Best Basis Compressive Sensing

Gabriel Peyré

▶ **To cite this version:**

Gabriel Peyré. Best Basis Compressive Sensing. 2007. hal-00365017v1

HAL Id: hal-00365017
https://hal.science/hal-00365017v1

Preprint submitted on 2 Mar 2009 (v1), last revised 10 Jan 2010 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Best Basis Compressed Sensing

Gabriel Peyré, *Member, IEEE*

***Abstract*—This paper proposes an extension of compressed sensing that allows to express the sparsity prior in a dictionary of bases. This enables the use of a fixed set of non-adaptive linear measurements and an adaptive recovery process. This reconstruction optimizes the basis to the structure of the sensed signal. An iterative thresholding algorithm is used in order to perform both the recovery and the estimation of the best basis. Numerical experiments on sounds and geometrical images show that adaptivity is indeed crucial to capture the regularity of complex natural signals.**

***Index Terms*—Compressed sensing, best basis, wavelet packets, cosine packets, bandlets.**

I. Introduction

COMPRESSED sensing is a new sampling strategy that uses a fixed set of linear measurements together with a non-linear recovery process. In order for this scheme to work with a low number of measurements, compressed sensing theory requires the sensed signal to be sparse in a given orthogonal basis and the sensing vectors to be incoherent with this basis. This theory of compressed acquisition of data has been proposed jointly by Candes, Tao and Romberg [1], [2] and Donoho [3], [4].

This paper is focussed on enhancing the compressive sensing technology by using a adaptive recovery process. While the original theory uses a fixed basis to express the sparsity of a given signal, we propose to use a more general dictionary of bases. In this framework, the signal to recover is assumed to be sparse in at least one basis of the whole dictionary.

A. Compressed Sensing and Sparsity

In a series of papers, Candes, Tao and Romberg [1], [2] and Donoho [3], [4] have proposed the idea of directly acquiring signal in a compressed form. Instead of performing the acquisition with a high sampling rate and then compressing the data in an orthogonal basis, the signal is simply projected on a reduced set of linear vectors. The compressibility of the signal is only exploited during the reconstruction phase, where one needs to use the sparsity of the signal in an orthogonal basis.

Compressed sensing acquisition of data might have an important impact for the design of imaging devices where data acquisition is expensive. Takhar et al. [5] detail a single pixel camera that acquires random projections from the visual scene through a digital micromirror array. A similar acquisition strategy can be used in MRI imaging [6] to reduce the acquisition time and increase the spatial resolution.

Gabriel Peyré is with the Ceremade, Université Paris Dauphine, 75775 Paris Cedex 16 France, email: gabriel.peyre@ceremade.dauphine.fr

B. Previous Works

In order for the compressed sensing approach to be efficient, one needs to focus on the recovery phase and exploit the structure of the signal to recover.

Redundancy to increase sparsity. Fixed orthogonal bases are not flexible enough to capture the complex regularity of sounds or natural images. For instance the orthogonal wavelet transform lacks of translation and rotation invariance and a fixed local cosine basis fail to capture transient part of musical sounds [7].

A first way to increase the sparsity of the representation is to increase the redundancy of the basis. One can can for instance use a frame of translation invariant wavelets which has proven useful for image denoising. The curvelet frame [8] allows to better represent edges in images than wavelets. The resulting lack of orthogonality might however deteriorate the coherence of the basis with the sensing vectors. Incoherence in the case of union of orthogonal bases is studied in [9] and more general frame are studied in [10].

Best basis to increase sparsity. These fixed bases or frames are however not efficient enough to compress optimally complex geometric images or musical sounds. Instead of increasing the redundancy of a single set of vectors, one can consider several orthogonal bases that compose a large dictionary of atoms. In this framework, one has to choose a best basis adapted to the signal to process.

Local cosine bases [7] divide the time axis in segments that are adapted to the local frequency content of the sound. Other kinds of dictionaries of 1D bases have been proposed, such as the wavelet packets dictionary [11] and non stationary wavelet packets [12].

The set of cartoon images is a simple model that captures the sketch content of natural images. The curvelet frame of Candès and Donoho [8] can deal with such a regularity and enjoys a better approximation rate than traditional isotropic wavelets. This result can be enhanced using a dictionary of locally elongated functions that follow the image geometry. Bandlets bases of Le Pennec and Mallat [13] provide such a geometric dictionary together with a fast optimization procedure to compute a basis adapted to a given image.

