SNR in dB	Texture part	Cartoon part	
mMCA with $\Xi = I$	15.7	22.9	
Adaptive mMCA	16.3	23.6	

## 4.2 Adaptive Color Image Inpainting

In the previous section, we emphasized on the improvement led by adapting the spectral dictionary to data. Hereafter, we consider the particular case of color image inpainting. The data X are assumed to be made of three channels (i.e. corresponding to each color layer). We apply exactly the same spectral dictionary update described in Sect. 4.1. In the context of inpainting, the mMCA algorithm of Sect. 3.2 is then adapted as follows:

1. Set the number of iterations  $\mathit{I}_{max}$  and threshold  $\lambda^{(0)}.$ 

2. While  $\lambda^{(h)}$  is higher than a given lower bound  $\lambda_{min}$  (e.g. can depend on the noise variance),

a. Compute  $\mathbf{Y}^{(h)} = \mathbf{Y} + \mathcal{M}^c \odot \tilde{\mathbf{X}}^{(h-1)}$ .

b. Initialize to zero each residual morphological components  $\{\tilde{\varpi}_i\}^{(h-1)}$ .

For 
$$j = 1, ..., D$$

• Compute the residual term  $\mathbf{R}_{j}^{(h)}$  assuming the current estimates of  $\varpi_{\{p\}\neq\{j\}}, \, \tilde{\sigma}_{p\neq j}^{(h-1)}$  are fixed:

Fig. 11 Recovering color images. (a) Original Barbara color image. (b) Masked image—90% of the color pixels are missing. (c) Inpainted image using the original MCA algorithm on each color channel. (d) Inpainted image using the adaptive mMCA algorithm



$$\begin{split} \mathbf{R}_{j}^{(h)} &= \mathbf{Y}^{(h)} - \sum_{p \neq j} \tilde{\varpi}_{p}^{(h-1)}. \\ \bullet \text{ Estimate the current coefficients of } \tilde{\varpi}_{j}^{(h)} \text{ by thresholding} \\ \text{with threshold } \lambda^{(h)}: \\ \tilde{\alpha}_{i}^{(h)} &= \Delta_{\lambda^{(h)}} (\boldsymbol{\Xi}^{(h)}^{T} \mathbf{R}_{i}^{(h)} \boldsymbol{\Phi}_{i}^{T}). \end{split}$$

• Get the new estimate of  $\overline{\varpi}_j$  by reconstructing from the selected coefficients  $\tilde{\alpha}_i^{(h)}$ :

$$\begin{split} \tilde{\varpi}_{j}^{(h)} &= \Xi^{(h)} \tilde{\alpha}_{j}^{(h)} \Phi_{j}. \\ \text{c. Update the hypercube } \tilde{\mathbf{X}}^{(h)} &= \sum_{j=1}^{D} \tilde{\varpi}_{j}^{(h)} \\ \text{d. Update the spectral basis } \Xi: \\ &\Xi^{(h+1)} = \mathbf{Y}^{(h)} [\sum_{j=1}^{D} \varpi_{j}^{(h)} \Phi_{j}]^{\dagger}. \end{split}$$

3. Decrease the threshold  $\lambda^{(h)}$  following mMOM strategy.

Figure 11(a) shows the original *Barbara* color image. Figure 11(b) depicts the masked color image where 90% of the color pixels are missing. Figure 11(c) portrays the recovered image using mMCA in the original RGB color space (which amounts to perform a monochannel MCA-based inpainting on each channel). Figure 11(d) shows the image recovered with the color space-adaptive mMCA algorithm. The zoom on the recovered images in Fig. 12 shows that adapting the color space avoids chromatic aberrations and hence produces a better visual result. This visual impression is quantitatively confirmed by SNR measures, where the color space-adaptive mMCA improves the SNR by 1 dB.

