Morphological Diversity and Sparsity for Multichannel Data Restoration

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

Morphological Diversity and Sparsity for Multichannel Data Restoration

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

[1] jerome.bobin@cea.fr, ymoudden@cea.fr, jstarck@cea.fr - CEA-DAPNIA/SEDI, Service d'Astrophysique, CEA/Saclay, 91191 Gif sur Yvette, France

[2] jalal.fadili@greyc.ensicaen.fr - GREYC CNRS UMR 6072, Image Processing Group, ENSICAEN 14050, Caen Cedex, France

This version contains a corrigendum to the proofs in the previous version of the paper

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 a set of multichannel inverse problems such as morphological component decomposition and inpainting.

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 colour layer in colour 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 $\{x_i\}_{i=1,\cdots,m}$. For convenience, those m vectors are stacked in a $m \times t$ matrix $\mathbf{X} = \left[x_1^T \cdots x_m^T\right]^T$. In a wide range of applications, the data $\mathbf{X}$ 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 modelled as follows :

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

where $\mathcal{F}$ is the *degradation* mapping, $\mathbf{X}$ is the original multichannel data to be recovered and $\mathbf{N}$ 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* $\{\phi_i\}_{i=1,\cdots,T}$: $x = \sum_{i=1} \alpha_i \phi_i$, where $\alpha_i = \langle \phi_i, x \rangle$ are called the decomposition coefficients of x in the dictionary $\mathbf{\Phi} = [\phi_1^T, \cdots, \phi_T^T]^T$ (the $T \times t$ matrix whose lines are the atoms normalized to a unit ℓ_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 Λ_x, the *support* of x in $\mathbf{\Phi}$, as the set of indices of the non-zero entries in α :

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

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 Equation (1), a first and somewhat naive approach would consist in treating separately m single-channel restoration problems[3]. However, this is clearly suboptimal since, as we pointed out earlier, the m single channel observations share some *joint structure* requiring a coherent processing of the m channels simultaneously. For instance, the colour layers of colour images have often similar patterns. Thus recovering each observation separately is far from being optimal. Such *inter-observation* structures have to be properly modelled in order to enhance multichannel data restoration. A first sparsity-based solution consists in assuming that each channel must share the same sparsity

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

pattern *i.e.* the same support in $\mathbf{\Phi}$. 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 modelled 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 1 introduces a new sparsity-based framework for analyzing multichannel signals coined multichannel Morphological Component Analysis (mMCA). In Section 1.2 we enlighten the connections between mMCA and other extensions of sparse decomposition problems to the multichannel case. Section 1.3 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 Section 2.1 and inpainting in Section 2.2. We also put forward a mMCA-based adaptive restoration algorithm to better match the sparse representation to the multichannel data at hand.

Notations and definitions

We here define some useful quantities and notations. In the multichannel case, the data $\mathbf{X}$ 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 $\mathbf{X} \in \underbrace{\mathbb{R}^{t_1} \otimes \cdots \otimes \mathbb{R}^{t_p}}_{p}$ with $p \geq 2$. Multichannel data often consist of m "observations" $\{x_i\}_{i=1,\cdots,m}$ from m channels, each lying in $\mathbb{R}^t$. For convenience, we use the following matrix notation:

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

where each channel $\{x_i\}_{i=1,\cdots,m}$ 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 $\mathbf{\Xi}$) and a basis of $\mathbb{R}^{1,t}$ (say $\mathbf{\Phi}$). Projecting $\mathbf{X}$ onto each atom of the multichannel basis $\mathbf{\Psi} = \mathbf{\Xi} \otimes \mathbf{\Phi}$ 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 $\mathbf{\Psi}$. This atom can be written in matrix form as : $\psi_{ij} = \xi_i \phi_j$ which is an $m \times t$ rank-1 matrix. If $\mathbf{\Xi}$ or $\mathbf{\Phi}$ is orthonormal then $\mathbf{\Psi}$ is also orthonormal. In fact, the scalar product between two multichannel atoms is such that [4]:

$$\left\langle \psi_{ip}, \psi_{jq} \right\rangle = \left\langle \xi_i, \xi_j \right\rangle \left\langle \phi_p, \phi_q \right\rangle \tag{5}$$

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

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

is a measure of how its atoms *look like each other*. Handling subsets of elements that belong to $\mathbf{\Psi}$ will also be needed. Let $\mathit{\Lambda} = \{\{i_1, j_1\}, \cdots, \{i_T, j_T\}\}$ be a set of index couples. The active subdictionary obtained by restricting the dictionary $\mathbf{\Psi}$ to the atoms whose indices are the elements of $\mathit{\Lambda}$ is written $\mathbf{\Psi}_{\mathit{\Lambda}}$. This notation will be useful to define the support of a signal $\mathbf{X}$ in $\mathbf{\Psi}$. Assume that $\mathbf{X}$ is K-sparse in $\mathbf{\Psi}$ then $\mathbf{X} = \sum_{\{i,j\} \in \mathit{\Lambda}_x} \alpha_{i,j} \psi_{ij} = \sum_{\{i,j\} \in \mathit{\Lambda}_x} \alpha_{i,j} \xi_i \phi_j$ where $\mathit{\Lambda}_x$ is the support of $\mathbf{X}$, and $\mathbf{\Psi}_{\mathit{\Lambda}_x}$ the corresponding active subdictionary.
According to the definition of the scalar product between two multichannel atoms given in Equation (5), the mutual coherence for multichannel dictionaries $\mathbf{\Psi} = \mathbf{\Xi} \otimes \mathbf{\Phi}$ is as follows :

$$0 \leq \mu_{\mathbf{\Psi}} = \max\left\{\mu_{\mathbf{\Xi}}, \mu_{\mathbf{\Phi}}\right\} < 1 \tag{7}$$

for orthonormal $\mathbf{\Xi}$ and $\mathbf{\Phi}$. In the next, the Frobenius norm of a matrix $\mathbf{X}$ is $\|\mathbf{X}\|^2 = \text{ Trace}\left(\mathbf{X}^T \mathbf{X}\right)$. The ℓ_1 norm of $\mathbf{X}$ is defined as the sum of the absolute values of the entries of matrix $\mathbf{X}$.

1 Morphological Component Analysis For Multichannel Data

1.1 Morphological Diversity and Morphological Component Analysis

An introduction to morphological diversity : A monochannel signal x is said to be sparse in a waveform dictionary $\mathbf{\Phi}$ if it can be well represented from a few dictionary elements. More precisely, let us define α such that :

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

The entries of α are commonly called "coefficients" of x in $\mathbf{\Phi}$. In that setting, x is said to be sparse in $\mathbf{\Phi}$ if most entries of α are nearly zero and only a few have "significant" amplitudes. Particular ℓ_0-sparse signals are generated from a few non-zero dictionary elements. Note that this notion of sparsity is strongly

[4] 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, $\mathbf{G}_{\mathbf{\Psi}} = \mathbf{G}_{\mathbf{\Xi}} \otimes \mathbf{G}_{\mathbf{\Phi}}$.

dependent on the dictionary $\mathbf{\Phi}$; 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 $\mathbf{\Phi}$ is overcomplete (*i.e.* $T > t$), the solution of Equation (8) is generally not unique. In that case, the authors of [19] were the first to seek the sparsest α, in terms of ℓ_0-pseudo-norm, such that $x = \alpha\mathbf{\Phi}$. This approach leads to the following minimization problem :

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

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 ℓ_1 norm for the ℓ_0 pseudo-norm leading to the following linear program :

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

This problem can be solved for instance using interior-point 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, 26, 20, 27, 28]. See also [29] for an extensive review.

In [2] and [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:

$$x = \sum_{i=1}^{D} \varphi_i = \sum_{i=1}^{D} \alpha_i \mathbf{\Phi}_i \tag{11}$$

where $\{\mathbf{\Phi}_i\}_{i=1,\cdots,D}$ are orthonormal bases of $\mathbb{R}^t$. Morphological diversity then relies on the sparsity of those morphological components in specific bases. In terms of ℓ_0 norm, this morphological diversity can be formulated as follows:

$$\forall\{i,j\} \in \{1,\cdots,D\}; \quad j \neq i \Rightarrow \|\varphi_i \mathbf{\Phi}_i^T\|_0 < \|\varphi_i \mathbf{\Phi}_j^T\|_0 \tag{12}$$

In other words, MCA relies on the incoherence between the sub-dictionaries $\{\mathbf{\Phi}_i\}_{i=1,\cdots,D}$ to estimate the morphological components $\{\varphi_i\}_{i=1,\cdots,D}$ by solving the following convex minimization problem:

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

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.

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 $\{x_i\}_{i=1,\cdots,m}$ is the linear combination of D morphological components:

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

where each morphological component φ_{ij} is sparse in a specific basis $\mathbf{\Phi}_j$. Then each channel $\{x_i\}_{i=1,\cdots,m}$ is assumed to be sparse in the overcomplete dictionary $\mathbf{\Phi}$ made of the union of the D bases $\{\mathbf{\Phi}_i\}_{i=1,\cdots,D}$.
We further assume that each column of the data matrix $\mathbf{X}$ is sparse in the dictionary $\mathbf{\Xi}$ made of the union of D' bases $\{\mathbf{\Xi}_i\}_{i=1,\cdots,D'}$ to account for inter-channel structures. The multichannel data $\mathbf{X}$ are then assumed to be sparse in the multichannel dictionary $\mathbf{\Psi} = [\mathbf{\Xi_1} \cdots \mathbf{\Xi_{D'}}] \otimes [\mathbf{\Phi_1} \cdots \mathbf{\Phi_D}]$. The multichannel data are then modelled as the linear combination of $D \times D'$ multichannel morphological components:

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

where ϖ_{jk} is sparse in $\mathbf{\Xi}_k \otimes \mathbf{\Phi}_j$. In this setting, separating two multichannel morphological components ϖ_{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 $\mathbf{\Xi}_p \otimes \mathbf{\Phi}_i$ and $\mathbf{\Xi}_p \otimes \mathbf{\Phi}_j$ (resp. $\mathbf{\Xi}_p \otimes \mathbf{\Phi}_i$ and $\mathbf{\Xi}_q \otimes \mathbf{\Phi}_i$) is upper-bounded by $\mu_{\mathbf{\Phi}}$ (resp. $\mu_{\mathbf{\Xi}}$).
- 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 $\mathbf{\Xi}_p \otimes \mathbf{\Phi}_i$ and $\mathbf{\Xi}_q \otimes \mathbf{\Phi}_j$ is lower than $\mu_{\mathbf{\Xi}} \mu_{\mathbf{\Phi}} \leq \max\{\mu_{\mathbf{\Xi}}, \mu_{\mathbf{\Phi}}\}$.