All these dictionaries allow to increase the sparsity of complex signals and have direct applications in data compression and denoising. They have however not yet been used to solve inverse problems such as compressed sensing reconstruction, since such problems make the estimation of the best basis a difficult task. This paper shows how to actually do this estimation using an iterative method which allows to extend compressed sensing to a best basis framework.

Enhancing compressed sensing. Other approaches to enhance

compressed sensing reconstruction consist in imposing further constraints beyond sparsity. For instance, wavelets coefficients can be optimized in a scale-by-scale fashion [14] and positivity or low total variation can be enforced as well [15]. One can also use more advanced signals models not based on sparsity but rather on low dimensional smooth manifolds [16]. All these ideas can be incorporated in our framework as additional constraints that helps the recovery process.

C. Contributions

The first contribution of this paper is the extension of the compressed sensing to the setting of a dictionary of bases. This implies considering an energy to optimize both on the signal to recover and on the basis that sparsifies this signal. The exploration of both the set of signals and the set of bases is however not tractable numerically. The second contribution of this paper is a fast iterative algorithm that progressively estimates the best basis. This algorithm is derived using a relaxed variational energy that approximates the true energy by a series of surrogate functionals easy to optimize using thresholdings and fast best basis searches.

II. Compressed Sensing

Sparsity and incoherence. In the following, the original signal f to be recovered is assumed to be finite dimensional $f \in \mathbb{R}^N$ with $N \gg n$. The compressed sensing acquisition scenario uses a fixed set of n linear measurements

$$y = \Phi f = (\langle f, \varphi_i \rangle)_{i=1}^n \in \mathbb{R}^n \quad \text{with} \quad \varphi_i \in \mathbb{R}^N.$$

The price to pay for this compressed sampling strategy is a non-linear reconstruction procedure to recover f from the compressed representation $y = \Phi f$. This recovery process makes use of an orthogonal basis $\Psi = (\psi_m)_{m=0}^{N-1}$ of $\mathbb{R}^N$ and is defined as

$$f^\star = \operatorname*{argmin}_{g \in \mathbb{R}^N} \|\Psi^{\mathrm{T}} g\|_{\ell^1} \qquad \text{subject to} \qquad \Phi g = y, \quad (1)$$

where $\Psi^{\mathrm{T}} g = (\langle g, \psi_m \rangle)_m$ and $\|x\|_{\ell^1} = \sum_m |x_m|$.

In order for this recovery to be efficient, the compressed sensing theory requires two constraints:

- *Sparsity:* the signal f should be sparse in the basis Ψ. It means that f can be represented using only a small number $s \ll N$ of atoms from Ψ

$$\|\Psi^{\mathrm{T}} f\|_{\ell^0} \overset{\text{def.}}{=} \#\{m \setminus \langle f, \psi_m \rangle \neq 0\} \leqslant s. \quad (2)$$

- *Incoherence:* the sensing vectors $(\varphi_i)_i$ should be as different as possible from the sparsity vector $(\psi_m)_m$. This is ensured by monitoring the s-restricted isometry constant δ_s, which is the smallest $0 < \delta < 1$ such that

$$(1-\delta)\|c\|_{\ell^2}^2 \leqslant \|(\Phi\Psi)_T c\|_{\ell^2}^2 \leqslant (1+\delta)\|c\|_{\ell^2}^2. \quad (3)$$

  for all subsets T such that $\#T \leqslant s$ and for all coefficients $(c_m)_{m \in T}$, where A_T is the sub-matrix extracted from A by selecting the columns in T.

One should note that although the sparity assumption (2) constraints the ℓ^0 norm of $\Psi^{\mathrm{T}} f$, the actual recovery process (1) optimizes the ℓ^1 norm. This is important since the ℓ^1 norm is convex, which leads to a tractable optimization problem with fast algorithms, see [17].

The following recovery theorem allows to assess the perfect recovery using the ℓ^1 minimization.

Theorem 1. *([1], [2]) If f is s-sparse in Ψ as defined in (2) and if the sensing matrix Φ satisfies $\delta_{2s} + \delta_{3s} < 1$, where δ_s is defined in (3), then the solution of (1) satisfies $f^\star = f$.*

Random sensing matrices. The allowable sparsity s for which $\delta_{2s} + \delta_{3s} < 1$ needs to be large enough for theorem 1 to be useful. Hopefully, one can exhibit sets of matrices for which s can be taken of the order of n, up to logarithmic factors. These sets of matrices include

- Gaussian matrices: the entries of Φ are drawn independently from a gaussian distribution of variance $1/N$.
- Fourier sub-matrices: Φ is obtained by selecting n rows at random from the Fourier matrix $(\exp(2i\pi/nk\ell)_{k,\ell}$.

