ENHANCED SOURCE SEPARATION BY MORPHOLOGICAL COMPONENT ANALYSIS

J. Bobin, Y. Moudden, J.-L. Starck

DAPNIA/SEDI-SAP, Service d'Astrophysique,
CEA/Saclay, 91191 Gif sur Yvette, France

ABSTRACT

This paper describes two extensions of the recent Morphological Component Analysis (MCA) method to multichannel data. MCA takes advantage of the sparse representation of structured data in large overcomplete dictionaries to separate features in the data based on their morphology. It was shown to be an efficient technique in such problems as separating an image into texture and piecewise smooth parts or for inpainting applications. A first extension, MMCA, achieves a similar source separation objective based on morphological diversity. A second extension, GMMCA, takes advantage of the highly sparse representations of the sources that can be built using MCA. Indeed, parsity is now generally recognized as a valuable property for blind source separation. The efficiency of MMCA and GMMCA is confirmed in numerical experiments.

1. INTRODUCTION

A common assumption in signal or image processing is that measurements $\mathbf{X}$ made typically using an array of sensors, often consists of mixtures of contributions from various possibly independent underlying physical processes $\mathbf{S}$. The simplest mixture model is linear and instantaneous and takes the form :

$$\mathbf{X} = \mathbf{A}\,\mathbf{S} \;+\; \mathbf{N} \tag{1}$$

where $\mathbf{X}$ and $\mathbf{S}$ are random vectors of respective sizes $m \times 1$ and $n \times 1$ and $\mathbf{A}$ is an $m \times n$ matrix. Multiplying $\mathbf{S}$ by $\mathbf{A}$ linearly mixes the n sources into m observed processes. In some cases, an $m \times 1$ random vector $\mathbf{N}$ is included to account for instrumental noise.The problem is then to invert the mixing process so as to separate the data back into its constitutive elementary building blocks leading to a more concise and possibly more interpretable representation of the data. In a blind approach assuming minimal prior knowledge of the mixing process, source separation is merely about devising quantitative measures of diversity or contrast. A first simple example is the separation of sources with disjoint supports in a given representation such as time and/or frequency. In a second example, the strict orthogonality of the supports can be relaxed provided the mixed source processes are now statistically independent. This is the framework of Independent Component Analysis (ICA), a growing set of multichannel data analysis techniques, which have proven successful in a wide range of applications [1]. Indeed, although statistical independence is a strong assumption, it is in many cases physically plausible.

ICA algorithms for blind component separation and mixing matrix estimation depend on the *a priori* model used for the probability distributions of the sources [2, 1]. In a first set of blind techniques, the components are modeled as Gaussian processes and, in a given representation (time, Fourier, wavelet, etc.), separation requires that the sources have diverse, *i.e.* non proportional, variance profiles. These methods generally lead to criteria expressing the joint diagonality of a set of matrix statistics which can be optimized efficiently [3, 4].

In a second set of techniques, source separation is achieved based on the non-Gaussianity of all but possibly one of the components. Most mainstream ICA techniques belong to this category: JADE, FastICA, Infomax (see [1] and references therein). An especially important case is when the mixed sources are highly sparse, meaning that each source is only rarely active and mostly nearly zero. The independence assumption then ensures that the probability for two sources to be significant simultaneously is extremely low so that the sources may again be treated as having nearly disjoint supports. This is exploited for instance in Sparse Component Analysis [5]. And it is shown in [6] that first moving the data into a representation in which the sources are assumed to be sparse will greatly enhance the quality of the separation. Possible dictionaries include Fourier and related bases, wavelet bases, etc. Working with combinations of several bases or with very redundant dictionaries such as undecimated wavelet frames or the more recent ridgelets, curvelets [7], etc. could lead to even more efficient representations. However, selecting from a large dictionary, the smallest subset of elements, that will linearly combine to reproduce a given signal or image, is a hard combinatorial problem. Nevertheless, several algorithms have been proposed that can help build very sparse decompositions [8, 9] and in fact, a number of recent results prove that these algorithms will recover the unique optimal decomposition provided this solution is sparse enough and the dictionnary is sufficiently incoherent [10, 11].

