SCHOOL OF ENGINEERING - STI
SIGNAL PROCESSING INSTITUTE
Lorenzo Peotta, Philippe Jost, Pierre Vandergheynst and Pascal Frossard

CH-1015 LAUSANNE
Telephone: +4121 6932601
Telefax: +4121 6937600
e-mail: lorenzo.peotta@epfl.ch

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

SPARSE APPROXIMATION WITH BLOCK INCOHERENT DICTIONARIES

Lorenzo Peotta, Philippe Jost, Pierre Vandergheynst and Pascal Frossard
Swiss Federal Institute of Technology Lausanne (EPFL)

Signal Processing Institute Technical Report

TR-ITS-2003.007

December 1st, 2003

Submitted to IEEE ISIT 2004

Sparse Approximation with Block Incoherent Dictionaries

Lorenzo Peotta, Philippe Jost, Pierre Vandergheynst and Pascal Frossard
Signal Processing Institute
Swiss Federal Institute of Technology
Lausanne, Switzerland
{lorenzo.peotta,philippe.jost,pierre.vandergheynst,pascal.frossard}@epfl.ch

I. Introduction

Building good sparse approximations of functions is one of the major themes in approximation theory. When applied to signals, images or any kind of data, it allows to deal with basic building blocks that essentially synthesize all the information at hand. It is known since the early successes of wavelet analysis that sparse expansions very often result in efficient algorithms for characterizing signals in noise or even for analyzing and compressing signals. The very strong links between approximation theory and computational harmonic analysis on one hand and data processing on the other hand, resulted in fruitful cross-fertilizations over the last decade, from fundamental results (near optimal rate of non-linear approximations for wavelets and other basis [1]) to practical ones (like the JPEG2000 image compression standard).

Natural signals however do not generally lend themselves to simple models, for which orthonormal basis are generally near optimal. Images for example do contain smooth parts and regular contours that could be efficiently represented by a curvelet tight frame [2], but they also contain various kind of irregular edges together with a plethora of textures. Audio signals contain sharp transients and smooth parts that are suitable for wavelet basis, but they also contain stationary oscillatory parts that are better suited for local trigonometric basis [3]. Bearing in mind the multiple components of natural data, one is tempted to approximate them with mixtures of basis functions. Approximating data with general dictionaries seemed a daunting task, and raised many questions concerning the unicity and optimality of sparse representations. Fortunately there have been recently an intense activity in this field, showing that constructive results can be obtained on all fronts. The possibility of recovering optimal sparse representations using Basis Pursuit (BP) opened the way [4]–[7]. When an exact sparse representation is not needed, approximation results become more useful, and recent results have shown that variations around greedy algorithms such as Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP) are promising [8], [9].

One of the key properties in the above-mentioned results lies in the characteristics of the dictionary, and one could roughly say that in most cases the latter is required to be sufficiently incoherent, i.e *close enough* to an orthogonal basis. Putting strong restrictions on the dictionary though may damage the original goal in the sense that we loose flexibility in designing it. In this paper, we basically relax some of these strong hypotheses by allowing more redundancy in the dictionaries, through the concept of block incoherence, which basically describes a dictionary that can be represented as the union of incoherent blocks. We show that even pure greedy algorithms can strongly benefit from such design by proving a recovery condition under which Matching Pursuit will always pick up correct atoms during the signal expansion. Based on this result, we design an algorithm that constructs a near block incoherent dictionary starting from any initial dictionary. A tree structured greedy algorithm is then proposed as a way of constructing sparse approximations with block incoherent dictionaries. This algorithm presents the important advantage of being much faster than a classical Matching Pursuit. In the same time, it only minimally degrades the quality of approximation thanks to the recovery condition, derived for block incoherent dictionaries. The performance of the proposed algorithm are demonstrated in the context of image representation.

This paper is organized as follows. Sec. II first proposes definitions on coherence between generic subsets of basis functions. It then proposes theorems which show that Matching Pursuit picks the correct atoms during signal expansion, provided that the dictionary is block incoherent. Sec. III presents a generic method to build block incoherent dictionaries from any set of functions. It finally shows the benefits of the recovery condition in the context of image representation using block incoherent dictionary.