For these matrix ensembles, the following theorem holds.

Theorem 2. *([1], [2], [3], [4], [18]) If the sparsity obeys*

$$s \leqslant Cn/\log(N)^\kappa,$$

where C is a constant and with $\kappa = 1$ for gaussian and $\kappa = 4$ for Fourier matrices, then Φ satisfies $\delta_{2s} + \delta_{3s} < 1$ with a probability that increases toward 1 exponentially fast with N.

The logarithmic factor is actually $1/\log(N/s)$ for the gaussian ensemble [4]. Other matrices ensembles have been shown to exhibit similar behaviors, see the review [19].

Robust Compressed Sensing. To deal with noisy measurements $y = \Phi f + w$, where w is a Gaussian noise of variance t^2, one can turn the constrained formulations (1) into a penalized variational problem

$$f^\star = \operatorname*{argmin}_{g \in \mathbb{R}^N} E(g, \mathcal{B}, t) \quad \text{where} \quad (4)$$

$$E(g, \mathcal{B}, t) \overset{\text{def.}}{=} \frac{1}{2}\|\Phi g - y\|_{\ell^2}^2 + t\|\Psi^{\mathrm{T}} g\|_{\ell^1}. \quad (5)$$

The Lagrange multiplier t accounts both for stabilization against noise and approximate sparsity, which is common in practical applications.

Compressed sensing theory has been extended to this noisy setting [20], [4], where it is proved that the recovery error $\|f - f^\star\|$ is of the order of the noise level t.

An other interest of the formulation 4 is that during an iterative algorithm, the Lagrange multiplier t can be decreased toward 0 (or to a positive value in the noisy case). Indeed the solution of (4) approaches the noiseless solution of (1) when $t \to 0$. We use such Lagrange multiplier formulation in the fast recovery algorithm detailed in section III-B.

III. Best Basis Compressed Sensing

The recovery theorem 1 holds for a fixed orthogonal basis of $\mathbb{R}^N$. In order to deal with complex natural sounds and images, a fixed basis is not enough and one needs more basis elements. This paper thus proposes to extend the recovery process (4) to a library of orthogonal bases, the union of which contains a large collection of atoms.

A. Dictionaries and Lagrangian

Dictionaries of Orthogonal bases. A dictionary is a set $\mathcal{D}_\Lambda = \{\mathcal{B}(\lambda)\}_{\lambda\in\Lambda}$ of orthogonal bases $\mathcal{B}(\lambda) = \{\psi_m^\lambda\}_m$ of $\mathbb{R}^N$. The parameter $\lambda \in \Lambda$ is an index that needs to use in order to process some function f using $\mathcal{B}(\lambda)$. Instead of using an a priori fixed basis such as the wavelet or Fourier basis, one can freely use any basis from the dictionary.

An alternative to this dictionary of orthogonal basis $\mathcal{D}_\Lambda$ consists in using directly the union of all basis vectors $D = \{\psi_m^\lambda \setminus \lambda \in \Lambda, m\}$. Such non-orthogonal dictionaries can be used in conjunction with sparse decomposition solvers such as matching pursuit [7] or basis pursuit [17]. This method is however intractable in real applications where D is too large. In contrast, the dictionaries of orthogonal bases we consider all come with fast algorithms to compute a basis adapted to the function to process.

Lagrangian and adapted basis. A best basis $\mathcal{B}(\lambda^\star)$ adapted to a signal or an image $f \in \mathbb{R}^N$ should lead to a good approximation of f with a small number of atoms, as measured by the ℓ^1 norm. Such an adapted basis minimizes the Lagrangian $\mathcal{E}$

$$\lambda^\star = \operatorname*{argmin}_{\lambda\in\Lambda} \mathcal{E}(f, \mathcal{B}(\lambda), t) \tag{6}$$

$$\mathcal{E}(f, \mathcal{B}(\lambda), t) \overset{\text{def.}}{=} \min_{g\in\mathbb{R}^N} \frac{1}{2}\|f-g\|^2 + t\|\Psi^{\lambda\,\mathrm{T}} g\|_{\ell^1}$$

and where $\Psi^\lambda = (\psi_0^\lambda, \ldots, \psi_{N-1}^\lambda) \in \mathbb{R}^{N\times N}$ is the orthogonal transform matrix defined by $\mathcal{B}(\lambda)$. The variable t is a Lagrange multiplier that weights the quality of approximation in the chosen basis with the sparsity of the expansion. One can add a penalty term to $\mathcal{E}$ that depends on the complexity of λ in order to avoid using too complicated bases. We do not consider such penalty in this article.

