Morphological Component Analysis: An Adaptive Thresholding Strategy

Jérôme Bobin, Jean-Luc Starck, Jalal M. Fadili, Yassir Moudden, and David L. Donoho

Abstract—In a recent paper, a method called morphological component analysis (MCA) has been proposed to separate the texture from the natural part in images. MCA relies on an iterative thresholding algorithm, using a threshold which decreases linearly towards zero along the iterations. This paper shows how the MCA convergence can be drastically improved using the mutual incoherence of the dictionaries associated to the different components. This modified MCA algorithm is then compared to basis pursuit, and experiments show that MCA and BP solutions are similar in terms of sparsity, as measured by the ℓ_1 norm, but MCA is much faster and gives us the possibility of handling large scale data sets.

Index Terms—Feature extraction, morphological component analysis (MCA), sparse representations.

I. INTRODUCTION

I N a series of recent papers [1]–[4], the morphological component analysis (MCA) concept was developed and it has been shown that MCA can be used for separating the texture from the piecewise smooth component [2], for inpainting applications [3] or more generally for separating several components which have different morphologies. MCA has also been extended to the multichannel case in [5], [6]. The main idea behind MCA is to use the morphological diversity of the different features contained in the data, and to associate each morphology to a dictionary of atoms for which a fast transform is available. Thanks to recent developments in harmonic analysis, many new multiscale transforms are now available [7]–[10], which greatly increases the potentiality of MCA.

In Section I, we recall the MCA methodology. In [2], the authors introduced the MCA algorithm based on an iterative thresholding scheme depending on two key parameters: the threshold and the number of iterations. In Section I-B, we introduce a new way of tuning the threshold. We show that this new strategy called MOM, for "mean of max," is very practical as it is adaptive to the data, and has a finite stopping-behaviour

since it alleviates us from the delicate choice of the number of iterations. We also prove conditions that guarantee no false detections with overwhelming probability. Numerical examples are provided in Section II-A.

Furthermore, MCA/MOM also turns out to be a practical way to decompose a signal in an overcomplete representation made of a union of bases. In Section II-B, we provide a comparison between MCA and Basis Pursuit (BP) which clearly indicates that MCA/MOM is a practical alternative to classical sparse decomposition algorithms such as BP [11].

II. NEW THRESHOLDING STRATEGY IMPROVING MCA

A. Overview of MCA

For a given n-sample signal y consisting of a sum of K signals $\{y_i\}_{i=1,...,K}, y = \sum_{i=1}^{K} y_i$, having different morphologies, MCA assumes that a dictionary of bases $\{\Phi_1, \ldots, \Phi_K\}$ exists such that, for each i, y_i is sparse in Φ_i and not, or at least not as sparse, in other Φ_j $(j \neq i) : \forall j \neq i, \left\| \Phi_i^T y_i \right\|_0 < \left\| \Phi_j^T y_i \right\|_0$, where $\|x\|_0$ denotes the ℓ_0 pseudo-norm of the vector x (i.e., the number of nonzero coefficients of x). For the sake of simplicity, we will consider in this paper a number of morphological components K equal to two, but our results can be easily extended to any $K \geq 2$. We note $\alpha_{i=1,2}$ the decomposition of $y_{i=1,2}$ in $\Phi_{i=1,2} : \alpha_{i=1,2} = \Phi_{i=1,2}^T y_{i=1,2}$. In [1] and [2], it was proposed to estimate the components y_1 and y_2 by solving the following constrained optimization problem:

$$\min_{y_1, y_2} \left\| \Phi_1^T y_1 \right\|_1 + \left\| \Phi_2^T y_2 \right\|_1 \text{ s.t. } \|y - y_1 - y_2\|_2 \leq \sigma \quad (1)$$

where σ is the noise standard deviation in the noisy case. For now, we assume that no noise perturbs the data y, $\sigma = 0$ (i.e., equality constraint). Extension of our method to deal with noise is discussed in Section I-D.

The MCA algorithm given below relies on an iterative alternate matched filtering and thresholding scheme.

1) Set the number of iterations $I_{\max}$ and thresholds $\lambda_1^{(1)}$ and $\lambda_2^{(1)}$

2) While $\lambda_{i=1,2}^{(k)}$ is higher than a given lower bound $\lambda_{\min}$ (e.g., can depend on the noise variance),

— Proceed with the following iteration to estimate components $\{y_j\}_{j=1,\ldots,2}$ at iteration k:

For $j = 1, \ldots, 2$

• Compute the residual term $r_j^{(k)}$ assuming the current estimates of $y_{i \neq j}$, $\tilde{y}_{i \neq j}^{(k-1)}$ is fixed

$$r_j^k = y - \tilde{y}_{i \neq j}^{(k-1)}.$$

• Estimate the current coefficients of $\tilde{y}_j^{(k)}$ by hard thresholding with threshold $\lambda_j^{(k)}$

$$\tilde{\alpha}_j^{(k)} = \delta_{\lambda_j^{(k)}} \left(\Phi_j^T r_j^{(k)} \right).$$