II. Block Sparse Approximation

A. Preliminary Definitions

This section proposes a set of novel definitions for coherence between generic subsets of functions, that will be used in the remaining of the paper. Given a redundant dictionary $\mathcal{D}$, we consider the following L-subset decomposition $\mathcal{D} = \bigcup_{l=1}^{L} B_l$, and we call *blocks* the L subsets of atoms B_l, $l = 1, \cdots, L$. The *block coherence* is defined as the maximum coherence between any two atoms, taken from different blocks.

Definition 1: The block coherence μ_B, given a block decomposition $\mathcal{D} = \bigcup_{l=1}^{L} B_l$, is

$$\mu_B \triangleq \max_{i \neq j} \max_{k,l} | < g_k^i, g_l^j > |, \tag{1}$$

where g_k^i is the k^{th} atom from the block B_i.

Definition 2: A dictionary is said *block incoherent* if there exists a decomposition such that the *block coherence* μ_B is small. The block coherence considers similarities between atoms from two different blocks. In order to refine the analysis of the coherence, we introduce another function, called the $Babel\ block$ function, that represents the coherence between sets of m blocks $B_I = \bigcup_{i \in I} B_i$, with $Card(I) = m$.

Definition 3: Let $\mathcal{D} = \bigcup_{l=1}^{L} B_l$ denote a decomposition, and $B_I = \bigcup_{i \in I} B_i$ represent a set of m blocks. The $Babel\ block$ function is

$$\mu_{1_B}(m) \triangleq \max_{I, s.t. |I| = m} \max_{j \notin I, l} \sum_{i \in I} \max_k | < g_k^i, g_l^j > |. \tag{2}$$

Definition 4: A given dictionary $\mathcal{D}$ is said to be *block quasi-incoherent*, if we can find a block decomposition such that $\mu_{1_B}(m)$ grows slowly with m.

The *block coherence* μ_B considers coherence between two blocks, and the *Babel block* function $\mu_{1_B}(m)$ measures coherence between m blocks. Notice that the $Babel\ block$ function is bounded by the $block\ coherence$: $\mu_{1_B}(m) \leq m\mu_B$. The definitions of the previous functions is the extension of the coherence μ and the $Babel$ function $\mu_1(m)$ introduced by Donoho, Huo, Elad and Tropp [5], [6], [9]. We need now also to consider the coherence within a single block. Generally, a single block B_i has a strong coherence (i.e., the *Babel function* grows quickly). For a more detailed analysis, we are however interested in a function that represents the coherence of a particular subset of functions in B_i, and we call it the $Bore$ function $\xi(B_i)$.

Definition 5: The $Bore$ function related to a block B_i is

$$\xi(B_i) \triangleq \min_{\iota, s.t. |B_\iota| = rank(B_i)} \max_k \sum_{l \neq k} | < g_l^i, g_k^i > |, \tag{3}$$

where B_ι is a set of independent atoms from B_i, i.e., $B_\iota \subset B_i$ and $span(B_\iota) = span(B_i)$.

The $Bore$ function $\xi(B_i)$ indicates how close a basis constructed with atoms from block B_i is to an orthogonal basis that spans the range $\mathcal{R}(B_i)$. The set of atoms, i.e. B_ι, where the Bore function is minimal, is called $B_{i^\star}$. If $\xi(B_i) = 0$, we can find a set $B_{i^\star} \subset B_i$ that is an orthogonal basis for $span(B_i)$. The extension of the $Bore$ function to the dictionary $\mathcal{D}$ is finally defined as $\xi(\mathcal{D}) = \max_i \xi(B_i)$.

B. Exact Block Selection

Using the definitions defined in Sec. II-A, we prove in this section that, given a block incoherent dictionary $\mathcal{D}$ and a signal f, the Matching Pursuit (MP) algorithm can recover a block-sparse representation of f. We consider here the restricted problem $(\mathcal{D}, B_I)$-*SPARSE*, which means that f is a linear combination of atoms belonging to a subset of m blocks, $B_I = \bigcup_{i \in I} B_i$.