The following lemma characterize the best basis together with the best ℓ^1-penalized approximation in this basis.

Lemma 1. *The minimizer*

$$(f^\star, \lambda^\star) = \operatorname*{argmin}_{(g,\lambda)\in\in\mathbb{R}^N\times\Lambda} \frac{1}{2}\|f-g\|^2 + t\|\Psi^{\lambda\,\mathrm{T}} g\|_{\ell^1}$$

$$\textit{is given by} \quad \begin{cases} \lambda^\star = \operatorname*{argmin}_{\lambda\in\Lambda} \mathcal{E}(f, \mathcal{B}(\lambda), t), \\ f^\star = S_t(f, \mathcal{B}(\lambda^\star)), \end{cases}$$

where the soft thresholding operator is defined as

$$S_t(f, \mathcal{B}) = \sum_m s_t(\langle f, \psi_m\rangle)\, \psi_m \tag{7}$$

$$\textit{where} \quad s_t(x) = \begin{cases} x + \operatorname{sign}(x)t & \textit{if} \quad |x| > t, \\ 0 & \textit{if} \quad |x| \leqslant t. \end{cases}$$

Proof: The best basis parameter solves

$$\begin{aligned}\lambda^\star &= \operatorname*{argmin}_{\lambda\in\Lambda} \min_{g\in\mathbb{R}^N} \frac{1}{2}\|f-g\|^2 + t\|\Psi^{\lambda\,\mathrm{T}} g\|_{\ell^1} \\ &= \operatorname*{argmin}_{\lambda\in\Lambda} \mathcal{E}(f, \mathcal{B}(\lambda), t).\end{aligned}$$

The best ℓ^1 approximation $f^\star$ is found by using the following standard optimization result, see for instance [21]

$$\operatorname*{argmin}_g \frac{1}{2}\|f-g\|^2 + t\|\Psi^{\mathrm{T}} g\|_{\ell^1} = S_t(f, \mathcal{B}). \tag{8}$$

■

As noted in equation (8), the thresholding S_t introduced in (7) solves the best ℓ^1 approximation of a function f. The Lagrangian $\mathcal{E}(f, \mathcal{B}(\lambda), t)$ can thus be written as a sum over the atoms $(\psi_m^\lambda)_m$ of $\mathcal{B}(\lambda)$

$$\mathcal{E}(f, \mathcal{B}(\lambda), t) = \sum_m \gamma(|\langle f, \psi_m^\lambda\rangle|) \tag{9}$$

$$\text{where} \quad \gamma(x) = \begin{cases} x^2/2 & \text{if } |x| \leqslant t, \\ t|x| - t^2/2 & \text{otherwise.} \end{cases}$$

B. Best Basis Compressed Sensing Reconstruction

Best basis energy. The compressed sensing machinery is extended to a dictionary of bases $\mathcal{D}_\Lambda$ by imposing that the recovered signal is sparse in at least one basis of $\mathcal{D}_\Lambda$. The original recovery procedure (4) is replaced by

$$(f^\star, \lambda^\star) \overset{\text{def.}}{=} \operatorname*{argmin}_{(g,\lambda)\in\mathbb{R}^N\times\Lambda} E(g, \mathcal{B}(\lambda), t), \tag{10}$$

where the energy E is defined in equation (4).

Surrogate functional. Searching in the whole dictionary $\mathcal{D}_\Lambda$ for the best basis parameter $\lambda^\star \in \Lambda$ that minimizes (10) is not feasible for large dictionaries, which typically contain of the order of 2^N bases. Furthermore, the under-determinancy of Φ creates coupling in the non-linear set of equations involved in the minimization of E.

To solve these issues, one can relax the energy minimization (10) and use an approximate energy that is simpler to minimize. If one has some estimate h of the solution $f^\star$, the energy E can be replaced by the following surrogate functional that depends on h

$$E_h(g, \mathcal{B}(\lambda), t) \overset{\text{def.}}{=} E(g, \mathcal{B}, t) + \frac{\mu}{2}\|g-h\|^2 - \frac{1}{2}\|\Phi g - \Phi h\|^2.$$

where the constant $\mu > 0$ is chosen to ensure $\|\Phi g\|^2 \leqslant \mu\|g\|^2$ for all $g \in \mathbb{R}^N$.

If the estimate h is fixed, one can replace the optimization (10) by

$$(f^\star(h), \lambda^\star(h)) \overset{\text{def.}}{=} \operatorname*{argmin}_{(g,\lambda)\in\mathbb{R}^N\times\Lambda} E_h(g, \mathcal{B}(\lambda), t). \tag{11}$$

The following theorem, which is the main theoretical result of this paper, ensures that the minimization (11) is indeed easy to solve.