Morphological Component Analysis (MCA) is a method described in [12] that constructs a sparse representation of a signal or an image considering that it is a combination of features which are sparsely represented in different dictionaries. For instance, images commonly combine contours and textures : the former are well accounted for using *e.g.* curvelets while the latter may be well represented using local cosine functions. A brief account of MCA is given in section 2. The purpose of this contribution is to extend MCA to the case of multi-channel data. This is described in section 3. Then we show in section 4 how the sparse representation of data obtained using MCA can be used to enhance blind source separation.

2. MCA

In searching a sparse decomposition of a signal or image s, MCA makes the specific assumption that s is a sum of K components φ_k where a possibly overcomplete dictionary $\mathbf{\Phi}_k$ is given for each k, in which φ_k admits a sparse representation, $\varphi_k = \mathbf{\Phi}_k \alpha_k$ while its sparsest decomposition over the other $\mathbf{\Phi}_{k'\neq k}$ is essentially diffuse. The different $\mathbf{\Phi}_k$ can be seen as acting as discriminants between the different components of the initial signal s. Ideally, the α_k are the solutions of:

$$\min_{\{\alpha_1,\dots,\alpha_K\}} \sum_{k=1}^{K} \|\alpha_k\|_0 \quad \text{subject to} \quad s = \sum_{k=1}^{K} \mathbf{\Phi}_k \alpha_k. \tag{2}$$

However, the L_0 norm is non-convex and optimizing the above criterion is combinatorial by nature. Substituting an L_1 sparsity measure to the L_0 norm, as motivated by recent equivalence results *e.g.* in [10], and relaxing the equality constraint, the MCA algorithm seeks a solution to the following minimization problem:

$$\min_{\varphi_1,\dots,\varphi_K} \sum_{k=1}^{K} \lambda_k \|\alpha_k\|_1 + \|s - \sum_{k=1}^{K} \varphi_k\|_2^2 \ \text{ with } \ \varphi_k = \mathbf{\Phi}_k \alpha_k \tag{3}$$

with again $\varphi_k = \mathbf{\Phi}_k \alpha_k$. In the case where each $\mathbf{\Phi}_k$ is an orthonormal basis, the above is equivalent to the following set of coupled equations:

$$\forall k, \varphi_k = r_k - \frac{\lambda_k}{2} \mathbf{\Phi}_k \text{sign}(\mathbf{\Phi}_k^{-1} \varphi_k) \text{ with } r_k = s - \sum_{k' \neq k} \varphi_{k'} \tag{4}$$

This can be solved efficiently using the iterative *Block-Coordinate Relaxation Method* [13] in conjunction with, at a given k, a soft-thresholding of the decomposition of r_k over $\mathbf{\Phi_k}$. When non-unitary or redundant transforms are used, the above is no longer strictly valid. Nevertheless, simple shrinkage does give satisfactory results when practiced with non-unitary transforms and in fact this is rather well understood theoretically [14]. Finally, denoting by $\mathbf{T}_k$ and $\mathbf{R}_k$ the forward and inverse transforms associated with the redundant dictionary $\mathbf{\Phi_k}$, MCA finds a solution to problem (3) with the following algorithm:

1. Set # of iterations $L_{\max}$ & thresholds $\forall k, \delta_k = L_{\max} \cdot \lambda_k/2$
2. While $\delta_k > \lambda_k/2$,
   For $k = 1, \dots, K$:
      Update of φ_k assuming all $\varphi_{k'\neq k}$ are fixed:
      – Compute the residual $r_k = s - \sum_{k'\neq k} \varphi_{k'}$
      – Compute $\alpha_k = \mathbf{T}_k r_k$
      – Soft threshold α_k with threshold $= \delta_k$ gives $\hat{\alpha}_k$
      – Reconstruct φ_k by $s_k = \mathbf{R}_k \hat{\alpha}_k$
   Lower the thresholds: $\delta_k = \delta_k - \lambda_k/2$

In the above, soft thresholding results from the use of an L_1 sparsity measure, which as explained earlier comes as a good approximation to the desired L_0 norm. Towards the end of the iterative process, applying a hard threshold instead may lead to better results. The final threshold should vanish in the noise-less case. However, when equation 3 is interpreted in a probabilistic framework, then the last threshold depends on the noise variance and on the widths λ_k of the Laplacian priors set on the generative model of the data. A detailed description of MCA is given in [12] along with results of experiments in contour/texture separation and image inpainting.