## 4.3 Relations with BSS

In the previous sections, the adaptive mMCA-based algorithm has been devised to adapt the spectral dictionary  $\Xi$  to the data. Recall that this adaptive scheme can be recast as a two-step iterative algorithm:

- *Sparse coding*: the first step amounts to get a sparse decomposition of the data **X** in the multichannel dictionary  $\Xi \otimes \Phi$ .
- Spectral dictionary update: update the spectral dictionary Ξ.

Such a two-step iterative algorithm have a flavor of the GMCA algorithm we proposed for solving Blind Source Separation problems in [68]. In this paper, we also showed that this kind of adaptive scheme is likely to provide sparser representations. Even if the notion of source is a non-sense (for instance in color imaging), it is always worth looking for sparser representations. Indeed, we emphasized on the mMCA's ability to adapt to data leads to better recovery results.

## 5 Conclusion

We recalled the tremendous effectiveness of sparsity-based methods in signal restoration. In this paper we empha-

#### J Math Imaging Vis

Fig. 12 Zoom on recovered Barbara color image.
(a) Original Barbara color image. (b) Masked image—90% of the color pixels are missing. (c) Inpainted image using the original MCA algorithm on each color channel.
(d) Inpainted image using the adaptive mMCA algorithm



size on multichannel sparse dictionaries to better represent multichannel data. Indeed, more than a concatenation of monochannel signals, multichannel data are spatially and spectrally structured. We introduce a general sparsity-based framework coined multichannel Morphological Component Analysis (mMCA) that accounts for the specific structure of multichannel data. We enlighten the links between mMCA and extensions of general sparse decomposition problems to the multichannel case. New theoretical results are put forward that prove the efficiency of mMCA in providing sparse multichannel decompositions in dictionaries built as a union of tensor products of orthonormal bases. Experiments are given showing that the mMCA framework provides an effective tool for devising sparsity-based solutions to some restoration problems such as multichannel morphological component separation and color image inpainting. We also extend the mMCA algorithm to solve hyperspectral data inpainting issues. An adaptive scheme is also proposed to adapt the sparse representation to data. We illustrate the astounding enhancement provided by such an adaptive algorithm. Future work will be devoted to i) extending the mMCA framework to deal with other inverse problems such as multichannel deconvolution, ii) devising more adaptive mMCA-based algorithm for multichannel data decomposition.

Acknowledgement The authors want to thank O. Forni for providing the *Mars Observer* hyperspectral data.

#### **Appendix: Proofs**

Let us first simplify a few notations. We define  $\gamma = \{j, k\}$  as a couple of indices. The multichannel subdictionary  $\Xi_k \otimes \Phi_j$  will be written  $\Psi_{\gamma}$  with  $\gamma = \{j, k\}$ . The notation  $\psi_{\gamma}[i]$ will refer to the *i*-th atom of the multichannel subdictionary  $\Psi_{\gamma}$ . Furthermore, the sparse representation of the multichannel data **X** will be written as follows:

$$\mathbf{X} = \sum_{\gamma} \sum_{i \in \Lambda_{\gamma}} \alpha_{\gamma}[i] \psi_{\gamma}[i].$$
(34)

Equivalently, the sparse decomposition of the residual term at iteration h is written as follows:

$$\mathbf{R}^{(h)} = \sum_{\gamma} \sum_{i \in \Lambda_{\gamma}} \beta_{\gamma}^{(h)}[i] \psi_{\gamma}[i].$$
(35)

Note that the following proofs are still valid in the monochannel case which corresponds to the MCA algorithm.