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 :

$$\min_{\{\varpi_{jk}\}} \left\| \mathbf{X} - \sum_{j=1}^{D} \sum_{k=1}^{D'} \varpi_{jk} \right\|_2^2 + 2\lambda \sum_{j=1}^{D} \sum_{k=1}^{D'} \|\mathbf{\Xi}_k^T \varpi_{jk} \mathbf{\Phi}_j^T\|_1 \tag{16}$$

In Section 1.2 we enlighten the connections between the problem in Equation (16) and the extension of BPDN to the multichannel case. We also provide straightforward multichannel extensions of well-known recovery results. In Section 1.3 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 Equation (16).

1.2 Multichannel overcomplete sparse recovery

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

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

where α is an $M \times T$ matrix and $\|\alpha\|_{\ell_0}$ refers to the number of non-zero entries in α. The convex ℓ_1 minimization problem (10) can also be rewritten in the multichannel setting :

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

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 $\mu_{\mathbf{\Psi}}$, the two aforementioned problems share the same unique solution when the following condition holds [19]:

$$K < \frac{1}{2} \left(1 + \frac{1}{\mu_{\mathbf{\Psi}}} \right)$$

The uniqueness and equivalence condition of the sparse multichannel decomposition problem in Equation (17) is then similar to the monochannel case. Assume that $\mathbf{X}$ is K-sparse in the multichannel dictionary $\mathbf{\Psi} = \mathbf{\Xi} \otimes \mathbf{\Phi}$. The ℓ_0 sparse decomposition problem in Equation (17) has a unique solution and problems in Equation (17) and (18) are equivalent when :

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

In this framework, most results in the monochannel case [27, 20, 28, 31, 18, 26] can be straightforwardly extended to the multichannel case. Previous work on multichannel sparse decomposition includes [32] which introduced the concept of multichannel dictionary.

1.3 Multichannel Morphological Component Analysis

The problem at stake in Equation (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 ℓ_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 ℓ_0 sparse recovery problem in Equation (9) with random dictionaries under mild conditions.

Owing to the particular structure of the problem in Equation (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 $\mathbf{X}$ are assumed to be the linear combination of $D \times D'$ morphological components $\{\varpi_{jk}\}_{j=1,\cdots,D;k=1,\cdots,D'}$. We define Λ_{jk} as the *support* (*i.e.* the indices of active atoms) of ϖ_{jk} in the subdictionary $\mathbf{\Psi}_{jk} = \mathbf{\Xi}_k \otimes \mathbf{\Phi}_j$. As $\mathbf{X}$ is K-sparse in the whole dictionary, $\sum_{j,k} \text{Card}(\Lambda_{jk}) = K$. The data can be decomposed as follows:

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

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

$$\alpha_{jk} = \arg\min_{\alpha_{jk}} \|\mathbf{R}_{jk} - \mathbf{\Xi}_k\ \alpha_{jk}\mathbf{\Phi}_j\|^2 + 2\lambda\|\alpha_{jk}\|_{\ell_1} \tag{20}$$

where $\mathbf{R}_{jk} = \mathbf{X} - \sum_{p,q\neq j,k} \mathbf{\Xi}_q\alpha_{pq}\mathbf{\Phi}_p$ is a residual term.
Since we are assuming that the subdictionaries $\{\mathbf{\Phi}_j\}_j$ and $\{\mathbf{\Xi}_k\}_k$ are orthonormal, the problem in Equation (20) is equivalent to the following:

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

which has a unique solution $\alpha_{jk} = \Delta_\lambda \left(\mathbf{\Xi}_k^T \mathbf{R}_k \mathbf{\Phi}_j^T \right)$ known as soft-thresholding with threshold λ as follows:

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

For a fixed λ, mMCA selects groups of atoms based on their scalar product with the residual $\mathbf{R}_{jk}$. Assuming that we select only the most coherent atom (with the highest scalar product) with the residual $\mathbf{R}_{jk}$ 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 λ at each iteration. The mMCA algorithm is summarized below:

1. Set the number of iterations $I_{\max}$ 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, see Section 1.5),
   For $j = 1, \cdots, D$ and $k = 1, \cdots, D'$
   - 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:
     $\mathbf{R}_{jk}^{(h)} = \mathbf{X} - \sum_{pq \neq jk} \tilde{\varpi}_{pq \neq jk}^{(h-1)}$.
   - Estimate the current coefficients of $\tilde{\varpi}_{jk}^{(h)}$ by thresholding with threshold $\lambda^{(h)}$:
     $\tilde{\alpha}_{jk}^{(h)} = \Delta_{\lambda^{(h)}} \left(\mathbf{\Xi}_k^T \mathbf{R}_{jk}^{(h)} \mathbf{\Phi}_j^T \right)$.
   - Get the new estimate of ϖ_{jk} by reconstructing from the selected coefficients $\tilde{\alpha}_{jk}^{(h)}$:
     $\tilde{\varpi}_{jk}^{(h)} = \mathbf{\Xi}_k \tilde{\alpha}_k^{(h)} \mathbf{\Phi}_j$.
3. Decrease the threshold $\lambda^{(h)}$ following a given strategy.

The thresholding strategy : In [30] we proposed a thresholding strategy that is likely to provide the solution to the ℓ_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 :

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

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

$$\lambda^{(h)} = \frac{1}{2} \left\{ m_{j_0 k_0}^{(h-1)} + m_{j_1 k_1}^{(h-1)} \right\} \tag{24}$$

In the next section, we show conditions under which mMCA/mMOM selects atoms without error and converges asymptotically to the solution of the multichannel ℓ_0 sparse recovery problem in Equation (17).

1.4 Recovering sparse multichannel decompositions using mMCA

The mMOM rule defined in Equation (23)-(24) is such that mMCA will select, at each iteration, atoms belonging to the same subdictionary $\mathbf{\Psi}_{jk} = \mathbf{\Xi}_k \otimes \mathbf{\Phi}_j$. 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 ℓ_0 sparse recovery problem (Exact Selection Property), ii) mMCA/mMOM converges exponentially to $\mathbf{X}$ and its sparsest representation in $\mathbf{\Psi}$. Let's mention that the mMCA/mMOM exhibits an auto-stopping behaviour, and requires only one parameter $\lambda_{\min}$ whose choice is easy and discussed in Section 1.5.
The next proposition states that mMCA/mMOM verifies the Exact Selection Property (ESP) at each iteration.

Proposition 1 (Exact Selection Property). *Suppose that* $\mathbf{X}$ *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 $K = \sum_{j,k} Card(\Lambda_{jk})$ *satisfying* $K < \frac{\mu_{\Psi}^{-1}}{2}$*. At the* h*-th iteration, assume that the residual* $\mathbf{R}^{(h)}$ *is* K*-sparse such that :*

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

Then mMCA/mMOM picks up coefficients belonging to the support of $\mathbf{X}$ *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 $\mathbf{X}$ and its sparsest representation in $\mathbf{\Psi} = [\mathbf{\Xi}_1 \cdots \mathbf{\Xi}_{D'}] \otimes [\mathbf{\Phi}_1 \cdots \mathbf{\Phi}_D]$.

Proposition 2 (Convergence). *Suppose that* $\mathbf{X}$ *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 $K = \sum_{j,k} Card(\Lambda_{jk})$*.*
If $K < \frac{\mu_{\Psi}^{-1}}{2}$ *then mMCA/mMOM converges exponentially to* $\mathbf{X}$ *and its sparsest representation in* $\mathbf{\Psi}$*. 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.

1.5 Handling bounded noise with mMCA

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

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

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

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

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 ℓ_2 ball centered on the exact solution to the ℓ_0 sparse recovery problem with a radius on the order of ϵ. 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 $\lambda_{\min} = 3\epsilon$.

1.6 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 ℓ_1 sparse recovery problem can be seen in the light of a Bayesian framework. Solving the following problem

$$\min_{\{\alpha_{jk}\}} \left\| \mathbf{X} - \sum_{j=1}^{D} \sum_{k=1}^{D'} \mathbf{\Xi}_k \alpha_{jk} \mathbf{\Phi}_j \right\|^2 + 2\lambda \sum_{j=1}^{D} \sum_{k=1}^{D'} \|\alpha_{jk}\|_{\ell_1} \tag{27}$$

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][5]. 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 $\boldsymbol{\Xi}$ is chosen based on a spectral sparsity assumption. The choice of the dictionary $\boldsymbol{\Xi}$ relies on sparsity prior information.

2 Applications To Some Sparse Multichannel Image Inverse Problems

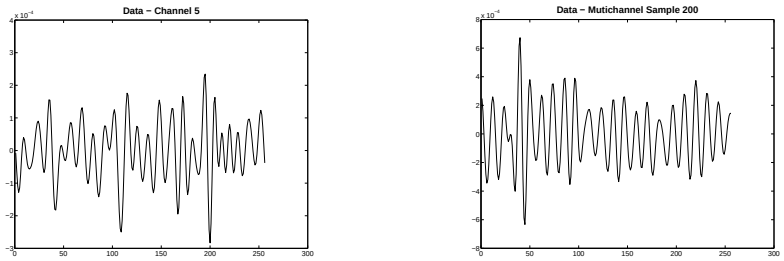
2.1 Multichannel Morphological Component Separation

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 $\boldsymbol{\Psi} = \boldsymbol{\Xi} \otimes \boldsymbol{\Phi}$ is such that $\boldsymbol{\Xi}$ and $\boldsymbol{\Phi}$ are both the union of the DCT-based basis ($\boldsymbol{\Xi}_2$ and $\boldsymbol{\Phi_2}$) and an orthogonal wavelet transform ($\boldsymbol{\Xi}_1$ and $\boldsymbol{\Phi_1}$). Hereafter, the data $\mathbf{X}$ are assumed to be the linear combination of 4 multichannel morphological components corresponding to the following multichannel bases : $\boldsymbol{\Xi}_1 \otimes \boldsymbol{\Phi}_1$, $\boldsymbol{\Xi}_2 \otimes \boldsymbol{\Phi}_1$, $\boldsymbol{\Xi}_2 \otimes \boldsymbol{\Phi}_1$ and $\boldsymbol{\Xi}_2 \otimes \boldsymbol{\Phi}_2$.
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 $\{\boldsymbol{\Xi}_i \otimes \boldsymbol{\Phi}_j\}_{i=1,2;j=1,2}$ respectively. The related coefficients $\{\alpha_{jk}\}_{j=1,2;k=1,2}$ have been drawn according 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 Figures 2 to 5 show the 5^{th} channels (lines of ϖ) and 200^{th} columns (columns of ϖ) of these synthetic morphological components.
The multichannel MCA algorithm is then used to estimate the 4 aforementioned morphological components. The recovered multichannel morphological components are then depicted on the right panels of Figures 2 to 5. At first sight, the pictured samples of the multichannel morphological components are visually well estimated. More quantitatively, the following tabular provides the recovery error :

[5] An alternative choice would be the wave atoms [7].