3. MULTICHANNEL MCA

We consider the mixing model (1) and make the additional assumption that each source s_k is well (*i.e.* sparsely) represented in a specific dictionary as in section 2. Again, assigning a Laplacian prior with precision λ_k to the decomposition coefficients of the k^{th} source s_k in dictionary Φ_k is a practical way to implement this property. Here, s_k denotes the $1 \times n$ array of the k^{th} source samples. Classically, we assume Gaussian white noise with known covariance Γ_n. This leads to the following joint estimator of the source processes $\mathbf{S} = \{s_1, \dots, s_{n_s}\}$ and the mixing matrix $\mathbf{A}$:

$$\{\hat{\mathbf{S}}, \hat{\mathbf{A}}\} = \text{Arg} \min_{\mathbf{S},\mathbf{A}} \|\mathbf{X} - \mathbf{AS}\|_{2,\mathbf{\Gamma}_n}^2 + \sum_k \lambda_k \|s_k \mathbf{T}_k\|_1 \tag{5}$$

where $\|\mathbf{M}\|_{2,\mathbf{\Gamma}_n}^2 = \text{trace}\left(\mathbf{M}^T \mathbf{\Gamma}_n^{-1} \mathbf{M}\right)$. Unfortunately, this minimization problem suffers from the lack of scale invariance of the objective function. Indeed, combining a scaling of the mixing matrix, $A \leftarrow \rho A$, and an inverse scaling of the source matrix, $S \leftarrow \frac{1}{\rho} S$, leaves the quadratic measure of fit unchanged whereas the term measuring sparsity is deeply altered by the same inverse scale factor $\frac{1}{\rho}$. Consequently, the minimization will probably drive us to trivial solutions, $A \rightarrow \infty$ and $S \rightarrow 0$, since the sparsity term can be minimized *ad libitum* as ρ goes to $+\infty$. Nevertheless, scale-invariance can be artificially recovered by normalizing the columns a^k of the mixing matrix $\mathbf{A}$ at each iteration ($a^{k+} \leftarrow a^{k-}/\|a^{k-}\|_2$) and propagating the scale factor to the corresponding source, $s_k{}^+ \leftarrow \|a^{k-}\|_2 s_k{}^-$, and precision $\lambda_k{}^+ \leftarrow \|a^{k-}\|_2 \lambda_k{}^-$.

Define the k^{th} multichannel residual $\mathbf{D}_k = \mathbf{X} - \sum_{k'\neq k} a^{k'} s_{k'}$ as corresponding to the part of the data unexplained by the other couples $\{a^{k'}, s_{k'}\}_{k'\neq k}$. Then, the minimization problem (5) is equivalent to jointly minimizing the following set of elementary criteria :

$$\forall k, \{\hat{s}_k, \hat{a}^k\} = \text{Arg} \min_{s_k, a^k} \|\mathbf{D}_k - a^k s_k\|_{2,\mathbf{\Gamma}_n}^2 + \lambda_k \|s_k \mathbf{T}_k\|_1 \tag{6}$$

Zeroing the gradient with respect to s_k and a^k of this criterion leads to the following coupled equations:

$$\begin{cases} s_k & = \frac{1}{a^{k^T} \mathbf{\Gamma}_n^{-1} a^k} \left(a^{k^T} \mathbf{\Gamma}_n^{-1} \mathbf{D}_k - \frac{\lambda_k}{2} \text{Sign}(s_k \mathbf{T}_k) \mathbf{R}_k \right) \\ a^k & = \frac{1}{s_k s_k{}^T} \mathbf{D}_k s_k{}^T \end{cases} \tag{7}$$

Although the above holds for unitary transforms for which $\mathbf{R}_k = \mathbf{T}_k^T$, we make the same approximation as in the previous section and consider that it is still valid for redundant transforms. Then, for a fixed a^k, the source process s_k is estimated by soft-thresholding the coefficients of the decomposition of a *coarse version* $\widetilde{s_k} = (1/a^{k^T} \mathbf{\Gamma}_n^{-1} a^k) a^{k^T} \mathbf{\Gamma}_n^{-1} \mathbf{D}_k$ with threshold $\lambda_k/(2a^{k^T} \mathbf{\Gamma}_n^{-1} a^k)$. Considering a fixed s_k, the update on a^k follows from a simple least squares linear regression. The MMCA algorithm is given below :