Firstly, we find a single sufficient condition under which Matching Pursuit recovers atoms from a given set of incoherent blocks B_I. In this case, we say that MP chooses atoms from *correct* blocks B_i $i \in I$. Let represent B_I as an operator or matrix, and let B_I^+ denote its pseudoinverse.

Theorem 1: Let $\mathcal{D}$ a block incoherent dictionary and $B_I = \bigcup_{i \in I} B_i$. If the signal $f \in \mathcal{V}_I = span(B_I)$, under the recovery condition

$$\eta(B_I) \triangleq \max_{g \notin B_I} \|B_I^+ g\|_1 < 1 \tag{4}$$

then we have that MP:

1) picks up atoms only from correct blocks B_i $i \in I$,
2) converges exponentially to f.

Proof of Theorem 1. We follow the proof for "Exact Recovery" theorem [9], [10]. Suppose that $r_{n-1} \in \mathcal{V}_I$. If an atom g_{n-1} from B_I is selected by the Matching Pursuit algorithm, then $r_n = r_{n-1} - < g_{n-1} r_{n-1} > g_{n-1}$ belongs to $\mathcal{V}_I$, with $r_0 = f$. The vector $B_I r_{n-1}$ lists the inner products between the residual r_n and all the atoms from the blocks B_i, $i \in I$; taking the ∞ norm of this vector we have that $\|B_I^T r_{n-1}\|_\infty$ is the largest of these inner products in magnitude, where B_I^T represents the complex conjugate of B_I. The number $\|B_{\bar{I}}^T r_{n-1}\|_\infty$ corresponds to the largest inner product in magnitude between r_{n-1} and an atom that does not belong to B_I, that means $g \in B_{\bar{I}}$. An atom is selected from the correct block B_i, $i \in I$, when the following quotient is less than one

$$\rho(r_{n-1}) \triangleq \frac{\|B_{\bar{I}}^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} < 1 \tag{5}$$

By assumption, $r_{n-1} \in \mathcal{V}_I$, thus $r_{n-1} = B_I c_{n-1}$, where c_{n-1} is a vector of coefficients, and using the pseudoinverse we have $c_{n-1} = B_I^+ r_{n-1}$. Therefore, using the properties of the pseudoinverse, $r_{n-1} = (B_I^+)^T B_I^T r_{n-1}$ and

$$\begin{aligned}\rho(r_{n-1}) &= \frac{\|B_{\bar{I}}^T (B_I^+)^T B_I^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} \\ &\leq \|B_{\bar{I}}^T (B_I^+)^T\|_{\infty,\infty}\end{aligned}$$

where the matrix norm $\|\cdot\|_{p,p}$ is the norm "induced" by the vector norm $\|\cdot\|_p$. Using properties of the matrix norm we obtain

$$\begin{aligned}\rho(r_{n-1}) &\leq \|B_{\bar{I}}^T (B_I^+)^T\|_{\infty,\infty} \\ &= \|B_I^+ B_{\bar{I}}\|_{1,1} \\ &= \max_{g \in B_{\bar{I}}} \|B_I^+ g\|_1,\end{aligned}$$

so $\rho(r_{n-1}) \leq \eta(B_I) < 1$ which means that MP selects an atom from B_I. By induction the first part is proved.

To prove the second part, we just notice that MP is faced with a finite dimentional space $\mathcal{V}_I$, and we know that MP in a finite dimentional space is exponentially convergent.

□

As a corollary, the following theorem gives a condition under which right block selection is in force when f belongs to the $span$ of an arbitrary set of m incoherent blocks.

Theorem 2: Let $\mathcal{D}$ a block incoherent dictionary and B_I an arbitrary set of m blocks and $R = \max_i rank(B_i)$. If the signal $f \in \mathcal{V}_I$ and

$$R\,\mu_{1_B}(m) + \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) < 1 \qquad (6)$$

then we have that MP:

1) picks up atoms only from the correct blocks,
2) converges exponentially to f.