	ϖ_{11}	ϖ_{12}	ϖ_{21}	ϖ_{22}
Recovery error SNR in dB	47.3	61.9	47.9	58.2

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



(a) (b)

Fig. 1. The original data $\mathbf{X}$

.

2.2 Multichannel Inpainting

In this section, we address the problem of multichannel image inpainting. In this context, a set of samples are missing (set to zero) in the data. According to the general inverse problem framework in Equation (1), the observed data are modelled 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 $\mathbf{X}$ 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 $\mathbf{X}$; 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].
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 Equation (19). The sparsity driven

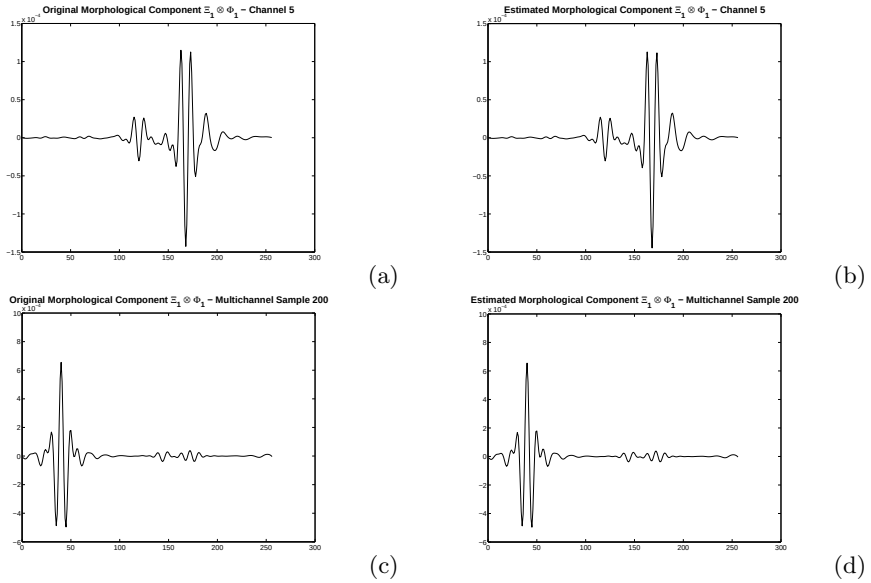


Fig. 2. Multichannel Morphological Component Separation - Component related to $\mathbf{\Xi}_1 \otimes \mathbf{\Phi}_1$: (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morphological component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the original morphological component.

inpainting objective can be written as follows :

$$\min_{\{\alpha_{jk}\}} \left\| \mathbf{Y} - \mathcal{M} \odot \left[\sum_{j=1}^{D} \sum_{k=1}^{D'} \mathbf{\Xi}_k \alpha_{jk} \mathbf{\Phi}_j \right] \right\|^2 + 2\lambda \sum_{j=1}^{D} \sum_{k=1}^{D'} \|\alpha_{jk}\|_{\ell_1} \tag{29}$$

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:

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

The mMCA-based inpainting algorithm is summarized as follows :

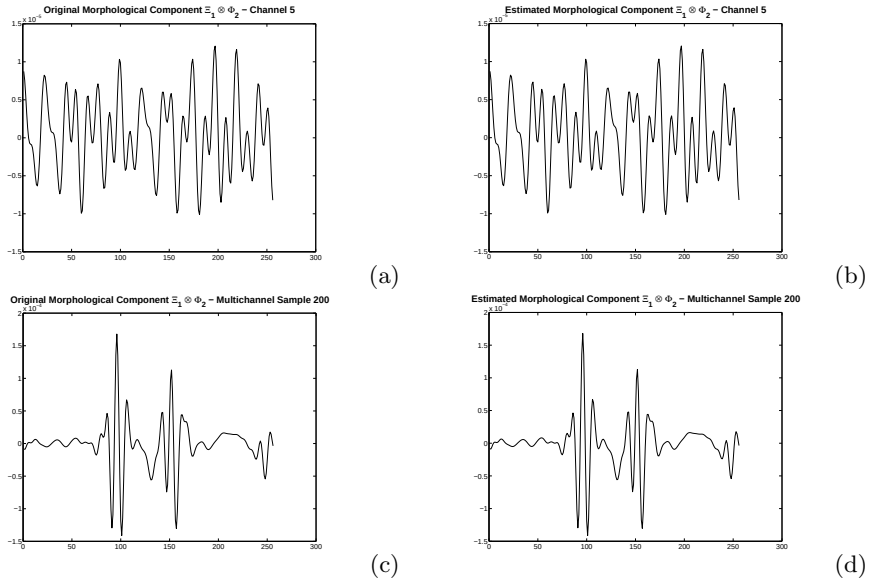


Fig. 3. Multichannel Morphological Component Separation - Component related to $\boldsymbol{\Xi}_1 \otimes \boldsymbol{\Phi}_2$: (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morphological component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the original morphological component.

1. Set the number of iterations $I_{\max}$ 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 as in Section 1.5),
   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 $\{\tilde{\varpi}_{jk}\}^{(h-1)}$.
   For $j = 1, \cdots, D$ and $k = 1, \cdots, D'$
   - 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 $\lambda^{(h)}$:
     $\tilde{\alpha}_{jk}^{(h)} = \Delta_{\lambda^{(h)}} \left(\boldsymbol{\Xi}_k^T \mathbf{R}_{jk}^{(h)} \boldsymbol{\Phi}_j^T \right)$.
   - Get the new estimate of ϖ_{jk} by reconstructing from the selected coefficients $\tilde{\alpha}_{jk}^{(h)}$:
     $\tilde{\varpi}_{jk}^{(h)} = \boldsymbol{\Xi}_k \tilde{\alpha}_k^{(h)} \boldsymbol{\Phi}_j$
   c. Update the hypercube $\tilde{\mathbf{X}}^{(h)} = \sum_{j=1}^{D} \sum_{j=1}^{D'} \tilde{\varpi}_{jk}^{(h)}$.
3. Decrease the threshold $\lambda^{(h)}$ following mMOM strategy.

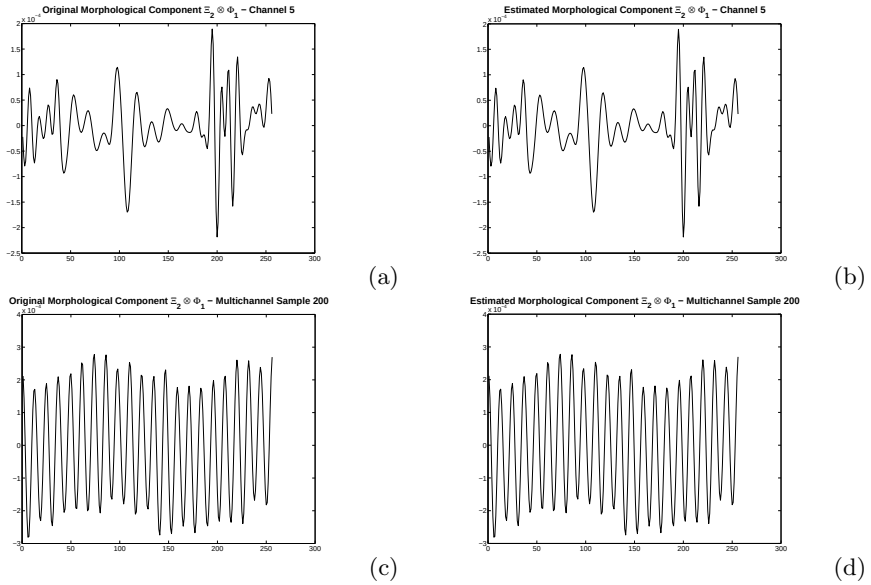


Fig. 4. Multichannel Morphological Component Separation - Component related to $\boldsymbol{\Xi}_2 \otimes \boldsymbol{\Phi}_1$: (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morphological component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the original morphological component.

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

Hyperspectral Data Inpainting : We deal with a hyperspectral data cube $\mathbf{X}$. Unfortunately, as usual when dealing with real data, multichannel pixels are missing. In the experiment we carried out, the data $\mathbf{X}$ is a *Mars Orbiter*[6] hyperspectral cube composed of a 128×128 spatial observations measured at 64 different frequencies (channels). $\mathbf{X}$ 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 $\boldsymbol{\Psi} = \boldsymbol{\Xi} \otimes \boldsymbol{\Phi}$. Each spectrum of the data is assumed to be sparse in the orthogonal one-dimensional wavelet basis $\boldsymbol{\Xi}$. Each spatial observation is nearly sparse in the orthogonal bidimensional wavelet basis $\boldsymbol{\Phi}$. Figure 6(a) displays the 10-th 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

[6] See the *Mars Orbiter* website at *http://mars.jpl.nasa.gov/mro/*.