*Proof of Proposition 1* Let's denote:

$$(\gamma^{\star}, i^{\star}) = \underset{\gamma, i \in \Lambda_{\gamma}}{\operatorname{arg\,max}} |\langle \mathbf{R}^{(h)}, \psi_{\gamma}[i] \rangle|,$$

$$(\gamma^{\dagger}, i^{\dagger}) = \operatorname*{arg\,max}_{\gamma \neq \gamma^{\star}, i \in \Lambda_{\gamma}} |\langle \mathbf{R}^{(h)}, \psi_{\gamma}[i] \rangle|,$$

$$\begin{split} \omega_{\gamma^{\star}}^{(h)} &= |\langle \mathbf{R}^{(h)}, \psi_{\gamma^{\star}}[i^{\star}] \rangle|, \\ \omega^{(h)} &= |\langle \mathbf{R}^{(h)}, \psi_{\gamma^{\dagger}}[i^{\dagger}] \rangle|. \end{split}$$

Moreover, let  $\beta^* = |\beta_{\gamma^*}^{(h)}[i^*]|$  and  $\beta^{\dagger} = |\beta_{\gamma^{\dagger}}^{(h)}[i^{\dagger}]|$ . By definition, the mMOM thresholding rule in (24) selects new coefficients such that:

$$\forall i, \gamma; \quad \omega_{\gamma^{\star}}^{(h)} \ge |\langle \mathbf{R}^{(h)}, \psi_{\gamma}[i] \rangle| > \omega^{(h)}.$$

To show that mMCA/mMOM agrees with the exact selection property, it is sufficient to show that:

$$|\langle \mathbf{R}^{(h)}, \psi_{\gamma^{\star}}[i] \rangle| < \omega^{(h)}, \quad \forall i \in \Lambda_{\gamma^{\star}}^{c}$$
(36)

where  $\Lambda_{\gamma^{\star}}^{c}$  is the complement of the support  $\Lambda_{\gamma^{\star}}$  restricted to the subdictionary  $\Psi_{\gamma^{\star}}$ .

On the one hand, we have:

$$\omega^{(h)} \ge \beta^{\dagger} - \sum_{p \neq \gamma^{\dagger}} \sum_{j \in \Lambda_{p}} |\beta_{p}^{(h)}[j]| |\langle \psi_{\gamma^{\dagger}}[i^{\dagger}], \psi_{p}[j] \rangle|$$
$$\ge \rho \beta^{\star} - K \beta^{\star} \mu_{\Psi}.$$
(37)

Here we used  $|\langle \psi_{\gamma^{\star}}[i], \psi_{p}[j] \rangle| \leq \mu_{\Psi}$ . We also supposed that  $\beta^{\dagger} = \rho \beta^{\star}$  for  $0 < \rho \leq 1$ . One can easily prove that this latter condition is sufficient for  $\omega_{\gamma^{\star}}^{(h)} > \omega^{(h)}$  to hold.

On the other hand, a straightforward calculation leads to:

$$\begin{aligned} |\langle \mathbf{R}^{(h)}, \psi_{\gamma^{\star}}[i] \rangle| \\ &\leq \sum_{p \neq \gamma^{\star}} \sum_{j \in \Lambda_{p}} |\beta_{p}^{(h)}[j]| |\langle \psi_{\gamma^{\star}}[i], \psi_{p}[j] \rangle|, \quad \forall i \in \Lambda_{\gamma^{\star}}^{c} \\ &\leq K \beta^{\dagger} \mu_{\Psi} \\ &\leq K \rho \beta^{\star} \mu_{\Psi} \end{aligned}$$
(38)

where all coefficients of  $\gamma^*$ -th component vanished because  $\Psi_{\gamma^*}$  is orthonormal.

From (37) and (38), we get that for (36) to hold, we need that:

$$\rho - K\mu_{\Psi} > K\rho\mu_{\Psi}$$

or equivalently  $K < \mu_{\Psi}^{-1} \frac{\rho}{1+\rho}$ . For  $0 < \rho \le 1$ ,  $\frac{\rho}{1+\rho}$  is a strictly increasing function bounded above by 1/2, yielding the bound of the proposition on *K*. This leads to (36) validating the Exact Selection Property (ESP).