Theorem 3. *The minimization* (11) *has a global minimum which is given by*

$$\begin{cases} \lambda^\star(h) = \operatorname*{argmin}_{\lambda\in\Lambda} \mathcal{E}(\tilde{h}, \mathcal{B}(\lambda), t), \\ f^\star(h) = S_{t/\mu}(\tilde{h}, \mathcal{B}(\lambda^\star(h))), \end{cases} \tag{12}$$

where $\tilde{h} \overset{\text{def.}}{=} h + \frac{1}{\mu}\Phi^{\mathrm{T}}(y - \Phi h)$*, where the Lagrangian* $\mathcal{E}$ *is defined in equation* (6) *and the soft thresholding operator is defined in equation* (7)*.*

Proof: The definition of μ makes the energy E_h strictly convex, it thus has a single global minimum. The energy E_h

is expanded as follow

$$\begin{aligned} E_h(g, \mathcal{B}(\lambda), f) &= \frac{\mu}{2}\|g\|^2 + \langle \Phi g,\, y\rangle + \langle \Phi g,\, \Phi h\rangle - \mu\langle g,\, h\rangle \\ &\qquad + t\|\Psi^{\lambda^{\mathrm{T}}} g\|_{\ell^1} + C \\ &= \frac{\mu}{2}\|g\|^2 - \langle g,\, \mu h + \Phi^{\mathrm{T}}(y - \Phi h)\rangle \\ &\qquad + t\|\Psi^{\lambda^{\mathrm{T}}} g\|_{\ell^1} + C \end{aligned}$$

where C is a constant independent of λ and g. Up to a multiplicative and additive constant, one can thus write as

$$E_h(g, \mathcal{B}(\lambda), t) \propto \frac{1}{2}\|h + \frac{1}{\mu}\Phi^{\mathrm{T}}(y - \Phi h) - g\|^2 + t\|\Psi^{\lambda^{\mathrm{T}}} g\|_{\ell^1}.$$

The result of the theorem follows from lemma 1. ∎

Equation (12) shows that the best basis parameter $\lambda^\star(h)$ corresponds to the best basis associated to the modified guess $\tilde{h}$. It also shows that $f^\star(h)$ is obtained by thresholding $\tilde{h}$ in that basis. The main interest of this theorem is that the best parameter $\lambda^\star(h)$ is found by optimizing a Lagrangian $\mathcal{E}$, which can be achieved with a fast algorithm, as explained in the section IV.

C. Best Basis Recovery Algorithm

In order to minimize $E(g, \mathcal{B}(\lambda), t)$ over (g, λ), one can minimize a set of surrogate functionals $E_{g_s}(g, \mathcal{B}(\lambda), t)$ where g_s is the current estimate of the solution at iteration s. Instead of using a fixed multiplier t, one decreases this threshold through iterations, which results in faster convergence.

Listing 1 Best-basis compressed sensing algorithm.

Initialization. Set $s = 0$, $f_0 = 0$ and $t_0 = t_{\max}$.
Step 1: Updating the estimate. Set

$$\tilde{f}_s = f_s + \frac{1}{\mu}\Phi^{\mathrm{T}}(y - \Phi\, f_s).$$

Step 2: Update best basis. Compute the best basis parameter

$$\lambda_s = \operatorname*{argmin}_{\lambda\in\Lambda} \mathcal{E}(\tilde{f}_s, \mathcal{B}(\lambda), t_s).$$

This minimization is carried out with a fast procedure, as detailed in section IV. **Step 3: Denoising the estimate.** Compute

$$f_{s+1} = S_{t_s/\mu}(\tilde{f}_s, \mathcal{B}(\lambda_s)),$$

where S_{t_s} is the threshold operator defined in (7).
Stopping criterion. If $s < s_{\max}$, set $t_{s+1} = t_s - (t_{\max} - t_\mu)/s_{\max}$, $s \leftarrow s+1$ and go to step 1, otherwise stop iterations.

The steps of the algorithm are detailed in pseudo-code 1. Similar iterative thresholding algorithms have been proposed to solve inverse problems such as (4), see for instance [22] and the references therein. Our procedure extends this kind of methods to a dictionary of bases and require, at each step, the estimation of a best basis.

In contrast to the fixed basis setting, the iterations with a best basis search are difficult to analyze. In particular, there is no recovery results such as theorem 1for the output of the algorithm. We thus rely on the numerical experiments of sections V-B, VI-B and VII-B to access the efficiency of this recovery process on both synthetic and natural signals and images.