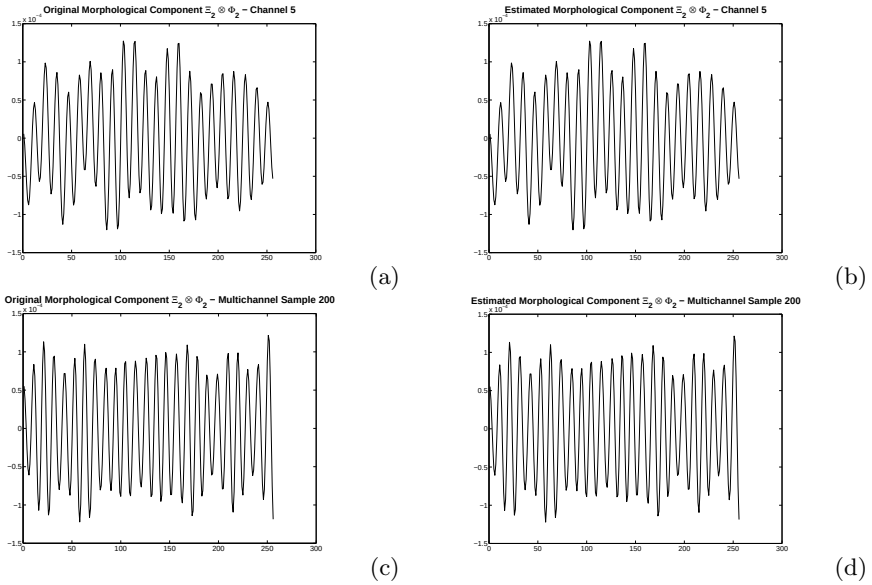


Fig. 5. Multichannel Morphological Component Separation - Component related to $\boldsymbol{\Xi}_2 \otimes \boldsymbol{\Phi}_2$: (a) The channel 5 of the original morphological component. (b) Channel 5 of the estimated morphological component. (c) Multichannel sample 200 of the original morphological component. (d) Multichannel sample 200 of the original morphological component.

and the original is 19.2dB. 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 Figure 7(b) shows the masked spectrum. From Figure 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 Figure 8 displays the behaviour 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 data as it performs well in terms of spectral and spatial feature recovery.

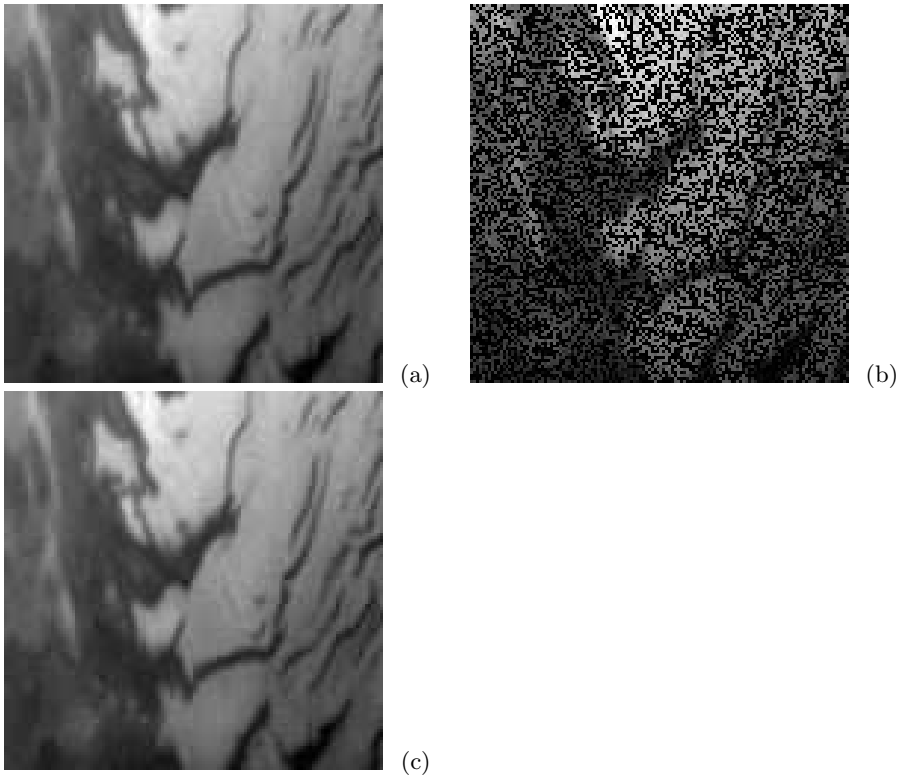


Fig. 6. Restoring Mars-Orbiter hyperspectral data: spatial feature recovery. (a) 10th original channel. (b) Masked channel - 50% of the pixels are missing. (c) Recovered channel using mMCA

.

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

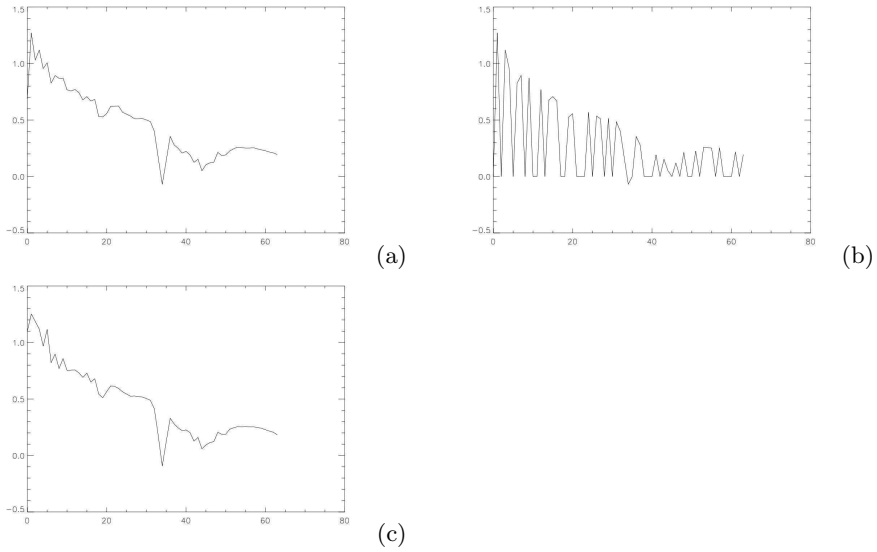


Fig. 7. Restoring Mars-Orbiter hyperspectral data: recovering the spectra. (a) Original spectrum at pixel $\{10, 10\}$. (b) Masked spectrum - 50% of the samples are missing. (c) Recovered spectrum using mMCA. **Abscissa :** channel number. **Ordinate :** spectrum amplitude

.

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 $\{y_k\}_{k=1,\cdots,M}$: $y_k = \langle x, \theta_k \rangle$. ymThese measurements are more conveniently represented in a matrix formulation:

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

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

$$\min_x \|x\mathbf{\Phi}^T\|_{\ell_1} \text{ s.t } y = x\mathbf{\Theta} \tag{31}$$

More precisely, Candès *et al.* showed in [56] that if $M > CK\log(t)$ then the previous ℓ_1 minimization problem provides the exact signal x with $C \simeq 22(\delta + 1)$ and probability of success $1 - \mathcal{O}(t^{-\delta})$. 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

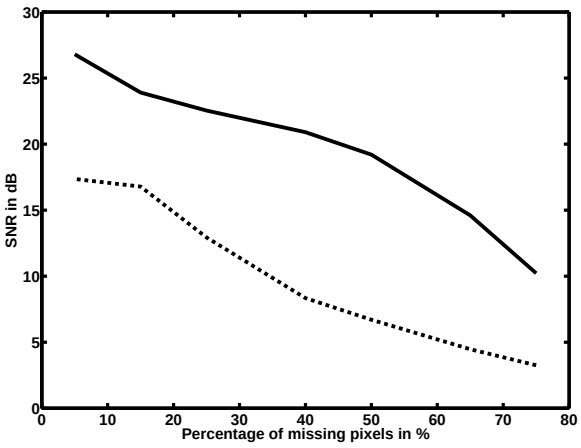


Fig. 8. Recovering the hyperspectral data : SNR in dB between the original data and the recovered hyperspectral data using mMCA (solid line) and MCA on each channel separately (dashed line).

CS, see [58–62] to name a few[7].
The observed data y defined in Equation (30) can be considered as subspace "projection" of the original data x. When the the *sensing* matrix $\mathbf{\Theta}$ 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.

3 Steps Ahead - Learning The Sparse Representation :

3.1 Adaptive Multichannel Morphological Component Separation

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

[7] A website at http://www.dsp.ece.rice.edu/cs/ is dedicated to Compressed Sensing Resources.

– Global approaches : in various fields, adaptive schemes have been proposed to globally update the representation (see e.g. [63–65]).
– 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 adpative 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 $\mathbf{\Xi}$ and the spatial dictionary $\mathbf{\Phi}$. Adapting the spatial dictionary to the data could be done by e.g choosing a decomposition tree assuming $\mathbf{\Phi}$ lies in a class of tree-based multiscale transforms. When applicable, the same adaptive process can be performed for $\mathbf{\Xi}$.

The data $\mathbf{X}$ are made of 3 observed channels corresponding to each colour 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 $\mathbf{X}$ are the linear combination of D multichannel morphological components. In this section, each channel of the data $\mathbf{X}$ is the linear combination of a texture part (assumed to be well sparsified by a Discrete Cosine Transform) and a contour part (sparse in the curvelet tight frame). Figure 9 displays a toy-example based colour image $\mathbf{X}$. We then propose recovering the colour morphological compoenents using the proposed mMCA method which seeks to adapt the colour space to the data $\mathbf{X}$. In this context, we assume that $\mathbf{\Xi}$ is a 3×3 invertible matrix. Adapting the spectral basis $\mathbf{\Xi}$ (*i.e.* the colour space) to the data then amounts to estimate an "optimal" matrix $\mathbf{\Xi}$. The mMCA algorithm is then adapted such that at each iteration h the matrix $\mathbf{\Xi}$ is updated by its least-squares estimate:

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

This problem has a unique minimizer defined as follows:

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

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

The mMCA algorithm of Section 2.2 is then adapted as follows :

1. Set the number of iterations $I_{\max}$ 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. For $j = 1, \cdots, D$
      • Compute the residual term $\mathbf{R}_j^{(h)}$ assuming the current estimates of $\varpi_{\{p\}\neq\{j\}}$, $\tilde{\varpi}_{p\neq j}^{(h-1)}$ are fixed:
      $$\mathbf{R}_j^{(h)} = \mathbf{X} - \sum_{p\neq j} \tilde{\varpi}_p^{(h-1)}.$$
      • Estimate the current coefficients of $\tilde{\varpi}_j^{(h)}$ by thresholding with threshold $\lambda^{(h)}$:
      $$\tilde{\alpha}_j^{(h)} = \Delta_{\lambda^{(h)}}\left(\mathbf{\Xi}^{(h)^T}\mathbf{R}_j^{(h)}\mathbf{\Phi}_j^T\right).$$
      • Get the new estimate of ϖ_j by reconstructing from the selected coefficients $\tilde{\alpha}_j^{(h)}$:
      $$\tilde{\varpi}_j^{(h)} = \mathbf{\Xi}^{(h)^T}\tilde{\alpha}_j^{(h)}\mathbf{\Phi}_j.$$
   b. Update the *spectral* basis $\mathbf{\Xi}$:
      $$\mathbf{\Xi}^{(h+1)} = \mathbf{Y}^{(h)}\left[\sum_{j=1}^{D}\varpi_j^{(h)}\mathbf{\Phi}_j\right]^{\dagger}.$$
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 MCA-based algorithm with the new adaptive mMCA algortihm decribed 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 : $\mathbf{\Xi} = \mathbf{I}$.
- Adaptive approach : a global adaptive mMCA is applied to the whole data $\mathbf{X}$.