Proof of Theorem 2. The proof is again given by induction. We suppose that $r_{n-1} \in \mathcal{V}_I$. If an atom from B_I is selected, then $r_n \in \mathcal{V}_I$. We indicate with $B_I' = \bigcup_{i \in I} B_{i^\star}$ the union of the m sets associated to the m blocks B_i in the definition (5) of the $Bore$ function. Now we define $B_{I^\star}$ to be a set of linear independent atoms from B_I' such that $|B_{I^\star}| = rank(B_I')$. It follows that $span(B_{I^\star}) = span(B_I) = \mathcal{V}_I$, $B_{I^\star}$ is a basis for $\mathcal{V}_I$, therefore $r_{n-1} = (B_{I^\star}^+)^T B_{I^\star}^T r_{n-1}$ and

$$\begin{aligned}\rho(r_{n-1}) &= \frac{\|B_{\bar{I}}^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty} \\ &= \frac{\|B_{\bar{I}}^T (B_{I^\star}^+)^T B_{I^\star}^T r_{n-1}\|_\infty}{\|B_I^T r_{n-1}\|_\infty}\end{aligned}$$

since $B_{I^\star} \subset B_I$ we have that $\|B_I^T r_{n-1}\|_\infty \geq \|B_{I^\star}^T r_{n-1}\|_\infty$ and

$$\begin{aligned}\rho(r_{n-1}) &\leq \frac{\|B_{\bar{I}}^T (B_{I^\star}^+)^T B_{I^\star}^T r_{n-1}\|_\infty}{\|B_{I^\star}^T r_{n-1}\|_\infty} \\ &\leq \|B_{\bar{I}}^T (B_{I^\star}^+)^T\|_{\infty,\infty} \\ &= \max_{g \in B_{\bar{I}}} \|B_{I^\star}^+ g\|_1.\end{aligned}$$

Now we can expand the pseudoinverse and apply the norm bound $\|Ax\|_1 \leq \|A\|_{1,1}\|x\|_1$

$$\begin{aligned}\rho(r_{n-1}) &\leq \max_{g \in B_{\bar{I}}} \|(B_{I^\star}^T B_{I^\star})^{-1} B_{I^\star}^T g\|_1 \\ &\leq \|(B_{I^\star}^T B_{I^\star})^{-1}\|_{1,1} \max_{g \in B_{\bar{I}}} \|B_{I^\star}^T g\|_1. \qquad (7)\end{aligned}$$

We can easily bound the second term of the right part of (7) using the $Babel\ block$ function

$$\begin{aligned}\max_{g \in B_{\bar{I}}} \|B_{I^\star}^T g\|_1 &= \max_{g \in B_{\bar{I}}} \sum_{\psi \in B_{I^\star}} |<\psi, g>| \\ &\leq R\,\mu_{1_B}(m) \qquad (8)\end{aligned}$$

where $R = \max_i rank(B_i)$. In order to bound the first term of the right part of (7) we use the Von Neumann series to compute the inverse $(B_{I^\star}^T B_{I^\star})^{-1}$. Writing $B_{I^\star}^T B_{I^\star} = \mathcal{I} + A$, where $\mathcal{I}$ is the identity matrix, and under the condition that $\|A\|_{1,1} < 1$, it follows that :

$$\begin{aligned}\|(B_{I^\star}^T B_{I^\star})^{-1}\|_{1,1} &= \|(\mathcal{I}+A)^{-1}\|_{1,1} = \|\sum_{k=0}^{\infty}(-A)^k\|_{1,1} \\ &\leq \sum_{k=0}^{\infty}\|A\|_{1,1}^k = \frac{1}{1-\|A\|_{1,1}}.\end{aligned}$$

The matrix A has zero diagonal and the values out of diagonal correspond to the inner product between atoms from $B_{I^\star}$, taking into account the structure of $B_{I^\star}$ (it is composed by m incoherent blocks) we can bound the norm using the $Bore$ and $Babel\ block$ function:

$$\begin{aligned}\|A\|_{1,1} &= \max_k \sum_{j\neq k} |< g_j^{I^\star}, g_k^{I^\star} >| \\ &\leq \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) \end{aligned} \tag{9}$$

and putting together the bounds (8),(9) into (7) we obtain for $\rho(r_{n-1})$ the bound

$$\rho(r_{n-1}) \leq \frac{R\,\mu_{1_B}(m)}{1-(\xi(\mathcal{D}) + R\,\mu_{1_B}(m-1))}.$$

So under the condition

$$R\,\mu_{1_B}(m) + \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) < 1$$

it follows that $\rho(r_{n-1}) < 1$ and MP selects an atom from the correct block B_i, by induction the first part is proved. For the second part, we are in the same condition as in theorem 1.