• Get the new estimate of y_j by reconstructing from the selected coefficients $\tilde{\alpha}_j^{(k)}$

$$\tilde{y}_j^{(k)} = \Phi_j \tilde{\alpha}_j^{(k)}.$$

— Decrease the thresholds $\lambda_{i=1,2}$ following a given strategy

At the kth iteration, we have two estimates $\tilde{y}_1^{(k)}, \tilde{y}_2^{(k)}$ of y_1 and y_2 ($\tilde{y}_1^{(0)} = 0, \tilde{y}_2^{(0)} = 0$). Then $\tilde{y}_1^{(k+1)}$ and $\tilde{y}_2^{(k+1)}$ are obtained as follows:

$$\tilde{y}_1^{(k+1)} = \Delta_{\Phi_1, \lambda_k} \left(y - \tilde{y}_2^{(k)} \right)$$
$$\tilde{y}_2^{(k+1)} = \Delta_{\Phi_2, \lambda_k} \left(y - \tilde{y}_1^{(k+1)} \right) \tag{2}$$

where the operator $\Delta_{\mathcal{D}, \lambda}(x)$ consists in decomposing x over the transform $\mathcal{D}$ ($\alpha = \mathcal{D}^T x$, i.e., matched filtering), thresholding the obtained coefficients α with the threshold λ ($\tilde{\alpha} = \delta_\lambda(\alpha)$), and reconstructing $\tilde{x}$ from $\tilde{\alpha}$ ($\tilde{x} = \mathcal{D}\tilde{\alpha}$) : $\Delta_{\mathcal{D}, \lambda}(x) = \mathcal{D}\left[\delta_\lambda\left(\mathcal{D}^T x\right)\right]$. The thresholding operator δ_λ can be either a hard or a soft thresholding. In practice, hard thresholding leads generally to better results [1], [2]. The threshold λ_k decreases linearly towards zero

$$\lambda_k = \lambda_1 - (k-1) \frac{\lambda_1 - \lambda_{\min}}{I_{\max} - 1} \tag{3}$$

where λ_1 is the first threshold and $I_{\max}$ is the number of iterations. Both are parameters of the method. The first threshold λ_1 can be set automatically to a large enough value (for instance the maximum in magnitude of all coefficients $\lambda_1 = \max \left\{ \left\| \Phi_1^T y \right\|_\infty, \left\| \Phi_2^T y \right\|_\infty \right\}$). For an exact representation of the data with the morphological components, $\lambda_{\min}$ must be set to zero. When noise is contained in the data, as discussed in Section I-D, $\lambda_{\min}$ should be set to few times the noise standard deviation. Intuitively, MCA distinguish between the morphological components by taking advantage of the mutual coherence of the subdictionaries Φ_1 and Φ_2. Quantitatively, the mutual coherence of a dictionary $\Phi = [\Phi_1 \quad \Phi_2]$ is defined, assuming that its columns are normalized to unit ℓ_2-norm, in terms of the Gram matrix $G = \Phi^T \Phi$

$$\mu_\Phi = \max_{i \neq j} |G(i, j)|.$$

The "mutual incoherence" of the dictionaries is a key assumption for the MCA algorithm; it states that the image we wish to decompose contains features with different morphologies which are sparse in different representations. MCA will provide a good components separation by using this morphological diversity concept when the transforms amalgamated in the dictionary are mutually incoherent enough.

Although the simplicity of this linear strategy is an advantage, there is no way to estimate the minimum number of iterations yielding a successful separation. Too small a number of iterations leads to a bad separation while a large number of iterations is computationally expensive. Experiments have clearly shown that the optimal number of iterations depends on the data (typically hundreds). Therefore, a good thresholding strategy must provide a fast decomposition with the least number of iterations $I_{\max}$.

B. "Mean-of-Max"—The Intuition

As we stated before, the choice of the threshold management strategy is vital towards good performance of the MCA separation mechanism. A "good" thresholding strategy is one that meets three main requirements: i) applicability to the data for a wide range of dictionary pairs, ii) low global computational cost and iii) no need to prespecify a number of iterations. Thus, in this section, we discuss a new threshold management method aiming at satisfying the above three requirements.