The input toy-example based data are displayed in Figure 9. The texture part is composed of a globally oscillating pattern; it will be assumed to be well-sparsified by a Discrete Cosine Transform ($\mathbf{\Phi_1}$ with be the DCT-based basis). The Decomposition results are illustrated in Figure 10. At first sight, both approaches (MCA and mMCA) perform similarly providing good visual results. More quantitatively, the following tabular summarizes the respective performances :

Recovery error SNR in dB	Texture Part	Contour Part
mMCA with $\mathbf{\Xi} = \mathbf{I}$	15.7	22.9
Adaptive mMCA	**16.3**	**23.6**

To conlude accounting for inter-channel structures greatly improves the decomposition task. We introduce an adaptive sparsity-driven algorithm that is able to adapt to the data. In the next section, we apply a similar adaptive scheme to the multichannel inpainting problem.

Fig. 9. Adaptive Multichannel Morphological Component Separation - The 256×256 original colour image

3.2 Adaptive Colour 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 colour image inpainting. The data $\mathbf{X}$ are assumed to be made of three channels (*i.e.* correponding to each colour layer). We apply exactly the same spectral dictionary update described in Section 3.1. In the context of inpainting, the mMCA algorithm of Section 2.2 is then adapted as follows :

1. Set the number of iterations $I_{\max}$ 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 $\{\tilde{\varpi}_{jk}\}^{(h-1)}$.
   For $j = 1, \cdots, D$
   - Compute the residual term $\mathbf{R}_j^{(h)}$ assuming the current estimates of $\varpi_{\{p\}\neq\{j\}}$, $\tilde{\varpi}_{p\neq j}^{(h-1)}$ are fixed:
     $$\mathbf{R}_j^{(h)} = \mathbf{Y}^{(h)} - \sum_{p\neq j} \tilde{\varpi}_p^{(h-1)}.$$
   - Estimate the current coefficients of $\tilde{\varpi}_j^{(h)}$ by thresholding with threshold $\lambda^{(h)}$:
     $$\tilde{\alpha}_j^{(h)} = \Delta_{\lambda^{(h)}}\left(\mathbf{\Xi}^{(h)^T} \mathbf{R}_j^{(h)} \mathbf{\Phi}_j^T\right).$$
   - Get the new estimate of ϖ_j by reconstructing from the selected coefficients $\tilde{\alpha}_j^{(h)}$:
     $$\tilde{\varpi}_j^{(h)} = \mathbf{\Xi}^{(h)^T} \tilde{\alpha}_j^{(h)} \mathbf{\Phi}_j.$$
   c. Update the hypercube $\tilde{\mathbf{X}}^{(h)} = \sum_{j=1}^{D} \tilde{\varpi}_j^{(h)}$.
   d. Update the *spectral* basis $\mathbf{\Xi}$:
   $$\mathbf{\Xi}^{(h+1)} = \mathbf{Y}^{(h)} \left[\sum_{j=1}^{D} \varpi_j^{(h)} \mathbf{\Phi}_j\right]^{\dagger}.$$
3. Decrease the threshold $\lambda^{(h)}$ following mMOM strategy.

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

3.3 Relations with BSS

In the previous sections, the adaptive mMCA-based algorithm has been devised to adapt the spectral dictionary $\mathbf{\Xi}$ 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 $\mathbf{X}$ in the multichannel dictionary $\mathbf{\Xi} \otimes \mathbf{\Phi}$.
- Spectral dictionary update : update the spectral dictionary $\mathbf{\Xi}$.

Such a two-step iterative algorithm have the flavour of the GMCA algorithm we proposed for solving Blind Source Separation Issues 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 colour 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.

4 Conclusion

We recalled the tremendous effectiveness of sparsity-based methods in signal restoration. In this paper we emphasize 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 classical recovery issues such as multichannel morphological component extraction and colour 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 convolution, 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.

References

1. Starck, J.L., Candès, E., Donoho, D.: The curvelet transform for image denoising. IEEE Transactions on Image Processing **11**(6) (2002) 670–684
2. Starck, J.L., Elad, M., Donoho, D.: Image decomposition via the combination of sparse representations and a variational approach. IEEE Transactions on Image Processing **14**(10) (2005) 1570–1582
3. Li, Y., Amari, S., Cichocki, A., Guan, C.: Underdetermined blind source separation based on sparse representation. IEEE Transactions on information theory **52** (2006) 3139–3152
4. Zibulevsky, M., Pearlmutter, B.: Blind source separation by sparse decomposition. Neural Computations **13/4** (2001)
5. Bobin, J., Moudden, Y., Starck, J.L., Elad, M.: Morphological diversity and source separation. IEEE Signal Processing Letters **13**(7) (2006) 409–412
6. Olshausen, B., Field, D.: Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision Research. **37** (2006) 3311–3325
7. Demanet, L., Ying, L.: Wave atoms and sparsity of oscillatory patterns. ACHA (2006) accepted.
8. Mallat, S.: A Wavelet Tour of Signal Processing. Academic Press (1998)
9. Candès, E., Donoho, D.: Curvelets. Technical report, Statistics, Stanford University (1999)

[8] Recall that $\mathbf{G}_{\mathbf{\Psi}_{I_{\gamma^\star}}}$ is the Gram matrix of the subdictionary $\mathbf{\Psi}_{I_{\gamma^\star}}$.

10. Candès, E., Demanet, L., Donoho, D., Ying, L.: Fast discrete curvelet transforms. SIAM Multiscale Model. Simul **5/3** (2006) 861–899
11. Cotter, S., Rao, B., Engan, K., Kreutz-Delgado, K.: Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Transactions on Signal Processing **53** (2005) 2477–2488
12. 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)
13. Tropp, J., Gilbert, A., Strauss, M.: Algorithms for simultaneous sparse approximation. part i: greedy pursuit. Signal Processing - special issue "Sparse approximations in signal and image processing" **86** (2006) 572–588
14. Tropp, J.: Algorithms for simultaneous sparse approximation. part ii: convex relaxation. Signal Processing - special issue "Sparse approximations in signal and image processing" **86** (2006) 589–602
15. Chen, J., Huo, X.: Sparse representations for multiple measurement vectors (MMV) in an over-complete dictionary. In: ICASSP '05. (2005)
16. Starck, J.L., Elad, M., Donoho, D.: Redundant multiscale transforms and their application for morphological component analysis. Advances in Imaging and Electron Physics **132** (2004) 287–348
17. Elad, M., Starck, J.L., Donoho, D., Querre, P.: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). ACHA **19**(3) (2005) 340–358
18. Tropp, J.: Greedy is good : algorithmic results for sparse approximation. IEEE Transactions on Information Theory **50**(10) (2004) 2231–2242
19. Donoho, D., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Transactions on Information Theory **47**(7) (2001) 2845–2862
20. Gribonval, R., Nielsen, M.: Sparse representations in unions of bases. IEEE Transactions on Information Theory **49**(12) (2003) 3320–3325
21. Candès, E., Donoho, D.: Ridgelets: the key to high dimensional intermittency? Philosophical Transactions of the Royal Society of London A **357** (1999) 2495–2509
22. Starck, J.L., Candès, E., Donoho, D.: The curvelet transform for image denoising. IEEE Transactions on Image Processing **11**(6) (2002) 131–141
23. LePennec, E., Mallat, S.: Sparse geometric image representations with bandelets. IEEE Transactions on Image Processing **14**(4) (2005) 423–438
24. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Transactions on Image Processing **14**(12) (2005) 2091–2106
25. Chen, S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing **20**(1) (1999) 33–61
26. Donoho, D., Elad, M.: Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization. Proc. Nat. Aca. Sci. **100** (2003) 2197–2202
27. Bruckstein, A., Elad, M.: A generalized uncertainty principle and sparse representation in pairs of $\mathbb{R}^N$ bases. IEEE Transactions on Information Theory **48** (2002) 2558–2567
28. Fuchs, J.J.: On sparse representations in arbitrary redundant bases. IEEE Transactions on Information Theory **50**(6) (2004) 1341–1344
29. Bruckstein, A., Donoho, D., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review (2007) to appear.
30. Bobin, J., Starck, J.L., Fadili, J., Moudden, Y., Donoho, D.: Morphological component analysis: An adaptive thresholding strategy. IEEE Trans. On Image Processing **16**(11) (November 2007) 2675 – 2681

31. Feuer, A., Nemirovsky, A.: On sparse representation in pairs of bases. IEEE Transactions on Information Theory **49**(6) (2003) 1579–1581
32. Gribonval, R., Nielsen, M.: Beyond sparsity : recovering structured representations by l_1-minimization and greedy algorithms. – application to the analysis of sparse underdetermined ICA –. Advances in Computational Mathematics (2005) in press.
33. Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing **41**(12) (1993) 3397–3415
34. Donoho, D., Tsaig, Y.: Fast solution of ℓ_1 minimization problems when the solution may be sparse. (2006) submitted.
35. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Annals of Statistics **32**(2) (2004) 407–499
36. Osborne, M.R., Presnell, B., Turlach, B.A.: A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis **20**(3) (2000) 389–403
37. Malioutov, D.M., Cetin, M., Willsky, A.S.: Homotopy continuation for sparse signal representation. In: ICASSP '05. Volume 5. (2005) 733–736
38. Plumbley, M.: Recovery of sparse representations by polytope faces pursuit. In: ICA06. (2006) 206–213
39. Donoho, D., Tsaig, Y., Drori, I., Starck, J.L.: Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. IEEE Transactions On Information Theory (2006) submitted.
40. Sardy, S., Bruce, A., Tseng, P.: Block coordinate relaxation methods for nonparametric wavelet denoising. Journal of Computational and Graphical Statistics **9**(2) (2000) 361–379
41. Donoho, D., Elad, M., Temlyakov, V.: Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. On Information Theory **52** (2006) 6–18
42. Fuchs, J.J.: Recovery conditions of sparse representations in the presence of noise. ICASSP '06 **3**(3) (2006) 337–340
43. Tropp, J.: Just relax: Convex programming methods for subset selection and sparse approximation. IEEE Transactions on Information Theory **52**(3) (2006) 1030–1051
44. M.Elad: Why simple shrinkage is still relevant for redundant representations? IEEE Transactions on Information Theory **52**(12) (2006) 5559–5569
45. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. SIAM Journal on Multiscale Modeling and Simulation **4**(4) (2005) 1168–1200
46. Chan, T., Shen, J.: Mathematical models for local non-texture inpainting. SIAM J. Appl. Math **62**(3) (2001) 1019–1043
47. Bertalmìo, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. Proceedings of SIGGRAPH 2000, New Orleans, USA (2000)
48. Masnou, S., Morel, J.M.: Disocclusion : a variational approach using level lines. IEEE Trans. On Image Processing **11**(2) (2002) 68–76
49. Tschumperlé, D., Deriche, R.: Vector-valued image regularization with PDE's : A common framework for different applications. CVPR'2003, Computer Vision and Pattern Recognition, Madison/USA (2003)
50. Demanet, L., Song, B., Chan, T.: Image inpainting by correspondence maps: a deterministic approach. Proc. VLSM conf., Nice, France (2003)
51. Guleryuz, O.: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I: theory. IEEE Trans. On Image Processing **15**(3) (2006) 539 – 554
52. Guleryuz, O.: Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II: adaptive algorithms. IEEE Trans. On Image Processing **15**(3) (2006) 555 – 571