*Proof of Proposition* 2 The proof is similar to that of MP and OMP [18]. In fact, our mMCA/mMOM strategy and its proof is a stagewise extension of the greedy MP. mMCA/mMOM begins by setting  $\mathbf{R}^{(0)} = \mathbf{X}$ . By hypothesis in the proposition, at each step  $h \ge 0$  and component  $\gamma^*$ ,

mMCA/mMOM chooses correct atoms from  $\Lambda_{\gamma^*}$  by hard-thresholding, and then calculates a new residual such that:

$$\mathbf{R}^{(h)} = \sum_{i \in I_{\gamma^{\star}}} \langle \mathbf{R}^{(h)}, \psi_{\gamma^{\star}}[i] \rangle \psi_{\gamma^{\star}}[i] + \mathbf{R}^{(h+1)}$$

where  $I_{\gamma^{\star}} \subseteq \Lambda_{\gamma^{\star}}$ , and  $\Psi_{I_{\gamma^{\star}}}$  is the restriction of  $\Psi$  to the atoms indexed by  $I_{\gamma^{\star}}$ . From the orthonormality of the atoms in  $\Psi_{I_{\gamma^{\star}}}$ , it is easy to see that,

$$\|\mathbf{R}^{(h+1)}\|^{2} = \|\mathbf{R}^{(h)}\|^{2} - \left\|\sum_{i \in I_{\gamma^{\star}}} \langle \mathbf{R}^{(h)}, \psi_{\gamma^{\star}}[i] \rangle \psi_{\gamma^{\star}}[i]\right\|^{2} \le \sigma^{2} \|\mathbf{R}^{(h)}\|^{2}$$
(39)

where  $\sigma^2$  is the highest eigenvalue of  $\mathbf{I} - \mathbf{G}_{\Psi_{I_{\gamma^*}}}$ ,<sup>6</sup> **I** is the identity matrix of appropriate dimensions. From [18, Lemma 2.3] with arguments relying on Gershgorin Disc Theorem, and [18, Proposition 2.1], the squared singular values of  $\Psi_{I_{\gamma^*}}$  are bounded below by  $1 - (\operatorname{Card}(I_{\gamma^*}) - 1)\mu_{\Psi}$ . It follows that  $\sigma^2 \leq (\operatorname{Card}(I_{\gamma^*}) - 1)\mu_{\Psi}$ . Then, as  $K\mu_{\Psi} < 1/2$ ,

$$\|\mathbf{R}^{(h+1)}\|^{2} \leq (\operatorname{Card}(I_{\gamma^{\star}}) - 1)\mu_{\Psi} \|\mathbf{R}^{(h)}\|^{2}$$
  
$$\leq (K - 1)\mu_{\Psi} \|\mathbf{R}^{(h)}\|^{2}$$
  
$$\leq 1/2 \|\mathbf{R}^{(h)}\|^{2}$$
  
$$\leq \cdots$$
  
$$\leq 2^{-(h+1)} \|\mathbf{X}\|^{2}$$
(40)

which completes the proof.

#### References

- Starck, J.L., Candès, E., Donoho, D.: The curvelet transform for image denoising. IEEE Trans. Image Process. 11(6), 670–684 (2002)
- Starck, J.L., Elad, M., Donoho, D.: Image decomposition via the combination of sparse representation and a variational approach. IEEE Trans. Image Process. 14(10), 1570–1582 (2005)
- Li, Y., Amari, S., Cichocki, A., Guan, C.: Underdetermined blind source separation based on sparse representation. IEEE Trans. Inf. Theory 52, 3139–3152 (2006)
- Zibulevsky, M., Pearlmutter, B.: Blind source separation by sparse decomposition. Neural Comput. 13(4) (2001)
- Bobin, J., Moudden, Y., Starck, J.L., Elad, M.: Morphological diversity and source separation. IEEE Signal Process. Lett. 13(7), 409–412 (2006)
- Olshausen, B., Field, D.: Sparse coding with an overcomplete basis set: a strategy employed by v1? Vis. Res. 37, 3311–3325 (2006)
- Demanet, L., Ying, L.: Wave atoms and sparsity of oscillatory patterns Appl. Comput. Harmon. Anal. 23(3), 368–387 (2007)

<sup>&</sup>lt;sup>6</sup>Recall that  $\mathbf{G}_{\Psi_{I,\star}}$  is the Gram matrix of the subdictionary  $\Psi_{I_{\gamma^{\star}}}$ .

- Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press, New York (1998)
- 9. Candès, E., Donoho, D.: Curvelets. Technical report, Statistics, Stanford University (1999)
- Candès, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. SIAM Multiscale Model. Simul. 5/3, 861– 899 (2006)
- Cotter, S., Rao, B., Engan, K., Kreutz-Delgado, K.: Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans. Signal Process. 53, 2477–2488 (2005)
- Fornasier, M., Rauhut, H.: Recovery algorithms for vector valued data with joint sparsity constraints. Preprint, available at http:// www.dsp.ece.rice.edu/cs/ (2006)
- Tropp, J., Gilbert, A., Strauss, M.: Algorithms for simultaneous sparse approximation, part i: greedy pursuit. Signal Process. 86, 572–588 (2006). Special issue "Sparse approximations in signal and image processing"
- Tropp, J.: Algorithms for simultaneous sparse approximation, part ii: convex relaxation. Signal Process. 86, 589–602 (2006). Special issue "Sparse approximations in signal and image processing"
- Chen, J., Huo, X.: Sparse representations for multiple measurement vectors (MMV) in an over-complete dictionary. In: ICASSP '05 (2005)
- Starck, J.L., Elad, M., Donoho, D.: Redundant multiscale transforms and their application for morphological component analysis. Adv. Imaging Electron. Phys. 132, 287–348 (2004)
- Elad, M., Starck, J.L., Donoho, D., Querre, P.: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). Appl. Comput. Harmon. Anal. 19(3), 340–358 (2005)
- Tropp, T.: Greed is good: algorithmic results for sparse approximation. IEEE Trans. Inf. Theory 50(10), 2231–2242 (2004)
- Donoho, D., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inf. Theory 47(7), 2845–2862 (2001)
- Gribonval, R., Nielsen, M.: Sparse representations in unions of bases. IEEE Trans. Inf. Theory 49(12), 3320–3325 (2003)
- Candès, E., Donoho, D.: Ridgelets: the key to high dimensional intermittency? Philos. Trans. R. Soc. Lond. A 357, 2495–2509 (1999)
- Starck, J.L., Candès, E., Donoho, D.: The curvelet transform for image denoising. IEEE Trans. Image Process. 11(6), 131–141 (2002)
- LePennec, E., Mallat, S.: Sparse geometric image representations with bandelets. IEEE Trans. Image Process. 14(4), 423–438 (2005)
- Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Process. 14(12), 2091–2106 (2005)
- Chen, S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. 20(1), 33–61 (1999)
- Donoho, D., Elad, M.: Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ<sup>1</sup> minimization. Proc. Natl. Acad. Sci. 100, 2197–2202 (2003)
- 27. Bruckstein, A., Elad, M.: A generalized uncertainty principle and sparse representation in pairs of  $\mathbb{R}^N$  bases. IEEE Trans. Inf. Theory **48**, 2558–2567 (2002)
- Fuchs, J.J.: On sparse representations in arbitrary redundant bases. IEEE Trans. Inf. Theory 50(6), 1341–1344 (2004)
- Bruckstein, A., Donoho, D., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. (2007, to appear)
- Bobin, J., Starck, J.L., Fadili, J., Moudden, Y., Donoho, D.: Morphological component analysis: an adaptive thresholding strategy. IEEE Trans. Image Process. 16(11), 2675–2681 (2007)
- Feuer, A., Nemirovsky, A.: On sparse representation in pairs of bases. IEEE Trans. Inf. Theory 49(6), 1579–1581 (2003)