We assume now that the entries of $\alpha_{i=1,2} = [\alpha_{i=1,2}(j)]_{j=1,\ldots,n}$ (where n is the number of entries of each vector $\alpha_{i=1,2}$) are identically and independently distributed (*iid*) random variables. We further assume a sparse distribution so that most of the samples of a given $\alpha_{i=1,2}$ are zeros. We can, thus, divide this vector into two subsets. The first one contains the nonzero coefficients $S_{i=1,2}$. This set is called the support of $\alpha_{i=1,2}$. The second one is complementary to the latter $\bar{S}_{i=1,2}$. During the MCA process, at the kth iteration in the Φ_1 transform, MCA selects coefficients from the following vector:

$$\gamma_1^{(k)} = \Phi_1^T \left(y - \tilde{y}_2^{(k-1)} \right) \tag{4}$$
$$= \alpha_1 + \epsilon_1^{(k)} \tag{5}$$

which is a perturbed version of α_1 with an error term $\epsilon_1^{(k)}$. This term is due to i) the imperfect estimation of y_2 (i.e., some of the coefficients in α_2 have not yet been recovered), and ii) the mutual coherence of the transforms Φ_1 and Φ_2. We define $\bar{\alpha}_{i=1,2}^{(k-1)} = \alpha_{i=1,2} - \tilde{\alpha}_{i=1,2}^{(k-1)}$, where $\tilde{\alpha}_{i=1,2}^{(k-1)} = \Phi_{i=1,2}^T \tilde{y}_{i=1,2}^{(k-1)}$. This quantity is the estimation error of the ith morphological component in the coefficients domain. Without loss of generality, let us assume that at iteration k, $\left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty > \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty$. This equation states that the most significant (in terms of ℓ_∞ norm) coefficient that has not been selected yet belongs to the support of α_1. We then discuss the way we ought to tune $\lambda_1^{(k)}$ so as to estimate α_1 at iteration k. The converse case $\left(\left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty < \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty \right)$ would yield the same discussion. Concretely, we would like to update $\tilde{\alpha}_1^{(k-1)}$ avoiding false detections and, thus, fix the threshold $\lambda_1^{(k)}$ in order to select new coefficients in the support S_1. These new coefficients are guaranteed to be in the support of α_1 with overwhelming probability if their amplitudes are higher
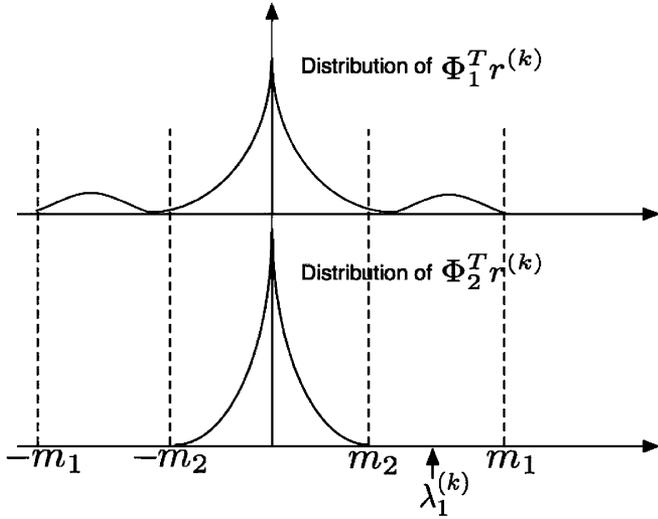
Fig. 1. Rationale governing the MCA/MOM thresholding strategy.

than the highest (in absolute value) amplitude of the entries of the nuisance term $\epsilon_1^{(k)}$

$$\epsilon_1^{(k)} = \Phi_1^T \Phi_2 \bar{\alpha}_2^{(k-1)}. \tag{6}$$

Owing to the last argument, $\lambda_1^{(k)}$ has to be chosen such that

$$\left\| \Phi_1^T \Phi_2 \bar{\alpha}_2^{(k-1)} \right\|_\infty < \lambda_1^{(k)} < \left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty. \tag{7}$$

Basically, (7) says that to qualify, the chosen threshold must avoid false detections. The upper bound in (7) ensures that at least one coefficient in the support of α_1 will be recovered at iteration k. Nonetheless, these two bounds depend on the true coefficients and cannot be computed in practice.

Instead, we intuitively guess that knowing the maximum values of the total residual $r^{(k)} = y - \tilde{y}_1^{(k-1)} - \tilde{y}_2^{(k-1)}$ in the Φ_1 and Φ_2 domains will give us some information about the *relative* values of the useful parts $\alpha_{i=1,2}$ and the perturbation terms $\epsilon_{1,2}^{(k)}$. Let us have a look at the following quantities: $\Phi_{i=1,2}^T r^{(k)} = \Phi_{i=1,2}^T \left(\Phi_1 \bar{\alpha}_1^{(k-1)} + \Phi_2 \bar{\alpha}_2^{(k-1)} \right)$. Note that these quantities can be easily computed online. The top (respectively, bottom) picture in Fig. 1 depicts an intuitive view of $\Phi_1^T r^{(k)}$ (respectively, $\Phi_2^T r^{(k)}$). In these plots, we define $m_{i=1,2} = \left\| \Phi_{i=1,2}^T r^{(k)} \right\|_\infty$.