IV. Best Basis Extraction

This section reviews a best basis framework common to all the dictionaries used in the numerical experiments of this paper. In order to have a fast algorithm for the best basis extraction, one has to impose some hierarchical structure on the set of bases. This can be done by using trees that parameterize the dictionary.

A. Tree structured dictionaries

This paper focusses on dictionaries $\mathcal{D}_\Lambda$ having a multiscale structure in order to have fast decomposition procedures. In particular, we suppose that each $\lambda \in \Lambda$ that parameterizes a basis $\mathcal{B}(\lambda) \in \mathcal{D}_\Lambda$ is a binary tree (for mono-dimensional signals, where $d = 1$) or a quad-tree (for bi-dimensional images, where $d = 2$). The set of nodes is denoted as $\mathcal{N}(\lambda)$ and each node $p_i^j \in \mathcal{N}(\lambda)$ is located at some level $0 \leqslant j < J \stackrel{\text{def.}}{=} \log_2(N)/d$ and position $0 \leqslant i < 2^{dj}$. A node p_i^j is thus located in the j^{th} row and the i^{th} column of the tree. Each interior node $p_i^j \in \mathcal{I}(\lambda) \subset \mathcal{N}(\lambda)$ has 2^d children $\{n_{2^d i}^{j+1}, \ldots, n_{2^d(i+1)-1}^{j+1}\}$. The leaves nodes $p_i^j \in \mathcal{L}(\lambda)$ have no child.

In order to incorporate additional information, each node $p_i^j \in \mathcal{N}(\lambda)$ is tagged with a token $\ell(p_i^j) \in \{0, \ldots, \kappa-1\} \cup \emptyset$. The special token $\emptyset$ indicates that no information is defined for this node. This token is used in the non-stationary wavelet packets to select a filter at each scale and in the bandlets dictionary to describe the local direction of an edge. The value of κ indicates the number of such tokens (for instance the number of edges directions in the bandlet framework). Figure 1 shows an example of such a tagged binary tree λ.

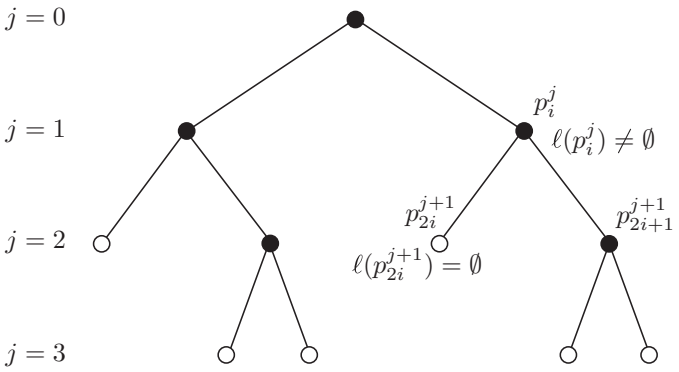

Fig. 1. *Examples of binary tree λ.*

The vectors of a basis $\mathcal{B}(\lambda)$ are grouped in the leaves of the tree λ so that one can write

$$\mathcal{B}(\lambda) = \left\{ \psi_{i,s}^j \setminus \forall\, p_i^j \in \mathcal{L}(\lambda) \quad \text{and} \quad 0 \leqslant s < 2^{dj} \right\}.$$

The variable s indexes the atoms of the basis that are clustered in the node p_i^j of the tree. When one does not care about the location of the basis elements in the tree, the basis is written

as $\mathcal{B}(\lambda) = \{\psi_m\}_m$ where the index is $m = (j, i, s)$ with $p_i^j \in \mathcal{L}(\lambda)$ and $0 \leqslant s < 2^{dj}$.

Examples of dictionaries for sounds and geometrical images. The following sections V, VI and VII detail three kinds of dictionaries adapted to various signals and images structures.