□

Using the bound for the $Babel\ block$ function $\mu_{1_B}(m-1) \leq \mu_{1_B}(m) \leq m\,\mu_B$ it follows from Theorem 2 that, if the signal f belongs to the $span$ of m blocks, then MP recovers atoms from the correct blocks when

$$m < \frac{1-\xi(\mathcal{D})}{R\,\mu_B} \ .$$

Another important factor that determines the quality of a signal expansion is the rate of convergence of the approximation. This rate can also be bounded with the help of the coherence defined previously, in the case of block incoherent dictionaries. We briefly state now two new theorems that estimates the energy decay in the residual component of a Matching Pursuit expansion. The proofs are omitted here because of space constraints, but the interested readers are referred to [10] for details.

Theorem 3: If the signal $f \in \mathcal{V}_I$ and $R\,\mu_{1_B}(m) + \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) < 1$, then MP picks up atoms only from the correct blocks at each step and

$$\|r_n\|_2^2 \leq \|f\|_2^2 \left(1 - \frac{1-\xi(\mathcal{D}) - R\,\mu_{1_B}(m-1)}{R\,m}\right)^n . \tag{10}$$

Theorem 4: If the signal $f \in \mathcal{V}_I$ and $R\,\mu_{1_B}(m) + \xi(\mathcal{D}) + R\,\mu_{1_B}(m-1) < 1$, then MP picks up atoms only from the correct blocks at each step and

$$\|r_n\|_2^2 \leq \|f\|_2^2 \left(1 - \frac{\beta^2}{m}\right)^n , \tag{11}$$

where $\beta = \min_i \beta_i$, and β_i is related to the redundancy and structure of block B_i.

Since the dimension of the vector spaces generated by $span(B_i)$ is supposed to be small, we expect β to be close to one. The term m that divides β^2 could be substituted, taking into account the block incoherent structure of the dictionary [10]. If we have β^2 close to one, and m replaced by $h(m) \ll m$, we thus prove the good approximation behavior of Matching Pursuit for structured signals, that we observe on experimental results.

III. Application: Sparse Image Representation

A. Generation of Near Block Incoherent Dictionaries

This section presents a method to generate near block incoherent dictionaries, from arbitrary ones. The algorithm groups atoms into clusters and creates a representative atom for each cluster. A tree representation of the dictionary $\mathcal{D}$ allows for a fast implementation of the Matching Pursuit expansion. The elements from the initial dictionary $\mathcal{D}$ form the leaves of the tree. The node $N_{l,n}$, at the l^{th} level and n^{th} position of the tree holds a subspace of $\mathcal{D}$, which is as orthogonal as possible to its siblings. Each node has M children and is fully characterized by $L_{l,n}$, the list of the atom indexes from $\mathcal{D}$ contained in the subtree spanned by $N_{l,n}$. $N_{l,n}$ is a leaf node if $L_{l,n}$ contains only one element. Each node $N_{l,n}$ is also assigned a centroid $c_{l,n}$, that represents the atoms from the initial dictionary $\mathcal{D}$ contained in the corresponding subtree:

$$c_{l,n} = \frac{\sum_{k \in L_{l,n}} g_k}{\sqrt{\| \sum_{k \in L_{l,n}} g_k \|}} \ , \tag{12}$$

where the bi-dimensional function $g_k(x, y)$ denotes a generic atom in $\mathcal{D}$. The distance $d(g_i, g_j)$ between two atoms g_i and g_j is defined as $d(g_i, g_j) = |\langle g_i, g_j \rangle|$.