Note that, for extremely incoherent dictionaries, the error terms $\epsilon_2^{(k)}$ are negligible, and $m_2 \approx \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty$. Furthermore, according to the incoherence of the subdictionaries Φ_1 and Φ_2, $\bar{\alpha}_2^{(k-1)}$ is sparser than $\Phi_1^T \Phi_2 \bar{\alpha}_2^{(k-1)}$. This entails that with high probability

$$m_2 > \left\| \Phi_1^T \Phi_2 \bar{\alpha}_2^{(k-1)} \right\|_\infty. \tag{8}$$

As a consequence, choosing $\lambda_1^{(k)}$ such that $m_1 > \lambda_1^{(k)} > m_2$ intuitively fulfills the conditions in (7).

C. MOM Strategy: Analysis

In this section, we give arguments to support and prove that the following choice:

$$\lambda_{i=1,2}^{(k)} = \frac{1}{2} [m_1 + m_2] \tag{9}$$

is sufficient for the set of inequalities (7) to hold true under some mild nonrestrictive conditions. This strategy coined MOM, is illustrated in Fig. 1 as explained in the previous section. A key advantage is that it only requires to assess the maximum absolute values of the total residual transformed in each basis Φ_1 and Φ_2 of the dictionary Φ (m_1 and m_2 in Fig. 1). Compared to the initial linear decrease strategy, MOM resorts to an additional use of the implicit transform operators associated to $\Phi_{i=1,2}$ which induces an additional computational cost (typically $O(n)$ or $O(n \log n)$ for most popular transforms). It does not require any other parameters and provides a data-adaptive strategy. We define $\phi_{1,j}$ by the jth column of Φ_1. Before proving the sufficiency of the bounds (7), let us first note that

$$\left\| \Phi_1^T \Phi_2 \bar{\alpha}_2^{(k-1)} \right\|_\infty = \max_{\phi_{1,j} \in \Phi_1} \left| \phi_j^T \Phi_2 \bar{\alpha}_2^{(k-1)} \right| \tag{10}$$

$$\leq \mu_\Phi \left\| \bar{\alpha}_2^{(k-1)} \right\|_1 \tag{11}$$

$$= R_{1,\infty} \left(\bar{\alpha}_2^{(k-1)} \right) \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty \tag{12}$$

where $R_{1,\infty} \left(\bar{\alpha}_2^{(k-1)} \right) = \mu_\Phi \left\| \bar{\alpha}_2^{(k-1)} \right\|_1 / \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty \leq \mu_\Phi \left\| \bar{\alpha}_2^{(k-1)} \right\|_0$. Through iterations of MCA, it is easy to estimate online the following quantities:

$$m_1 = \left\| \Phi_1^T \left(y - \tilde{y}_1^{(k-1)} - \tilde{y}_2^{(k-1)} \right) \right\|_\infty \tag{13}$$

$$m_2 = \left\| \Phi_2^T \left(y - \tilde{y}_1^{(k-1)} - \tilde{y}_2^{(k-1)} \right) \right\|_\infty. \tag{14}$$

Straightforward calculations give the following inequalities:

$$m_1 \geq \left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty - R_{1,\infty} \left(\bar{\alpha}_2^{(k-1)} \right) \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty$$

$$\text{and } m_1 \leq \left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty + R_{1,\infty} \left(\bar{\alpha}_2^{(k-1)} \right) \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty \tag{15}$$

and similarly for m_2. Let us denote $R = \max_i \left\{ R_{1,\infty} \left(\bar{\alpha}_{i=1,2}^{(k-1)} \right) \right\}$. As MOM selects new coefficients in Φ_1, at iteration k, we can write $\left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty = \rho \left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty$ with $0 < \rho < 1$. Then, the following inequalities are easy to derive:

$$\frac{m_1 + m_2}{2} \geq \frac{(1+\rho)(1-R)}{2\rho} \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty$$

$$\text{and } \frac{m_1 + m_2}{2} \leq \frac{(1+\rho)(1+R)}{2} \left\| \bar{\alpha}_1^{(k-1)} \right\|_\infty. \tag{16}$$