1. Set # of iterations $L_{\max}$ & thresholds $\forall k, \delta_k = L_{\max} \cdot \lambda_k/2$
2. While $\delta_k > \lambda_k/2$,
   For $k = 1, \ldots, n_s$:
   - Renormalize a^k, s_k and δ_k
   - Update of s_k assuming all $s_{k' \neq k}$ and $a^{k'}$ are fixed:
     – Compute the residual $\mathbf{D}_k = \mathbf{X} - \sum_{k' \neq k} a^{k'} s_{k'}$
     – Project $\mathbf{D}_k$: $\tilde{s}_k = \frac{1}{{a^k}^T \mathbf{\Gamma}_n^{-1} a^k} {a^k}^T \mathbf{\Gamma}_n^{-1} D_k$
     – Compute $\alpha_k = \tilde{s}_k \mathbf{T}_k$
     – Soft threshold α_k with threshold $= \delta_k$ gives $\hat{\alpha}_k$
     – Reconstruct s_k by $s_k = \hat{\alpha}_k \mathbf{R}_k$
   - Update of a^k assuming all $s_{k'}$ and $a^{k' \neq k}$ are fixed:
     $a^k = \frac{1}{s_k {s_k}^T} \mathbf{D}_k {s_k}^T$

   Lower the thresholds: $\delta_k = \delta_k - \lambda_k/2$.

In comparison to the algorithm in [6] which uses a single sparsifying transform and a quadratic programming technique, our method considers more than just one transform and a shrinkage-based optimization. Figure 1 illustrates a simple experiment. The two source signals at the top left were linearly mixed to form the three synthetic observations shown at the top right. Some Gaussian noise was also added to the mixtures. The two sources are morphologically different: one consists of four bumps and the other is a plain sinewave. Source separation was conducted using the above MMCA algorithm and, for the sake of comparison, with the publicly available implementation of the JADE algorithm[1]. MMCA is

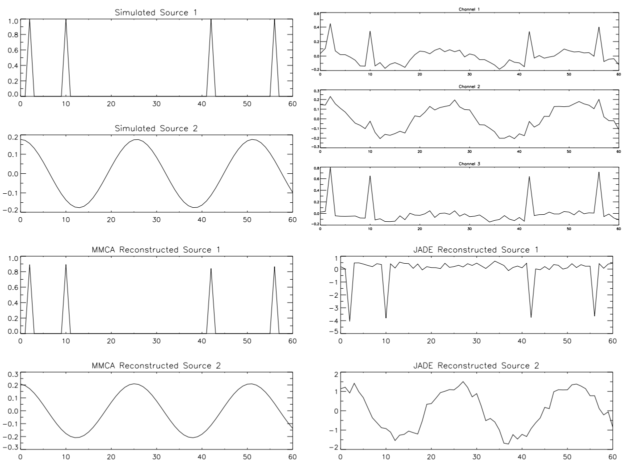

Fig. 1. **top left :** the two initial source signals. **top right :** three noisy *observed* mixtures. **bottom left :** the two source signals reconstructed using MMCA. **bottom right :** the two source signals reconstructed with Jade.

clearly able to effciently separate the intial source signals. Note that denoising is an intrinsic part of the algorithm as mentioned in section 2. However, the morphological diversity of the source processes is a condition for the good performance of MMCA. In the next section, we describe an extension to the case of similarly structured sources in which we exploit the sparse representations obtained with MCA to enhance blind source separation.

[1]http://www.tsi.enst.fr/ cardoso/guidesepsou.html

4. GENERALIZED MMCA

We assume now that the sources are statistically independent and that each source is a linear combination of morphologically different components as follows :

$$\forall k \in \{1, \ldots, n_s\}, \quad s_k = \sum_{i=1}^{n_k} \varphi_{k,i} = \sum_{i=1}^{n_k} \alpha_{k,i} \mathbf{R}_i \tag{8}$$

where each component $\varphi_{k,i}$ is well sparsified by only one transform $\mathbf{T}_k$. Estimating all the parameters of the GMMCA model leads to the following minimization problem :

$$\min_{A, \ldots, \varphi_{k,i}, \ldots} \left\{ \|\mathbf{X} - \mathbf{AS}\|^2_{2,\Gamma_n} + \sum_{k=1}^{n_s} \sum_{i=1}^{n_k} \lambda_k \|\varphi_{k,i} \mathbf{T}_i\|_1 \right\} \tag{9}$$