Let the mean distance between $c_{l,n}$ and its assigned atoms be written as $D_{l,n} = 1/n_{l,n} \sum_{i \in L_{l,n}} d(g_i, c_{l,n})$ where $n_{l,n}$ is the cardinality of $L_{l,n}$. For a given set $L_{l,n}$ of atoms, the quality $Q_{L_{l,n}}$ of a clustering is then defined as:

$$Q_{L_{l,n}} = \frac{1}{M} \sum_{w=0}^{M-1} D_{l+1, nM+w} \tag{13}$$

We choose to use the *k-means* algorithm [11] to build the tree. The *k-means* will try to maximize $Q_{L_{l,n}}$. The computation is over when the gain in term of $Q_{L_{l,n}}$ is less than a fixed ϵ, where ϵ can be made arbitrarily small. The generation of the tree is relatively computationally complex, but since it only depends on the dictionary, it can be done once and stored for multiple usage. As an example, Figure 1 represents a small part of a tree with $M = 4$. It shows how leaf atoms are groups together into parents which becomes more and more orthogonal to their sibling as the level in the tree augments. The clustering algorithm and the tree construction partition any initial dictionary $\mathcal{D}$ into near incoherent blocks. They are related to the *new paradigm* for dictionary design described in [12], that consists in creating structures from an arbitrary redundant dictionary in order to achieve low computational complexity.

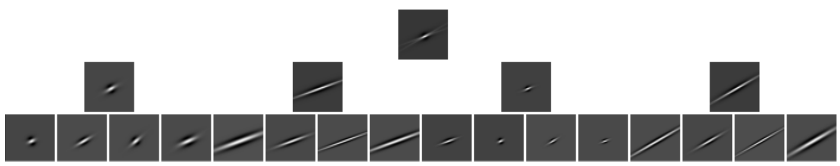

Fig. 1. Sample part of a tree.

B. Sparse Image Representation

Algorithm 1 Tree-based subspace pursuit

```
l=0, n=0 so that N_{l,n} is set to root node
repeat
   if N_{l,n} is root node then
      use fullsearch primitive to get best child w and best position
      l = l + 1 and n = nM + w
   else
      use localsearch primitive to get best child c and best position knowing last optimal position.
      l = l + 1 and n = nM + w
   end if
until N_{l,n} is a leaf
```

Based on the tree representation described above, we now propose a greedy algorithm that, at each stage, finds the best path through the tree down to the best leaf nodes (i.e., the atoms from $\mathcal{D}$). This provides a very fast alternative to the original Matching Pursuit method, as described in Algorithm 1.

Let $R^N f$ denote the residual of the signal f after the N^{th} call to the search algorithm. A primitive called *fullsearch* performs a full search over $R^N f$ for a set of m atoms, which represent the centroids of the M children of node. It outputs an atom and its respective position which best matches the residual signal. Another primitive called *localsearch* takes a list of atoms and an initial position, and performs a full search over a window of size $A \times A$ around the initial position (typically $A = 3$). It returns the atom and its corresponding position that best matches the residual signal. The computational complexity of this modified pursuit algorithm is clearly much lower than a complete full search method. For example, the first step already eliminates $\frac{M-1}{M}$ of the dictionary and additionally gives accurate spatial information on the position of the atom.

Figure 2 shows the performance of the modified tree-based pursuit algorithm, and the gain in computational complexity, as functions of M, the number of children per node of the tree. These results have been obtained with the dictionary described in [13], and a 128×128 *Lena* test image. The performance of the proposed method are compared to those of a full search that computes all the possible scalar products by multiplication in the Fourier domain.

Most of the complexity of the search procedure remains in the first M full searches to execute. The complexity therefore increases close to linearly with M (see Figure 2 (b)). A speed-up factor of about 150 is reached, compared to the full search method. On the other hand, if M is too small, the search becomes suboptimal because the block incoherence of the dictionary is not sufficient. It can be seen in Figure 2 (a) that the reconstruction error rapidly decreases a M grows. The total error is close to the reference full-search method, even if the blocks of the dictionary are not exactly orthogonal, which demonstrates the potential of block incoherent dictionaries. Note that when M is equal to the number of atoms in $\mathcal{D}$, or if the dictionary is built on orthogonal blocks, both methods perform identically. However, we cannot give any detailed conclusions on the performance of the rate of approximation of both methods; we can only provide bounds on the rate of convergence, because of the greediness of the algorithms.