Hence, defining $\lambda_1^{(k)} = m_1 + m_2/2$ is sufficient to guarantee the inequalities in (7) under the following conditions:

$$\frac{(1+\rho)(1-R)}{2\rho} \left\| \bar{\alpha}_2^{(k-1)} \right\|_\infty \geq \mu_\Phi \left\| \bar{\alpha}_2^{(k-1)} \right\|_1 \qquad (17)$$

$$\frac{(1+\rho)(1+R)}{2} \leq 1. \qquad (18)$$

A straightforward development of (17) provides the next condition

$$R \leq \frac{1+\rho}{1+3\rho}. \qquad (19)$$

As $0 < \rho < 1$, the right hand side of (19) is bounded below by $1/2$. Thus, (17) is always verified if $R \leq 1/2$. Recalling that $R \leq \max_i \mu_\Phi \left\| \bar{\alpha}_i^{(k-1)} \right\|_0$, it is worth noting that if at iteration k, we have:

$$\max_{i=1,2} \left\| \bar{\alpha}_i^{(k-1)} \right\|_0 \leq \frac{\mu_\Phi^{-1}}{2} \qquad (20)$$

then the condition $R \leq 1/2$ is necessarily satisfied. Note that the above inequality has a flavor of the sparsity criterion ensuring unique maximally sparse solutions of ℓ_1-norm under-determined linear systems in [12]–[15]. In words, the lower bound we arrive at stipulates that, as long as the rest of coefficients to be identified at iteration k, in the support S_1 (resp., S_2) is sparse enough, then with a high probability, we are ensured to avoid false detections (i.e., coefficients not belonging S_1). That is, the atoms newly selected by the MCA/MOM algorithm at each iteration to enter the active set are guaranteed (with over-whelming probability) to belong to the support of the true solution. Thus, for rather sparse signals, (17) is valid entailing the lower bound inequality $\lambda_1^{(k)} \geq \left\| \Phi_1^T \Phi_2 \bar{\alpha}_2^{(k-1)} \right\|_\infty$. The same correct term selection property with a condition similar to (20) has also been shown in a recent paper of Donoho and Tsaig [16], when studying the properties of the Homotopy algorithm. Note that the Homotopy method is different from MCA; the Homotopy is an exact path following that identifies *one* coefficient of the solution at a time entering or leaving the active set, while MCA/MOM is an adaptive stagewise iterative thresholding that follows the solution path by identifying *groups* of coefficients entering the active set at each iteration.

As far as the upper bound inequality in (7) is concerned, [18] implies that

$$R \leq \frac{1-\rho}{1+\rho}. \qquad (21)$$

Note that although this condition is much more restrictive [condition (18) is not always verified], it is also less important as it only guarantees MCA to pick at least one new coefficient at each iteration. Let us add that in the case where the number of morphological components $K > 2$, the threshold in (9) is computed as the mean of the two dominant terms of the set $\left\{ \left\| \Phi_i^T \left(y - \sum_{j=1}^K \tilde{y}_j^{(k)} \right) \right\|_\infty \right\}_{i=1,\dots,K}$.

D. Handling Additive Gaussian Noise

From a probabilistic point of view, the ℓ_2-norm constraint in (1) is equivalent to the anti-loglikelihood assuming the data are perturbed by an additive white Gaussian noise (AWGN). Hence, MCA can handle intrinsically data perturbed by additive Gaussian noise. Furthermore, MCA being a thresholding-based algorithm, it is, therefore, very robust to the noise since thresholding techniques belong to the best approaches for noise removal [17] in image processing. MCA is an iterative coarse-to-fine (in terms of coefficient amplitude in the dictionary) and AWGN noise can be handled by just stopping iterating when the residual is at the noise level. Assuming that the noise variance σ^2 is known, the algorithm stops at iteration k when the ℓ_2-norm of the residual satisfies $\|r^{(k)}\|_2 \leq \sqrt{n}\sigma$. Another way consists in using a similar strategy as in denoising methods, and the iteration will stop when $\lambda^{(k)}$ is smaller than $\tau\sigma$, where τ is a constant, typically between 3 and 5. Put formally, the convergence is achieved when the residual satisfies: $\max\left\{ \left\| \Phi_1^T r^{(k)} \right\|_\infty, \left\| \Phi_2^T y^{(k)} \right\|_\infty \right\} < \tau\sigma$. A similar criterion was used in the combined (wavelet/curvelet) filtering method [18].

For non-Gaussian noise, a similar strategy as before could also be used, but a noise modeling step in the transform domain must be accomplished in order to derive the probability density function for each coefficient of the dictionary to be due to the noise. Then the convergence criterion can be recast as a test of the null hypothesis: remaining α coefficients are not significant. This is formally written $\Pr\left(\max\left\{ \left\| \Phi_1^T r^{(k)} \right\|_\infty, \left\| \Phi_2^T r^{(k)} \right\|_\infty \right\} < \alpha_\epsilon \right) > \epsilon$. The distribution of many noise models in different dictionaries such as wavelets or curvelets have been proposed in the literature [19]. Another line of action to handle some specific noise distributions such as Poisson or Rayleigh noise is to use variance stabilizing transforms [19]–[21], and the problem is brought back to the Gaussian case.