- **The local cosine dictionary** [11], [7] is used to process highly oscillating signals ($d = 1$) such as music and speech audio data. The binary tree λ segments the time axis in order to match the variations of the local frequencies in the sound. In this case $\ell(p_i^j) = \emptyset$ for all nodes since there is no need for additional information beside the spatial segmentation.
- **The non-stationary wavelet packets dictionary** [12] is used to process signals ($d = 1$) that require an arbitrary tiling of the scale axis. The non-stationary cascade of filterings also allows to adapt the basis functions through the scales. In this case, $\ell(p_i^j) \in \{0, \ldots, \kappa - 1\}$ indicates the index of a wavelet filter that is used to subdivide the frequency axis. The particular case of wavelet packets [11] is obtained for $\kappa = 1$.
- **The bandlet dictionary** [13], [23] is used to process images ($d = 2$) with geometric features such as edges or directional textures. The quad-tree λ segments the square $[0, 1]^2$ into sub-squares $S_i^j \subset [0, 1]^2$ of size $2^{-j} \times 2^{-j}$. An interior node $p_i^j \in \mathcal{I}(\lambda)$ caries no information beside spatial segmentation and thus $\ell(p_i^j) = \emptyset$ for these nodes. A leaf node $p_i^j \in \mathcal{L}(\lambda)$ defines the local bandlet transform and caries a geometrical information $\ell(p_i^j) \in [0, \pi[\cup \emptyset$. A token $\ell(p_i^j) \in [0, \pi[$ is define over geometrical square in which the bandlet vector are elongated and follow the direction $\ell(p_i^j)$. A token $\ell(p_i^j) = \emptyset$ is defined in isotropic square that corresponds to regions where the image is uniformly regular.

Other classes of dictionaries include bases composed of atoms with rapidly varying oscillations such as the modulated bases of Coifman et al. [24] and the chirplets dictionary of Candès [25].

B. Dynamic programming for Best-basis Computation

Fast best basis search algorithms exploit the tree structure of the dictionaries of interest. The corresponding algorithm in the local cosine dictionary is explained in [11], [7], the search in the non-stationary wavelet packet dictionary is explained in [26] and the search in the bandlet dictionary in [13], [23]. All these algorithms correspond to particular instances of the Classification and Regression Tree (CART) algorithm of Breidman et al. [27] as explained by Donoho [28].

In order to give the details of this optimization algorithm, we restrict ourself to dictionaries $\mathcal{D}_\Lambda$ where $\ell(p_i^j) = \emptyset$ for any interior node $p_i^j \in \mathcal{I}(\lambda)$ and $\lambda \in \Lambda$. This does not cover the case of non-stationary wavelet packets, which requires a more complex optimization procedure described in [26]. In the following, the dependency of the Lagrangian $\mathcal{E}$ on both f and t is made implicit and we write $\mathcal{E}(\mathcal{B}) \stackrel{\text{def.}}{=} \mathcal{E}(f, \mathcal{B}, t)$.

Equation (9) shows that for a tree-structured basis $\mathcal{B}(\lambda) = \left\{\psi_{i,s}^j \setminus p_i^j \in \mathcal{L}(\lambda)\right\}$, the Lagrangian is decomposed as a sum over the Lagrangians associated to each leaf

$$\mathcal{E}(\mathcal{B}(\lambda)) = \sum_{p_i^j \in \mathcal{L}(\lambda)} \mathcal{E}(\mathcal{B}(p_i^j, \ell(p_i^j)))$$

$$\mathcal{B}(p_i^j, \ell(p_i^j)) \stackrel{\text{def.}}{=} \left\{\psi_{i,s}^j \setminus s = 0, \ldots, 2^{dj} - 1\right\}$$

Since we consider bases such that $\ell(p_i^j) = \emptyset$ on interior nodes $p_i^j \in \mathcal{I}(\lambda)$, it means that each set of vectors $\mathcal{B}(p_i^j, \ell(p_i^j)) \subset \mathcal{B}(\lambda)$ is parameterized by $p_i^j \in \mathcal{L}(\lambda)$ and $\ell(p_i^j) \in \{0, \ldots, \kappa - 1\} \cup \emptyset$.

One does not need to explore all the bases $\mathcal{B}(\lambda)$ for $\lambda \in \Lambda$ to compute the best basis $\mathcal{B}(\lambda^\star)$. It is enough to compute the decomposition of f on the elementary bases $\mathcal{B}(p_i^j, \ell)$ for all possible choices of p_i^j and ℓ. The best basis algorithm process as follow.

- **Initialization.** Set λ as the full tree of depth $J = \log_2(N)/d$. Set j=J.
- **Best token selection.** For each $0 \leqslant i < 2^{jd}$ and $\ell \in \{0, \ldots, \kappa - 1\} \cup \emptyset$ compute the Lagrangian $\mathcal{E}(\mathcal{B}(p_i^j, \ell))$ associated to the basis vectors corresponding to (p_i^j, ℓ). The best token is defined as

$$\ell^\star(p_i^j) = \operatorname*{argmin}_{\ell=0,\ldots,\kappa-1} \mathcal{E}(\mathcal{B}(p_i^j, \ell)).$$