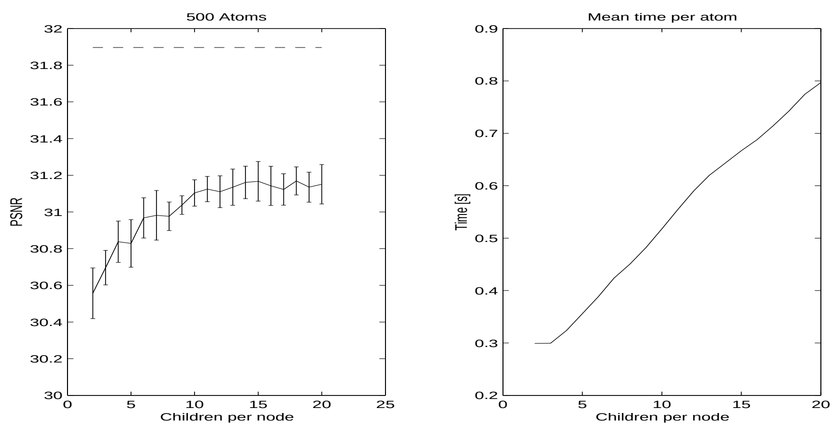

Fig. 2. (a) Reconstruction error as a function of the number of nodes M, for 500 atoms. The dashed line shows the results of a full search algorithm. (b) Mean time to find an atom as a function of the number of nodes M.

IV. Conclusions

This paper presented the potential of block incoherent dictionaries as efficient methods to limit the complexity of overcomplete signal expansions. It is shown to decrease the complexity of the signal decomposition with respect to Matching Pursuit, while the quality of the resulting approximation is kept quite satisfactory, depending however on the block incoherence. The future work will investigate more efficient clustering algorithms, taking into account inter-cluster distance, since the *k-means* clustering algorithm only considers intra-cluster distances. The design of block incoherent dictionaries from the beginning, and tighter bounds on approximation rate will also be investigated.

References

[1] DeVore R. A., “Nonlinear approximation,” *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[2] Candès, E. J. and Donoho, D. L., “Curvelets - a surprisingly effective non-adaptive representation for objects with edges.” *Curves and Surfaces, L. L. S. et al., ed., Nashville, TN, (Vanderbilt University Press)*, pp. 123–143, 1999.

[3] Daudet L. and Torrésani B., “Hybrid representations for audiophonic signal encoding,” *Signal Processing*, vol. 82, no. 11, pp. 1595–1617, November 2002.

[4] Chen, S. and Donoho, D., “Atomic Decomposition by Basis Pursuit,” in *SPIE International Conference on Wavelets*, Sant Diego, July 1995.

[5] Elad M. and Bruckstein A. M., “A generalized uncertainty principles and sparse representation in pairs of bases,” *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep 2002.

[6] Donoho D.L. and Huo, X., “Uncertainty principles and ideal atom decomposition,” *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[7] Gribonval R. and Nielsen M., “Sparse representations in unions of bases,” IRISA, Rennes (France), Tech. Rep. 1499, 2003.

[8] Gilbert A. C., Muthukrishnan S. and Strauss M. J., “Approximation of functions over redundant dictionaries using coherence,” in *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[9] Tropp, J., “Greed is good : Algorithmic results for sparse approximation,” Texas Institute for Computational Engineering and Sciences, Tech. Rep., 2003.

[10] Gribonval, R. and Vandergheynst, P. and Peotta, L., “Mp in block quasi-incoherent dictionaries,” Swiss Federal Institute of Technology, Tech. Rep., 2003.

[11] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proc. Fifth Berkley Symp. Math. Statistics and Probabiliy*, vol. 1, 1967, pp. 281–296.

[12] Neff, R. and Zakhor, A., “Dictionary approximation for matching pursuit video coding,” in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 2, Sept. 2000, pp. 828–831.

[13] Frossard P. and Vandergheynst P. and Figueras i Ventura R. M., “High flexibility scalable image coding,” in *Proceedings of VCIP 2003*, July 2003, pp. 281–296.