53. Fadili, J., Starck, J.L.: EM algorithm for sparse representation - based image inpainting. IEEE International Conference on Image Processing ICIP'05 **2** (2005) 61–63 Genoa,Italia.
54. Fadili, J., Starck, J.L., Murtagh, F.: Inpainting and zooming using sparse representations. The Computer Journal - in press (2006) in press.
55. D.Donoho: Compressed sensing. IEEE Trans. on Information Theory **52**(4) (2006) 1289–1306
56. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. on Information Theory **52**(2) (2006) 489–509
57. Candès, E.: Compressive sampling. International Congress of Mathematics, Madrid (2006)
58. Donoho, D., Tsaig, Y.: Extensions of compressed sensing. Signal Processing **86**(3) (2006) 5433–548
59. Candès, E., Tao, T.: Statistical estimation when p is much larger than n. To appear in Annals of Statistics - available at *http://www.dsp.ece.rice.edu/cs/* **52**(9) (2006) 4036–4048
60. Candès, E., Tao, T.: Near optimal signal recovery from random projections: Universal encoding strategies? IEEE Trans. on Information Theory **52**(12) (2006) 5406–5425
61. Candès, E., Romberg, J.: Practical signal recovery from random projections. Preprint - available at *http://www.dsp.ece.rice.edu/cs/* (2005)
62. Tropp, J., Gilbert, A.: Signal recovery from partial information via orthogonal matching pursuit. Preprint - available at *http://www.dsp.ece.rice.edu/cs/* (2005)
63. Aharon, M., Elad, M., Bruckstein, A.: k-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing **54**(11) (2006) 4311–4322
64. Sallee, P., Olshausen, B.: Learning sparse multiscale image representations. In: Advances Neural Information Processing Systems. (2003)
65. Plumbley, M.D.: Dictionary learning for l1-exact sparse coding. In Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D., eds.: Independent Component Analysis and Signal Separation. Volume 4666 of LNCS., Springer (2007) 406–413
66. 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. (2007) Preprint - available at *http://www.dsp.ece.rice.edu/cs/*.
68. J.Bobin, J-L.Starck, J.Fadili, Y.Moudden: Sparsity and morphological diversity in blind source separation. IEEE Transactions on Image Processing **16**(11) (November 2007) 2662 – 2674

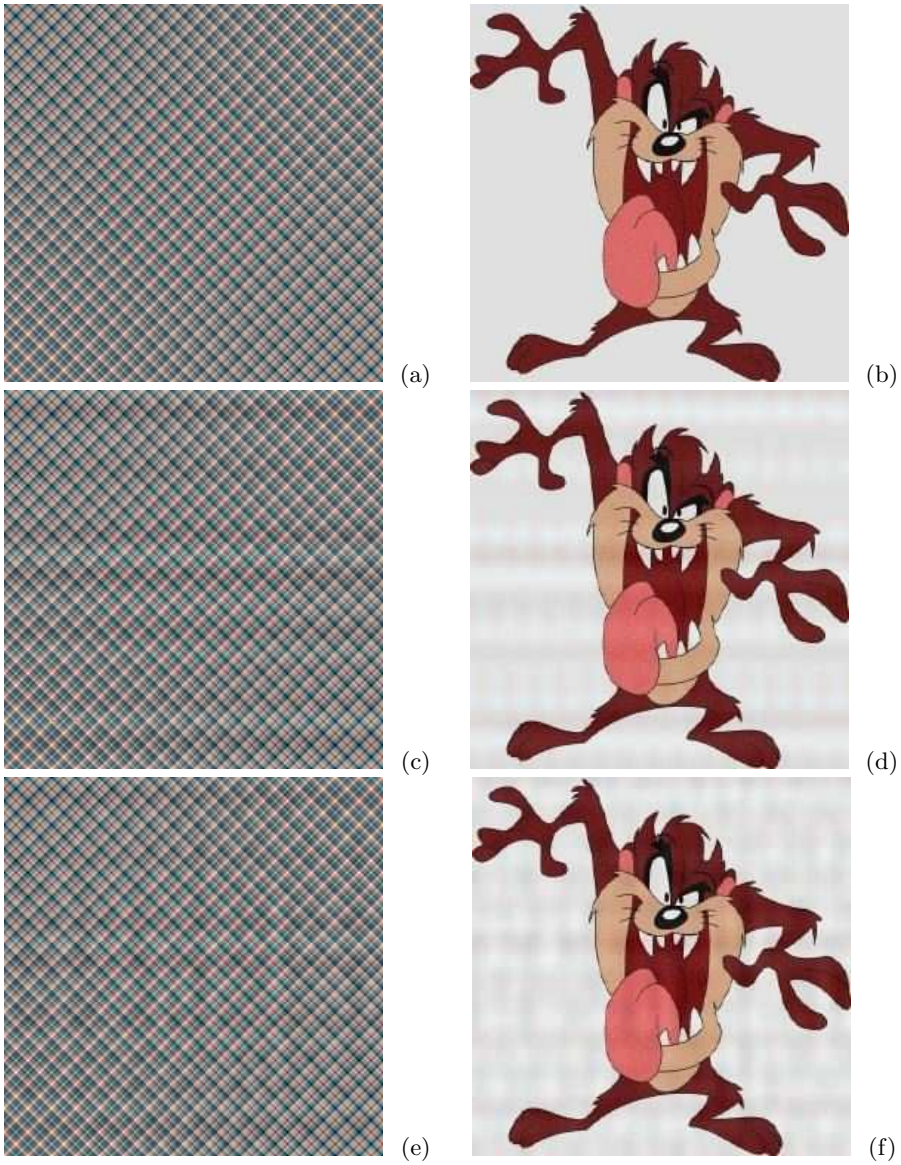


Fig. 10. Adpative Multichannel Morphological Component Separation - Components Recovery : (a) Original multichannel texture component. (b) Original multichannel pointwise component. (c) Multichannel texture component estimated with MCA. (d) Multichannel pointwise component estimated with MCA. (e) Multichannel texture component estimated with mMCA. (f) Multichannel pointwise component estimated with mMCA.

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

.

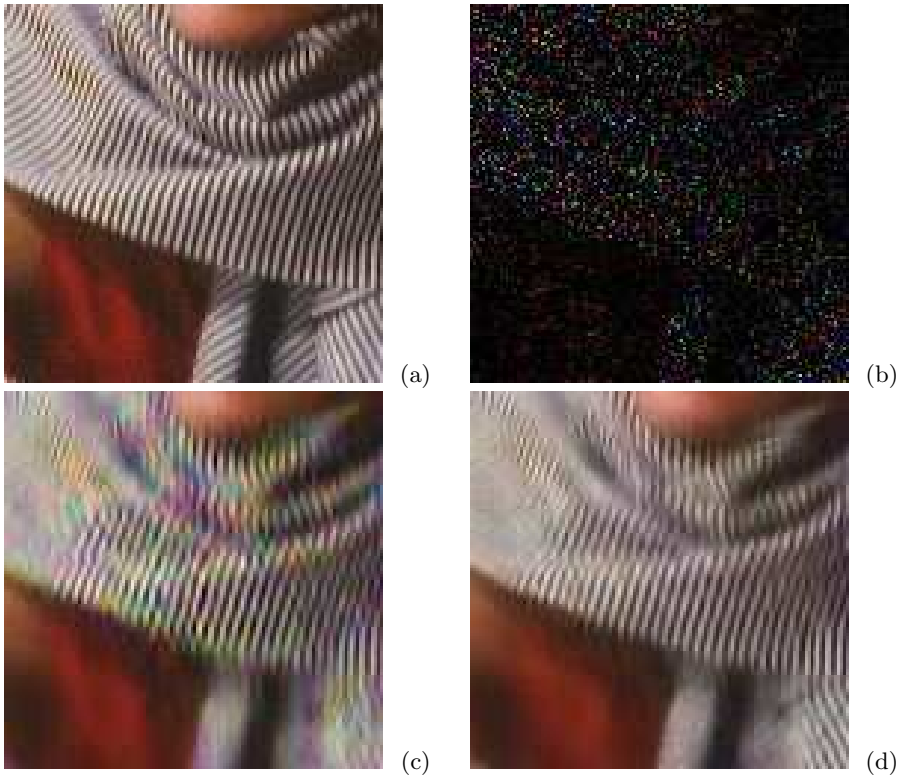


Fig. 12. Zoom on recovered Barbara colour image original colour image. (a) Original Barbara colour image. (b) Masked image - 90% of the colour pixels are missing. (c) Inpainted image using the original MCA algorithm on each color channel. (d) Inpainted image using the adaptive mMCA algorithm.

Erratum to "Morphological Diversity and Sparsity for Multichannel Data Restoration"

We first define precisely what we intend by correct sparse recovery.