E. Related Work

Donoho *et al.* in [22] have also recently focused on iterative thresholding scheme applied to solving under-determined linear sparse problems. Interestingly, models similar to theirs can be easily transposed to our MCA setting. In their context, the perturbation term $\epsilon_1^{(k)}$ [see (5)] is assumed to be a Gaussian process whose samples are *iid*. Estimating the standard deviation $\sigma_{\epsilon_1}^{(k)}$ at iteration k using a *median absolute deviation* estimate provides a straightforward thresholding strategy: $\lambda_1^{(k)} = t\sigma_{\epsilon_1}^{(k)}$ (in practice, $t = 3$) the validity of which stands as long as the gaussianity of $\epsilon_1^{(k)}$ is a valid assumption (i.e., for incoherent dictionaries with low mutual coherence μ_Φ).

F. Success Phase Diagram

In order to evaluate the success/failure of MCA, we define $\mathcal{BG}(p_1, p_2)$ as the class of signals such that if y belongs to $\mathcal{BG}(p_1, p_2)$, it is the linear combination $y = \Phi_1\alpha_1 + \Phi_2\alpha_2$ where α_1 and α_2 are Bernoulli-Gaussian random vectors the samples of which are nonzero with probabilities p_1 and p_2 respectively. Nonzero coefficients are generated from a Gaussian law with mean zero and variance 1. Fig. 1 illustrates
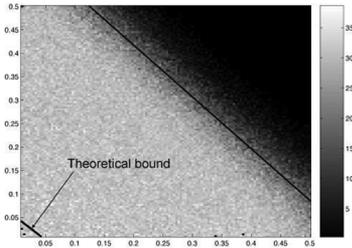
Fig. 2. Opposite ℓ_2 estimation error in decibels of the first morphological component extracted with MCA/MOM ($-20\log_{10}(\|y_1 - \tilde{y}_1\|_2)$). In abscissa: p_1 and in ordinate: p_2.

Fig. 3. Left: Original image (size 128×128) used to compare the (dashed) linear and the (solid) MOM strategies. Middle: Number of recovered coefficients $\|\alpha_1\|_0 + \|\alpha_2\|_0$ while the algorithm is in progress. In abscissa: Number of iterations in log scale. Right: Reconstruction error ℓ_2-norm $\|y - \bar{y}_1 - \bar{y}_2\|_2$ as a function of the number of iterations.

Fig. 4. Left: Original signal. Middle: First morphological component (curvelet transform) extracted with MCA/MOM. Right: Second morphological component (global DCT) extracted with MCA/MOM.

the efficiency of MCA/MOM. It shows the opposite ℓ_2 estimation error of the first morphological component extracted with MCA/MOM ($-20\log_{10}(\|y_1 - \tilde{y}_1\|_2)$) for different values of p_1 and p_2 between 0.01 and 0.5. Each point has been evaluated from 25 different decompositions of signals (the number of samples is $n = 512$). The redundant dictionary used in this experiment is the Spikes/Sines dictionary (in practice, Φ is the concatenation of the identity matrix and the DCT matrix). One observes what can be interpreted as a rather sharp phase transition. For couples (p_1, p_2) below the phase transition, the estimation of the morphological components is successful. The bottom-left bound is the theoretical recovery bound (in this case, this bound is equal to $p_1 + p_2 = 1/n\mu_\Phi$; see [23]). The top-right bound is the MCA empirical bound defined such that the natural logarithm of the estimation error is less than -20 dB. Clearly, the theoretical bound is too pessimistic and MCA/MOM has a wider successful area.

III. EXPERIMENTAL RESULTS

A. Texture/Natural Part Separation

Here, we compare the decompositions obtained with the linear strategy and the MOM. The redundant dictionary we used is composed of the union of the curvelet transform and the global DCT. Previous results (see Section II) have been proven assuming that Φ_1 and Φ_2 were orthogonal. In [24], the author shows that thresholding is still relevant for overcomplete representations for ℓ_1 sparse minimization. Furthermore, in [22], the authors developed a thresholding-based sparse decomposition algorithm in overcomplete representations. Dealing with redundant subdictionaries such as the curvelet representation is then still reliable. In practice, even if the curvelet transform is not orthogonal, the same MOM thresholding strategy is used and achieves astounding results.

Fig. 3 shows that MCA/MOM is more adaptive than MCA/Linear. Indeed, it continually selects new atoms (plot in the middle of Fig. 3) while MCA/Linear selects most of the coefficients in the last iterations. The same phenomenon is observed in the plot on the right. It shows the evolution of the ℓ_2-norm of the total residual: with MCA/MOM it decreases continually down to zero. These two plots shed light on how the MOM strategy adapts to the data. Indeed, the MOM strategy computes a new threshold at each iteration by taking into account online the evolution of the total residual term (in ℓ_∞ norm). Conversely, the linear strategy is not concerned

with the way the coefficients are distributed. Fig. 4 shows the result of a texture-cartoon separation experiment. The first morphological component is depicted in the middle of Fig. 4 and the second one on the right. The picture on the left depicts the original image. The overcomplete dictionary in this case is taken as the union of the curvelet transform and the global DCT. Hence, MCA/MOM extracts in the DCT basis a globally oscillatory part of the original image and leaves in the first morphological component edges and piece-wise smooth parts (i.e., cartoon). Visually, MCA/MOM performs well as a feature extraction/separation algorithm. Compared to MCA/linear that needs hundreds of iterations to converge, only 25 iterations were required for a similar separation quality. Note that the total variation (TV) constraint (see [2]) has not been taken into account here. Adding such a constraint would certainly improve slightly the result.