At any point in the minimization, the MCA algorithm can be used to estimate the underlying components $\{\hat{\varphi}_{k,i}\}_{i=1,\ldots,n_k}$ in the current estimate of the k^{th} source $\tilde{s}_k$. This builds very sparse representations of the different sources which will then have mostly disjoint supports: as mentioned earlier the probability for sparse independent sources to be simultaneously active is very low. This contrast is the key to a successful source separation as it enforces the necessary diversity to tell the sources apart. The estimation of the mixing matrix A should benefit from this increased diversity. Following similar derivations as for the minimization of (3) and (5), we propose that (9) can be solved using the following GMMCA algorithm:

1. Set # of iterations $L_{\max}$ & thresholds $\forall k, \delta_k = L_{\max} \cdot \lambda_k/2$
2. While $\delta_k > \lambda_k/2$:
   For $k = 1, \ldots, n_s$:
   3. Renormalize $\forall k$, a^k, s_k and δ_k as in MMCA.
   4. Update of s_k assuming all $s_{k' \neq k}$ and $a^{k'}$ are fixed:
      – Compute coarse *current* estimate $\tilde{s}_k$ of source s_k.
      – Perform an MCA decomposition of $\tilde{s}_k$ with final threshold δ_k and get coefficients $\{\alpha_{k,i}\}$ of its sparse representation.
   5. Update of $\mathbf{A}$ assuming the sources are fixed.
   6. Lower the thresholds, $\forall k$, $\delta_k = \delta_k - \lambda_k/2$.

At step four, a coarse version $\tilde{s}_k$ of s_k is computed. In the case where there are more sensors than sources to estimate, the current estimate of the mixing matrix generally admits a pseudo-inverse $\bar{\mathbf{A}}^{-1}$. We may then take $\tilde{s}_k = (\bar{\mathbf{A}}^{-1}\mathbf{X})_k$. In the under-determined case, one may take $\tilde{s}_k = 1/({a^k}^T \mathbf{\Gamma}_n^{-1} a^k) {a^k}^T \mathbf{\Gamma}_n^{-1} \mathbf{D}_k$ with $\mathbf{D}_k = \mathbf{X} - \sum_{k' \neq k} a^{k'} s_{k'}$ as in MMCA. We then build sparse representations of the $\tilde{s}_k$ using the MCA algorithm with final threshold $\delta_k / {a^k}^T \mathbf{\Gamma}_n^{-1} a^k$. Step five of the GMMCA algorithm consists in estimating the mixing matrix assuming the sources are fixed. Based on the most active atoms in the sparse representations $\alpha = \{\alpha_{k,i}\}$ of each of the current estimates of the source processes resulting from the MCA decomposition, the current estimate of the mixing matrix can be refined using the following update rule :

$$\hat{\mathbf{A}} \leftarrow \hat{\mathbf{A}} + \mu \hat{\mathbf{A}} (\mathbf{I} - \beta \Upsilon(\alpha) \alpha^T) \tag{10}$$

where β is a normalizing constant and $\mathbf{I}$ is the identity matrix in $\mathbf{R}^m$. This update is simply a descent along the rescaled gradient with respect to $\mathbf{A}$, of an approximation to the log likelihood as derived in [15]. Therefore, Υ is a vector of non-linear *score functions* related to the assumed probability distributions of the source coefficients. A line-search algorithm is used to find the optimal μ at each step. In the noise-less over-determined case, the minimization problem (9) could be rewritten in terms of the inverse of $\mathbf{A}$. The learning rule on the de-mixing matrix would then be what one typically encounters in noise-less ICA algorithms [1].