Definition 1. *We say that*

- *the GMCA/mMOM algorithm has the Correct Rejection Property, if at each iteration and for each morphological component $\gamma^\star$, the algorithm does not select new atoms to enter the active set that are outside the support of the $\gamma^\star$-th morphological component.*
- *the GMCA/mMOM algorithm has the Correct Selection Property, if at each iteration and for the morphological component $\gamma^\star$, the algorithm selects at least one atom to enter the active set from the support of the $\gamma^\star$-th morphological component.*
- *the GMCA/mMOM algorithm has the Exact Selection Property, if at each iteration and for the morphological component $\gamma^\star$, it has both the Correct Rejection Property and the Correct Selection Property.*

Theorem 1. *Suppose that $\mathbf{X}$ 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 $K = \sum_{j,k} |\Lambda_{jk}|$. At the h-th iteration, assume that the residual $\mathbf{R}^{(h)}$ is K-sparse such that :

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

Let ρ as defined in the appendix. Then we have,

(i) GMCA/mMOM has the Correct Rejection Property if $N = K\mu_\Psi < 1/4$ and $\rho \leq 1/5$.

(ii) GMCA/mMOM has the Correct Selection Property if $N = K\mu_\Psi < N_0 < 1$ and $\rho \leq \frac{1-N_0}{1+3N_0}$. For $N_0 = 1/2$, $\rho \leq 1/5$.

The proof is given the appendix.

As a consequence, taking $N_0 = 1/4$ in Theorem 1(ii) and combining with (i), we have the following corollary on Exact Selection Property,

Corollary 1 (Exact Selection Property). *GMCA/mMOM has the Exact Selection Property if* $K < \frac{\mu_{\Psi}^{-1}}{4}$ *and* $\rho \leq 1/5$.

When the Exact Selection Property holds, the next proposition shows that GMCA/mMOM converges exponentially to $\mathbf{X}$ and its sparsest representation in $\mathbf{\Psi} = [\mathbf{\Xi}_1 \cdots \mathbf{\Xi}_{D'}] \otimes [\mathbf{\Phi}_1 \cdots \mathbf{\Phi}_D]$.

Proposition 1 (Convergence). *Suppose that* $\mathbf{X}$ *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 $K = \sum_{j,k} |\Lambda_{jk}|$.
If $K < \frac{\mu_{\Psi}^{-1}}{4}$ *and* $\rho \leq 1/5$ *then GMCA/mMOM converges exponentially to* $\mathbf{X}$ *and its sparsest representation in* $\mathbf{\Psi}$*. More precisely, the residual converges to zero at an exponential rate.*

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

Appendix : Proofs

Let us first simplify a few notations. We define $\gamma = \{j, k\}$ as a couple of indices. The multichannel dictionary $\mathbf{\Xi}_k \otimes \mathbf{\Phi}_j$ will be written $\mathbf{\Psi}_\gamma$ with $\gamma = (j, k)$. The notation $\psi_\gamma[i]$ will refer to the i-th atom of the multichannel sub-dictionary $\mathbf{\Psi}_\gamma$. Furthermore, the sparse representation of the multichannel data $\mathbf{X}$ will be written as follows:

$$\mathbf{X} = \sum_{\gamma} \sum_{i \in \Lambda_\gamma} \alpha_\gamma[i] \psi_\gamma[i] \ , \tag{1}$$

with Λ_γ the support of the γ-th morphological component and $\Lambda = \cup_\gamma \Lambda_\gamma$. Under the Exact Selection Property, the sparse decomposition of the residual term at iteration h is written as follows:

$$\mathbf{R}^{(h)} = \sum_{\gamma} \sum_{i \in \Lambda_\gamma} \beta_\gamma[i] \psi_\gamma[i] \ . \tag{2}$$

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

Proof of Theorem 1

Proof. Let's denote :

$$(\gamma^\star, i^\star) = \underset{\gamma, i\in\Lambda_\gamma}{\arg\max} \ |\beta_\gamma[i]| \ , \beta^\star = \left|\beta^{(h)}_{\gamma^\star}[i^\star]\right| \ ,$$
$$(\gamma^\dagger, i^\dagger) = \underset{\gamma\neq\gamma^\star, i\in\Lambda_\gamma}{\arg\max} \ |\beta_\gamma[i]| \ , \beta^\dagger = \left|\beta^{(h)}_{\gamma^\dagger}[i^\dagger]\right| \ .$$

Let ρ such that $\beta^\dagger = \rho\beta^\star$ for $0 < \rho < 1$.

Definition 2. *We will say that the γ-th morphological component is ρ_0-separable $0 \leq \rho_0 < 1$, if $\|\beta_{\gamma'}\|_{\ell_\infty} \leq \rho_0 \|\beta_\gamma\|_{\ell_\infty}$, $\forall\gamma' \neq \gamma$.*

Obviously, the morphological component $\gamma^\star$ is ρ-separable, hence ρ_0-separable if $\rho \leq \rho_0$. The parameter ρ_0 can be interpreted as a separability threshold.

Let $\omega_\gamma = \left\|\mathbf{R}^{(h)}\mathbf{\Psi}_\gamma\right\|_{\ell_\infty} \ \forall\gamma$, $m_1 = \omega_{\gamma_1} = \max_\gamma \omega_\gamma$ and $m_2 = \max_{\gamma\neq\gamma_1} \omega_\gamma < m_1$, as defined in (24) in the paper.

Proof of (i) We want to show the Correct Rejection Property for the $\gamma^\star$-th morphological component. Then, we must first guarantee that this component is identifiable, i.e. m_1 occurs within $\gamma^\star$. The following lemma provides a sufficient condition under which the component $\gamma^\star$ is actually γ_1, hence $\omega_{\gamma^\star} = m_1$.

Lemma 1. *Let $0 < N = K\mu_\Psi < N_0 < 1$, and $\rho_0 = \frac{1-N_0}{1+N_0}$. If the $\gamma^\star$-th morphological component is ρ_0-separable, then it is identifiable, i.e. $\gamma^\star = \gamma_1$, $m_1 = \omega_{\gamma^\star}$ and $m_1 > m_2 \geq \omega_{\gamma^\dagger}$.*

Proof. We have,

$$\begin{aligned}
\omega_{\gamma^\star} &= \max_i \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i]\right\rangle\right| \\
&\geq \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i^\star]\right\rangle\right| \\
&\geq \beta^\star - \sum_{\gamma\neq\gamma^\star}\sum_{j\in\Lambda_\gamma} \left|\left\langle \psi_{\gamma^\star}[i^\star], \psi_\gamma[j]\right\rangle\right| |\beta_\gamma[j]| \\
&\geq \beta^\star - |\Lambda|\,\mu_\Psi\beta^\dagger \ , \qquad (3)
\end{aligned}$$

and $\forall\gamma \neq \gamma^\star$

$$\begin{aligned}
\omega_\gamma &= \max_i \left|\left\langle \mathbf{R}^{(h)}, \psi_\gamma[i]\right\rangle\right| \\
&\leq \beta^\dagger + \max_i \sum_{\gamma'\neq\gamma}\sum_{j\in\Lambda_{\gamma'}} \left|\left\langle \psi_\gamma[i], \psi_{\gamma'}[j]\right\rangle\right| |\beta_{\gamma'}[j]| \\
&\leq \beta^\dagger + |\Lambda|\,\mu_\Psi\beta^\star \ , \qquad (4)
\end{aligned}$$

where we used $\left|\left\langle \psi_\gamma[i], \psi_{\gamma'}[j] \right\rangle\right| \leq \mu_\Psi$, $\forall \gamma \neq \gamma'$ or $i \neq j$, orthonormality of each sub-dictionary $\mathbf{\Psi}_\gamma$, and recall that $\beta^\dagger = \rho\beta^\star$ for $0 < \rho < 1$. To have $\omega_{\gamma^\star} > \omega_\gamma \ \forall \gamma \neq \gamma^\star$, we need

$$\beta^\star - |\Lambda|\,\mu_\Psi \beta^\dagger > \beta^\dagger - |\Lambda|\,\mu_\Psi \beta^\star \Leftrightarrow \beta^\dagger < \frac{1-N}{1+N}\beta^\star \ . \tag{5}$$

$\frac{1-N}{1+N}$ is a strictly decreasing function of $N \in (0, N_0)$ strictly bounded below by ρ_0. As the $\gamma^\star$-th morphological component is ρ_0-separable, the bound (5) holds. Thus, $\omega_{\gamma^\star} > \omega_\gamma \ \forall \gamma \neq \gamma^\star \Leftrightarrow \gamma^\star = \gamma_1 \Leftrightarrow m_1 = \omega_{\gamma^\star} > m_2 \geq \omega_{\gamma^\dagger}$.

By definition, the mMOM thresholding rule selects new coefficients such that

$$\forall i, \gamma; \quad \left|\left\langle \mathbf{R}^{(h)}, \psi_\gamma[i] \right\rangle\right| \geq \lambda^{(h)} = \frac{m_1 + m_2}{2} \ .$$

To prove that GMCA/mMOM agrees with the Correct Rejection Property, we need to show that

$$\max_{(\gamma,i)\in C_{\gamma^\star}} \left|\left\langle \mathbf{R}^{(h)}, \psi_\gamma[i] \right\rangle\right| < \lambda^{(h)}, \quad C_{\gamma^\star} = \{\gamma^\star, i \in \Lambda^c_{\gamma^\star}\} \cup \{\gamma \neq \gamma^\star, \forall i\} \ , \tag{6}$$

where $\Lambda^c_{\gamma^\star}$ is the complement of the support $\Lambda_{\gamma^\star}$ restricted to the sub-dictionary $\mathbf{\Psi}_{\gamma^\star}$. Clearly, $C_{\gamma^\star}$ is the set of all atom indices but those that belong to the support of the $\gamma^\star$-th morphological component that we want to recover at iteration h.