B. Comparing MCA and Basis Pursuit

So far, MCA has been considered as an efficient feature extraction/separation algorithm based on sparse decompositions. In practice, we also observe that MCA gives rather good results as a sparse decomposition algorithm in a union of bases. Several references ([12], [23], and references therein) have shown that some algorithms such as BP solve the ℓ_1 minimization problem and even the ℓ_0 minimization problem provided that some conditions are verified. The results of some experiments we carried out in order to compare BP and MCA/MOM are pictured in Fig. 5. The graphs were computed as an average over 50 decompositions of signals belonging to the Bernoulli–Gaussian model $\mathcal{BG}(p_1, p_2)$ with random p_1 and p_2. The dictionary Φ used in this experiment is the union of an orthogonal 1-D wavelet transform and the global DCT. The top-left plot in Fig. 5 shows the ℓ_1/ℓ_2-curve. The top-right graph depicts the ℓ_0/ℓ_1-curve: the MCA/MOM curve is below the BP curves. Then, in terms of ℓ_1 norm sparsity, MCA/MOM and BP performs rather similarly.
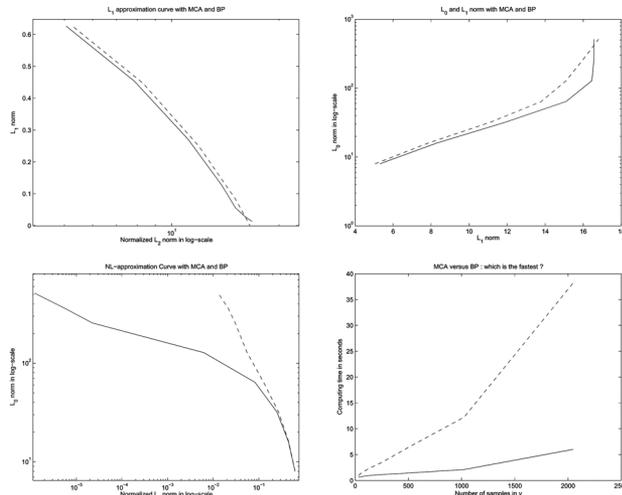
Fig. 5. Top left: ℓ_1/ℓ_2-curve of both (solid line) MCA/MOM and (dashed line) BP. Top right: ℓ_0/ℓ_1-curve of both (solid line) MCA/MOM and (dashed line) BP. Bottom left: Nonlinear approximation curve (ℓ_0/ℓ_2) of both (solid line) MCA/MOM and (dashed line) BP. Bottom right: Comparison of execution times (in seconds): (solid line) MCA/MOM and (dashed line) BP.

Finally, the bottom-left picture features the nonlinear approximation error curve computed from the set of coefficients estimated with both BP and MCA/MOM. According to the success phase diagram in Fig. 1, for most signals belonging to $\mathcal{BG}(p_1, p_2)$, ℓ_0 and ℓ_1 sparse problems are not equivalent. Indeed, as it uses a hard thresholding scheme, MCA/MOM is likely to converge to a ℓ_0 sparse solution. On the contrary, BP solves the ℓ_1 sparse problem. Then, in terms of ℓ_0 sparsity, MCA/MOM achieves better nonlinear approximation as illustrated in the bottom-left plot of Fig. 5. The bottom right of Fig. 5 shows how the computational cost increases as the number of samples n increases. MCA/MOM is clearly faster than BP. To conclude this section, MCA/MOM is a practical, fast and efficient way to decompose signals in a union of bases.[1]

IV. CONCLUSION

As a feature extraction/separation algorithm, MCA proved its efficiency. However, previous work did not propose any carefully-designed strategy to manage the threshold. This paper first insists on the impact of a particular choice of a thresholding scheme. Then, we exhibit a new advantageous way to tune the thresholds which we called MOM. The MOM strategy differs from other heuristics in the sense that it is fast, accurate and adaptive. It also no longer requires hand-tuning parameters as the MOM strategy is parameter free. Last, but not least, we put forward conditions that guarantee, with overwhelming probability, a good performance in the estimation of the morphological components as they particularly avoid false detections. We also showed that MCA is at least as efficient as BP in achieving sparse decompositions in redundant dictionaries. As it performs drastically faster than BP, MCA/MOM provides a practical alternative to this well-known sparse decomposition algorithm. In

the future, it would be interesting to extend MCA/MOM to be able to deal with other types of noise. A multichannel version of MCA has also recently been proposed in which MOM could be very useful [5], [6].