The GMMCA algorithm was applied on synthetic data. The top of Figure 2 shows 128 samples of two synthetic source signals with zero mean and unit variance. Both sources consist of bumps and sines. These were linearly mixed and white Gaussian noise of variance $\sigma = 0.01$ was added, resulting in the two *observed* signals shown on the middle row of Figure 2. The over-complete dictionary $\mathbf{T}$ we used was obtained as the concatenation of the Dirac basis and the DCT basis. The bottom row of figure 2 shows

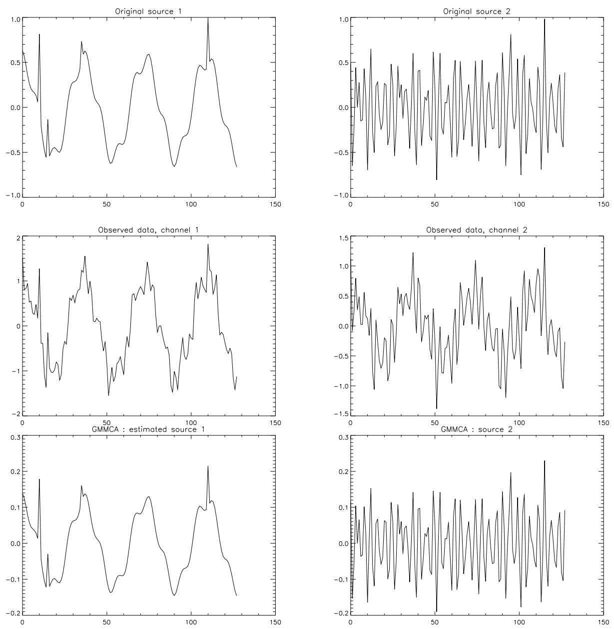

Fig. 2. **top row :** the two initial source signals. **middle row:** two noisy *observed* mixtures. **bottom row :** the two source signals reconstructed using GMMCA.

the two source signals estimated using GMMCA. Although more work is necessary to fully assess the performances of GMMCA, these preliminary results illustrate the efficiency of the proposed method for source separation. More results and applications to 2D data will be presented at the conference.

5. CONCLUSION

The MCA algorithme provides a powerful and fast sparse signal decomposition in a redundant dictionary. The MMCA algorithm described in this paper extends MCA to the multichannel case. For blind source separation, this first extension is shown to perform well provided the original sources are morphologically different meaning that the sources are sparsely represented in different bases. We also introduced a more general model, GMMCA, which assumes that the sources admit a sparse decomposition in a given over-complete dictionary $\mathbf{T}$ to be recovered using MCA. GMMCA is shown to enhance source separation by exploiting the sparsity of the sources.

6. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, New York, 2001, 481+xxii pages.

[2] Jean-François Cardoso, “The three easy routes to independent component analysis; contrasts and geometry,” in *Proc. ICA 2001, San Diego*, 2001.

[3] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique based on second order statistics,” *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[4] Dinh-Tuan Pham and Jean-François Cardoso, “Blind separation of instantaneous mixtures of non stationary sources,” *IEEE Trans. on Sig. Proc.*, vol. 49, no. 9, pp. 1837–1848, Sept. 2001.

[5] P. G. Georgiev, F. Theis, and A. Cichocki, “Sparse component analysis and blind source separation of underdetermined mixtures,” *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, 2005.

[6] M. Zibulevsky and B.A. Pearlmutter, “Blind source separation by sparse decomposition in a signal dictionary,” *Neural-Computation*, vol. 13, no. 4, pp. 863–882, April 2001.

[7] J.-L. Starck, E. Candès, and D.L. Donoho, “The curvelet transform for image denoising,” *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 131–141, 2002.

[8] S.G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[9] S.S. Chen, D.L. Donoho, and M.A. Saunder, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[10] D. L. Donoho and M. Elad, “Maximal sparsity representation via l_1 minimization,” *the Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, 2003.

[11] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[12] J.-L. Starck, M. Elad, and D.L. Donoho, “Redundant multiscale transforms and their application for morphological component analysis,” *Advances in Imaging and Electron Physics*, vol. 132, 2004.

[13] A.G. Bruce, S. Sardy, and P. Tseng, “Block coordinate relaxation methods for nonparametric signal de-noising,” *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 3391, pp. 75–86, 1998.

[14] M. Elad, “Why simple shrinkage is still relevant for redundant representations?,” Submitted to the IEEE Trans. On Information Theory on January 2005.

[15] M. S. Lewicki and T. J. Sejnowski, “Learning overcomplete representations,” *Neural Computation*, vol. 12, no. 2, pp. 337–365, 2000.