As $\rho < 1/5$ and $N_0 = 1/4$, $\gamma^\star$ is $\left(\frac{1-N_0}{1+N_0} = 3/5\right)$-separable. On the one hand, we have using Lemma 1,

$$\begin{gathered} m_1 = \omega_{\gamma^\star} = \max_i \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i] \right\rangle\right| \ , \\ \beta^\star - \sum_{\gamma\neq\gamma^\star} \sum_{j\in\Lambda_\gamma} \left|\left\langle \psi_{\gamma^\star}[i^\star], \psi_\gamma[j] \right\rangle\right| |\beta_\gamma[j]| \leq m_1 \leq \beta^\star + \max_i \sum_{\gamma\neq\gamma^\star} \sum_{j\in\Lambda_\gamma} \left|\left\langle \psi_{\gamma^\star}[i], \psi_\gamma[j] \right\rangle\right| |\beta_\gamma[j]| \ , \\ \beta^\star(1-\rho N) \leq m_1 \leq \beta^\star(1+\rho N) \ , \end{gathered} \tag{7}$$

and

$$\begin{aligned} \beta^\dagger - \sum_{\gamma\neq\gamma^\dagger} \sum_{j\in\Lambda_\gamma} \left|\left\langle \psi_{\gamma^\dagger}[i^\dagger], \psi_\gamma[j] \right\rangle\right| |\beta_\gamma[j]| \leq \omega_{\gamma^\dagger} \leq m_2 \leq \max_{\gamma\neq\gamma^\star} \Bigg\{ & \|\beta_\gamma\|_{\ell_\infty} \\ & + \max_i \sum_{\gamma'\neq\gamma} \sum_{j\in\Lambda_{\gamma'}} \left|\left\langle \psi_\gamma[i], \psi_{\gamma'}[j] \right\rangle\right| |\beta_{\gamma'}[j]| \Bigg\} \ , \end{aligned} \tag{8}$$

$$\begin{gathered} \beta^\dagger - |\Lambda|\,\mu_\Psi \beta^\star \leq m_2 \leq \beta^\dagger + |\Lambda|\,\mu_\Psi \beta^\star \ , \\ \beta^\star(\rho - N) \leq m_2 \leq \beta^\star(\rho + N) \ . \end{gathered}$$

From (7) and (8), we get

$$\frac{(1+\rho)(1-N)}{2} \leq \lambda^{(h)} \leq \frac{(1+\rho)(1+N)}{2} \ . \tag{9}$$

On the other hand, a straightforward calculation leads to

$$\left|\left\langle \mathbf{R}^{(h)}, \psi_\gamma[i] \right\rangle\right| = \begin{cases} \left|\sum_{\gamma\neq\gamma^\star}\sum_{j\in\Lambda_\gamma}\langle\psi_{\gamma^\star}[i],\psi_\gamma[j]\rangle\beta_\gamma[j]\right| & i\in\Lambda^c_{\gamma^\star}\ . \\ \left|\sum_{\gamma'\neq\gamma}\sum_{j\in\Lambda_{\gamma'}}\langle\psi_\gamma[i],\psi_{\gamma'}[j]\rangle\beta_{\gamma'}[j]\right| & \gamma\neq\gamma^\star, i\in\Lambda^c_\gamma\ . \\ \left|\beta_\gamma[i]+\sum_{\gamma'\neq\gamma}\sum_{j\in\Lambda_{\gamma'}}\langle\psi_\gamma[i],\psi_{\gamma'}[j]\rangle\beta_{\gamma'}[j]\right| & \gamma\neq\gamma^\star, i\in\Lambda_\gamma\ . \end{cases}$$

$$\leq \begin{cases} |\Lambda|\,\mu_\Psi\beta^\dagger & i\in\Lambda^c_{\gamma^\star}\ . \\ |\Lambda|\,\mu_\Psi\beta^\star & \gamma\neq\gamma^\star, i\in\Lambda^c_\gamma\ . \\ \beta^\dagger+|\Lambda|\,\mu_\Psi\beta^\star & \gamma\neq\gamma^\star, i\in\Lambda_\gamma\ . \end{cases}$$

$$\leq \begin{cases} N\rho\beta^\star & i\in\Lambda^c_{\gamma^\star}\ . \\ N\beta^\star & \gamma\neq\gamma^\star, i\in\Lambda^c_\gamma\ . \\ (\rho+N)\beta^\star & \gamma\neq\gamma^\star, i\in\Lambda_\gamma\ . \end{cases} \tag{10}$$

Therefore,

$$\max_{(\gamma,i)\in C_{\gamma^\star}}\left|\left\langle \mathbf{R}^{(h)}, \psi_\gamma[i]\right\rangle\right| \leq (\rho+N)\beta^\star\ . \tag{11}$$

This bound is attained for an atom that belongs to the support of an incorrect morphological component.

Combining the lower-bound of (9) and the upper-bound (11), we get that for (6) to hold, we need that

$$(\rho+N) < \frac{(1+\rho)(1-N)}{2}\ , \tag{12}$$

or equivalently $N < \frac{1-\rho}{3+\rho}$. For $0<\rho\leq 1/5$, this is a strictly decreasing function of ρ bounded below by $1/4$, yielding the bound of the theorem on K, hence validating the Correct Rejection Property.

Remark 1. There is a room for improving the sparsity bound of Theorem 1(i) if the Correct Rejection Property is relaxed. It is easy to see that :

- A sufficient condition for the GMCA/mMOM to reject all atoms indexed by $\Lambda^c_{\gamma^\star}$ is $N<1/2$ and $\rho\leq 1/3$.
- GMCA/mMOM will never select any atom $(\gamma,i)\in\{\gamma\neq\gamma^\star, i\in\Lambda^c_\gamma\}$ if $N<1/3$ and $\rho\leq 1/2$.

Proof of (ii) We will show that at the h-step, under the conditions of (ii),

$$\exists i\in\Lambda_{\gamma^\star} \text{ such that } \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i]\right\rangle\right| \geq \lambda^{(h)}\ . \tag{13}$$

As $\rho \leq \frac{1-N_0}{1+3N_0} < \frac{1-N_0}{1+N_0}$ for $0 < N_0 < 1$, the morphological component $\gamma^\star$ is $\frac{1-N_0}{1+N_0}$-separable. Thus, the upper-bound of (9) holds. Furthermore,

$$\begin{aligned}\max_{i\in\Lambda_{\gamma^\star}} \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i]\right\rangle\right| &\geq \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i^\star]\right\rangle\right| \\ &\geq \left|\beta^\star + \sum_{\gamma\neq\gamma^\star}\sum_{i\in\Lambda_\gamma} \left\langle \psi_\gamma[i], \psi_{\gamma^\star}[i^\star]\right\rangle \beta_\gamma[i]\right| \\ &\geq \beta^\star(1-\rho N) \ . \end{aligned} \tag{14}$$

For (13) to hold, it is sufficient to have

$$\frac{(1+\rho)(1+N)}{2} < 1-\rho N \ , \tag{15}$$

or equivalently $N < \frac{1-\rho}{1+3\rho}$. This upper-bound is a strictly decreasing function bounded below by N_0 for $\rho \in (0, \frac{1-N_0}{1+3N_0}]$. Since N was chosen $< N_0$, the bound (13) follows.

Proof of Proposition 1

Proof. Our GMCA/mMOM strategy and its proof is a stagewise extension of the greedy MP. As for MP and OMP, the proposition is shown by induction. GMCA/mMOM begins by setting $\mathbf{R}^{(0)} = \mathbf{X}$. By hypothesis in the proposition, at each step $h \geq 0$, GMCA/mMOM has the Exact Selection Property. Thus,

$$\forall h \geq 0;\ \mathbf{R}^{(h)} = \sum_\gamma \sum_{i\in\Lambda_\gamma} \psi_\gamma[i]\beta_\gamma[i] = \mathbf{\Psi}_\Lambda \beta[\Lambda] \in \text{span}\left(\mathbf{\Psi}_{\Lambda_\gamma}, \forall\gamma\right) \ .$$

where $\beta[\Lambda]$ is the restriction of β to its support Λ, and $\mathbf{\Psi}_\Lambda$ is the restriction of $\mathbf{\Psi}$ to the atoms indexed by Λ. GMCA/mMOM then chooses correct atoms only from $\Lambda_{\gamma^\star}$ by hard-thresholding, and then calculates a new residual such that :

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

From the orthonormality of the atoms in $\mathbf{\Psi}_{I_{\gamma^\star}}$, it is easy to see that,

$$\begin{aligned}\|\mathbf{R}^{(h+1)}\|^2 &= \|\mathbf{R}^{(h)}\|^2 - \sum_{i\in I_{\gamma^\star}} \left|\left\langle \mathbf{R}^{(h)}, \psi_{\gamma^\star}[i]\right\rangle\right|^2 \\ &\leq \left\|\mathbf{R}^{(h)}\right\|^2 - \left\|\mathbf{R}^{(h)}\mathbf{\Psi}_{I_{\gamma^\star}}\right\|_{\ell_\infty}^2 \ . \end{aligned} \tag{16}$$

The inequality follows from the fact that $I_{\gamma^\star} \neq \emptyset$ under the Correct Selection Property.

We also have,

$$\left\|\mathbf{R}^{(h)}\mathbf{\Psi}_{I_{\gamma^\star}}\right\|_{\ell_\infty} = \left\|\mathbf{R}^{(h)}\mathbf{\Psi}_{\Lambda_{\gamma^\star}}\right\|_{\ell_\infty} = \left\|\mathbf{R}^{(h)}\mathbf{\Psi}_{\Lambda}\right\|_{\ell_\infty} \ .$$

The first equality is a consequence of the Correct Selection Property (see (14)), and the second one follows from Lemma 1 as $\rho \leq 1/5 < 3/5$ the separability threshold of the morphological component $\gamma^\star$.

We then obtain,

$$\|\mathbf{R}^{(h+1)}\|^2 \leq \left\|\mathbf{R}^{(h)}\right\|^2 - \left\|\mathbf{R}^{(h)}\mathbf{\Psi}_{\Lambda}\right\|_{\ell_\infty}^2 \ . \tag{17}$$

The rest of the proof is merely the same as in [1, Theorem 2], with arguments relying on DeVore and Temlyakov Lemma [3], and Gershgorin Disc Theorem, see [2, Lemma 2.3 and Proposition 2.1]. We finally obtain,

$$\|\mathbf{R}^{(h+1)}\|^2 \leq \tau\|\mathbf{R}^{(h)}\|^2 \leq \ldots \leq \tau^{-(h+1)}\|\mathbf{X}\|^2 \ , \tag{18}$$

where $0 \leq \tau = \frac{(K-1)(1+\mu_\Psi)}{K} < 1 - \frac{3}{4K}$ as $K\mu_\Psi < 1/4$. This completes the proof.

References

1. Gribonval, R., Nielsen, M.: On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. IEEE Transactions on Information Theory **52** (2006) 255–261
2. Tropp, J.: Greedy is good : algorithmic results for sparse approximation. IEEE Transactions on Information Theory **50** (2004) 2231–2242
3. DeVore, R., Temlyakov, V.N.: Some remarks on greedy algorithms. Adv. Comput. Math. **5** (2007) 173–187