- Gribonval, R., Nielsen, M.: Beyond sparsity: recovering structured representations by l<sub>1</sub>-minimization and greedy algorithms. Adv. Comput. Math. 28(1) (2008)
- Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Trans. Signal Process. 41(12), 3397–3415 (1993)
- 34. Donoho, D., Tsaig, Y.: Fast solution of  $\ell_1$  minimization problems when the solution may be sparse (2006, submitted)
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. 32(2), 407–499 (2004)
- Osborne, M.R., Presnell, B., Turlach, B.A.: A new approach to variable selection in least squares problems. IMA J. Numer. Anal. 20(3), 389–403 (2000)
- Malioutov, D.M., Cetin, M., Willsky, A.S.: Homotopy continuation for sparse signal representation. In: ICASSP '05, vol. 5, pp. 733–736 (2005)
- Plumbley, M.: Recovery of sparse representations by polytope faces pursuit. In: ICA06, pp. 206–213 (2006)
- Donoho, D., Tsaig, Y., Drori, I., Starck, J.L.: Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. IEEE Trans. Inf. Theory (2006, submitted)
- Sardy, S., Bruce, A., Tseng, P.: Block coordinate relaxation methods for nonparametric wavelet denoising. J. Comput. Graph. Stat. 9(2), 361–379 (2000)
- Donoho, D., Elad, M., Temlyakov, V.: Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inf. Theory 52, 6–18 (2006)
- Fuchs, J.J.: Recovery conditions of sparse representations in the presence of noise. In: ICASSP '06, vol. 3, pp. 337–340 (2006)
- Tropp, T.: Just relax: convex programming methods for subset selection and sparse approximation. IEEE Trans. Inf. Theory 52(3), 1030–1051 (2006)
- Elad, M.: Why simple shrinkage is still relevant for redundant representations? IEEE Trans. Inf. Theory 52(12), 5559–5569 (2006)
- Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forwardbackward split ting. SIAM J. Multiscale Model. Simul. 4(4), 1168–1200 (2005)
- Chan, T., Shen, J.: Mathematical models for local non-texture inpainting. SIAM J. Appl. Math. 62(3), 1019–1043 (2001)
- Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of SIGGRAPH 2000, New Orleans, USA (2000)
- Masnou, S., Morel, J.M.: Disocclusion: a variational approach using level lines. IEEE Trans. Image Process. 11(2), 68–76 (2002)
- Tschumperlé, D., Deriche, R.: Vector-valued image regularization with PDE's: a common framework for different applications. In: CVPR'2003, Computer Vision and Pattern Recognition, Madison/USA (2003)
- Demanet, L., Song, B., Chan, T.: Image inpainting by correspondence maps: a deterministic approach. In: Proc. VLSM Conf., Nice, France (2003)
- Guleryuz, O.: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising—part I: theory. IEEE Trans. Image Process. 15(3), 539–554 (2006)
- Guleryuz, O.: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising part II: adaptive algorithms. IEEE Trans. Image Process. 15(3), 555–571 (2006)
- Fadili, J., Starck, J.L.: EM algorithm for sparse representation based image inpainting. In: IEEE International Conference on Image Processing ICIP'05, vol. 2, pp. 61–63, Genoa, Italia (2005)
- Fadili, J., Starck, J.L., Murtagh, F.: Inpainting and zooming using sparse representations. Comput. J. (2006, in press)
- Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006)
- Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory 52(2), 489–509 (2006)