REFERENCES

[1] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Adv. Imag. Electron. Phys.*, vol. 132, 2004.

[2] J.-L. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, pp. 1570–1582, 2005.

[3] M. Elad, J.-L. Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *ACHA*, 2005.

[4] M. J. Fadili and J.-L. Starck, "Sparse representations and bayesian image inpainting," presented at the SPARS, 2005.

[5] J. Bobin, Y. Moudden, J.-L. Starck, and M. Elad, "Morphological diversity and source separation," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 409–412, Jul. 2006.

[6] J. Bobin, Y. Moudden, and J.-L. Starck, "Enhanced source separation by morphological component analysis," presented at the ICASSP, 2006.

[7] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, vol. 5/3, pp. 861–899, 2006.

[8] J.-L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.

[9] E. L. Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.

[10] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[11] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[12] D. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2003.

[13] A. Bruckstein and M. Elad, "A generalized uncertainty principle and sparse representation in pairs of $\mathbf{r}^n$ bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[14] T. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1030–1051, 2006.

[15] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1341–1344, 2004.

[16] D. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *Preprint*, Oct. 2006.

[17] J.-L Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[18] J.-L Starck, D. L. Donoho, and E. J. Candès, A. F. Laine, M. A. Unser, and A. Aldroubi, Eds., "Very high quality image restoration by combining wavelets and curvelets," *Proc. SPIE*, vol. 4478, pp. 9–19, Dec. 2001.

[19] J.-L Starck and F. Murtagh, *Astronomical Image and Data Analysis*, 2nd ed. New York: Springer-Verlag, 2006.

[20] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, pp. 246–254, 1948.

[21] B. Z. M. Fadili and J.-L. Starck, "Multi-scale variance stabilizing transform for multi-dimensional poisson count image denoising," presented at the Int. Conf. Image Processing, Toulouse, France, May 14-19, 2006.

[22] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, to be published.

[23] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[24] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.

---

[1]More experiments with other images and sparse representations can be found at http://www.greyc.ensicaen.fr/~jfadili, http://jstarck.free.fr/mca.html or http://perso.orange.fr/jbobin/MCA2.html

Jérôme 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 the Université Paris XI, Orsay, France. He received the Agrégation de Physique in 2004. He is currently pursuing the Ph.D. degree at the CEA, France.

His research interests include statistics, information theory, multiscale methods and sparse representations in signal, and image processing.

Jean-Luc Starck received the Ph.D. degree from the University Nice-Sophia Antipolis, France, and the Habilitation degree from the University Paris XI, Paris, France.

He was a Visitor at the European Southern Observatory (ESO) in 1993 and at the Statistics Department, Stanford University, Stanford, CA, in 2000 and 2005, and at the University of California, Los Angeles, in 2004. He has been a Researcher at CEA, France, since 1994. 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 the author of two books entitled *Image Processing and Data Analysis: the Multiscale Approach* (Cambridge University Press, 1998) and *Astronomical Image and Data Analysis* (Springer, 2006, 2nd Ed.). His research interests include image processing, statistical methods in astrophysics, and cosmology. He is an expert in multiscale methods such as wavelets and curvelets.

Jalal M. Fadili graduated from the Ecole Nationale Supérieure d'Ingénieurs (ENSI) de Caen, Caen, France, and received the M.Sc. and Ph.D. degrees in signal and image processing from the University of Caen.

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 University of Technology, Brisbane, Australia, and Stanford University, Stanford, CA, 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.

Yassir Moudden graduated in electrical engineering from SUPELEC, Gif-sur-Yvette, France, and received the M.S. degree in physics from the Université de Paris VII, France, in 1997, and the Ph.D. degree in signal processing from the Université de Paris XI, Orsay, France.

He was a visitor at the University of California, Los Angeles, in 2004 and and is currently with the CEA, France, working on applications of signal processing to astronomy. His research interests include signal and image processing, data analysis, statistics, and information theory.

David L. Donoho received the A.B. degree (summa cum laude) in statistics from Princeton University, Princeton, NJ, and the Ph.D. degree in statistics from Harvard University, Cambridge, MA.

He is Professor of statistics at Stanford University, Stanford, CA. He was a Professor at the University of California, Berkeley, and a Visiting Professor at the Université de Paris, Paris, France, as well as a Sackfer Fellow at Tel Aviv University, Tel Aviv, Israel. His research interests are in harmonic analysis, image representation, and mathematical statistics.

Dr. Donoho is a member of the U.S. National Academy of Sciences and a Fellow of the American Academy of Arts and Sciences.