- 57. Candès, E.: Compressive sampling. In: International Congress of Mathematics, Madrid (2006)
- Donoho, D., Tsaig, Y.: Extensions of compressed sensing. Signal Process. 86(3), 5433–5548 (2006)
- Candès, E., Tao, T.: Statistical estimation when p is much larger than n. Ann. Stat. 52(9), 4036–4048 (2006). Available at http://www.dsp.ece.rice.edu/cs/
- Candès, E., Tao, T.: Near optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inf. Theory 52(12), 5406–5425 (2006)
- Candès, E., Romberg, J.: Practical signal recovery from random projections. Preprint, available at http://www.dsp.ece.rice.edu/cs/ (2005)
- Tropp, J., Gilbert, A.: Signal recovery from partial information via orthogonal matching pursuit. Preprint, available at http://www. dsp.ece.rice.edu/cs/ (2005)
- Aharon, M., Elad, M., Bruckstein, A.: k-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. 54(11), 4311–4322 (2006)
- Salee, P., Olshausen, B.: Learning sparse multiscale image representations. In: Advances in Neural Information Processing Systems (2003)
- Plumbley, M.D.: Dictionary learning for 11-exact sparse coding. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) Independent Component Analysis and Signal Separation. Lecture Notes in Computer Science, vol. 4666, pp. 406–413. Springer, Berlin (2007)
- Plumbley, M.D.: Tree-based orthogonal matching pursuit algorithm for signal reconstruction. In: La, C., Do, M.N. (eds.) Proc. of IEEE International Conference on Image Processing (2006)
- 67. Peyré, G.: Best basis compressed sensing. In: SSVM (2006). Preprint, available at http://www.dsp.ece.rice.edu/cs/
- Bobin, J., Starck, J.-L., Fadili, J., Moudden, Y.: Sparsity and morphological diversity in blind source separation. IEEE Trans. Image Process. 16(11), 2662–2674 (2007)



**J. Bobin** graduated from the Ecole Normale Superieure (ENS) de Cachan, France, in 2005 and received the M.Sc. degree in signal and image processing from ENS Cachan and Université Paris XI, Orsay, France. He received the Agrégation de Physique in 2004. Since 2005, he is pursuing his Ph.D. with J-L. Starck at the CEA. His research interests include statistics, information theory, multiscale methods and sparse representations in signal and image processing.



**Y. Moudden** graduated in electrical engineering from SUPELEC, Gif-sur-Yvette, France and obtained a M.S. 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 XI, Orsay, France. He was a visitor at UCLA in 2004 and and is currently with the CEA, Saclay, France, working on applications of signal processing to astronomy. His research interests include signal and

image processing, data analysis, statistics and information theory.



**J. Fadili** graduated from the Ecole Nationale Supérieure d'Ingénieurs (ENSI) de Caen, France, in 1996 and received the M.Sc. and Ph.D. degrees in signal and image processing from Caen University. He was a Research Associate with the University of Cambridge (MacDonnel-Pew Fellow), Cambridge, U.K., from 1999 to 2000. He has been an Associate Professor of signal and image processing since September 2001 at ENSI. He was a Visitor at the Queensland Uni-

versity of Technology, Brisbane, Australia and Stanford University CA, USA in 2006. His research interests include statistical approaches in signal and image processing, inverse problems in image processing, multiscale methods and sparse representations in signal and image processing. Areas of application include medical and astronomical imaging.



**J.-L. Starck** Jean-Luc Starck has a Ph.D. from University Nice-Sophia Antipolis and an Habilitation from University Paris XI. He was a visitor at the European Southern Observatory (ESO) in 1993, at UCLA in 2004 and at Stanford's statistics department in 2000 and 2005. He has been a Researcher at CEA since 1994. His research interests include image processing, statistical methods in astrophysics and cosmology. He is an expert in multiscale methods such

wavelets and curvelets, He is leader of the project Multiresolution at CEA and he is a core team member of the PLANCK ESA project. He has published more than 200 papers in different areas in scientific journals and conference proceedings. He is also author of two books entitled Image Processing and Data Analysis: the Multiscale Approach (Cambridge University Press, 1998), and Astronomical Image and Data Analysis (Springer, 2nd edition, 2006).