- **Split/merge decision.** Compute the Lagrangian of the sub-tree as the sum of the Lagrangians of the children nodes of p_i^j: if $j < J$,

$$\tilde{\mathcal{E}}(\mathcal{B}(\lambda_i^j)) \stackrel{\text{def.}}{=} \sum_{\varepsilon=0}^{2^d-1} \mathcal{E}(f, \mathcal{B}(\lambda_{2^d i+\varepsilon}^{j+1}), t),$$

with the convention that $\tilde{\mathcal{E}}(\mathcal{B}(\lambda_i^j)) = +\infty$ for $j = J$.
Split: If $\tilde{\mathcal{E}}(\mathcal{B}(\lambda_i^j)) \leqslant \mathcal{E}(\mathcal{B}(p_i^j, \ell^\star(p_i^j)))$ then the node p_i^j is declared as an interior node $p_i^j \in \mathcal{I}(\lambda)$ and $\ell(p_i^j) \stackrel{\text{def.}}{=} \emptyset$.
Merge: If $\tilde{\mathcal{E}}(\mathcal{B}(\lambda_i^j)) > \mathcal{E}(\mathcal{B}(p_i^j, \ell^\star(p_i^j)))$ then the node p_i^j is declared as an leaf node $p_i^j \in \mathcal{L}(\lambda)$ and $\ell(p_i^j) \stackrel{\text{def.}}{=} \ell^\star(p_i^j)$. The 2^d sub-trees starting below p_i^j are removed from λ.
The Lagrangian $\mathcal{E}(\mathcal{B}(\lambda_i^j))$ is set to

$$\min\left(\tilde{\mathcal{E}}(\mathcal{B}(\lambda_i^j)), \mathcal{E}(\mathcal{B}(p_i^j, \ell^\star(p_i^j)))\right).$$

- **Stopping criterion.** If $j > 0$, set $j \leftarrow j - 1$ and repeat the previous two steps.

This algorithm requires the decomposition of f onto each elementary basis $\mathcal{B}(p_i^j, \ell)$ for all values of (p_i^j, ℓ). The complexity of this best basis search is thus $O(\kappa\, C(N) \log_2(N))$ where $C(N)$ is the cost of transforming a signal of size N. One has $C(N) = N$ for the wavelet packet and bandlet dictionaries and $C(N) = N \log_2(N)$ for the cosine packet dictionary.

C. Settings for the Numerical Results

The following sections detail several dictionaries of orthogonal bases. The performance of these dictionaries is illustrated for compressed sensing recovery using the same numerical experiments. The recovery success is measured using

$$\text{PSNR}(f, f^\star) = -20 \log_{10}(\|f - f^\star\|),$$

where the signals are assumed to take values in $[0,1]$. This recovery error is measured for various values of the sensing rate $n/N \in [0,1]$. Three kinds of recoveries are compared:

- Recovery using a fixed basis (for instance fixed local DCT or fixed orthogonal basis) using the original optimization (1).
- Recovery using the oracle best-basis $\mathcal{B}(\lambda^\star(f))$ estimated from the original signal $f^\star$. This is an upper-bound for the performance of our method since in this basis is not available in practice.
- Recovery using the algorithm of section III-C, that estimates iteratively the best basis.

The numerical experiments are done in the noiseless setting, so the final threshold is set to $t_{\min} = 0$ in the iterative algorithm.

V. Best Local Cosine Basis Compressed Sensing

A. Adapted Local Cosine Transform

A local cosine basis $\mathcal{B}(\lambda)$ is parameterized with a binary tree λ that segments the time axis in dyadic intervals, see [11], [7]. Each leaf node $p_i^j \in \mathcal{L}(\lambda)$ corresponds to a selected interval $[x_i^j, x_{i+1}^j]$, where $x_i^j \overset{\text{def.}}{=} 2^{-j}Ni - 1/2$. For each of these leave node, the local cosine basis vectors are defined as

$$\forall\, p_i^j \in \mathcal{L}(\lambda), \quad \forall\, k, \quad \psi_{i,s}^j[k] = b\big(2^j(k - x_i^j)\big)\sqrt{\frac{2}{2^{-j}N}}\cos\left[\pi\big(s+\frac{1}{2}\big)\frac{k-x_i^j}{2^{-j}N}\right],$$

where b is a smooth windowing function that satisfies some compatibility conditions [7].

The decomposition of a signal f on the vectors of some basis $\mathcal{B}(\lambda)$ is computed in $O(N\log(N))$ using several fast Fourier transforms. A best basis $\mathcal{B}(\lambda^\star)$ that minimizes (6) can be extracted by a tree pruning procedure in $O(N\log(N)^2)$ time, see [11], [7]. Figure 2 shows some examples of basis vectors.

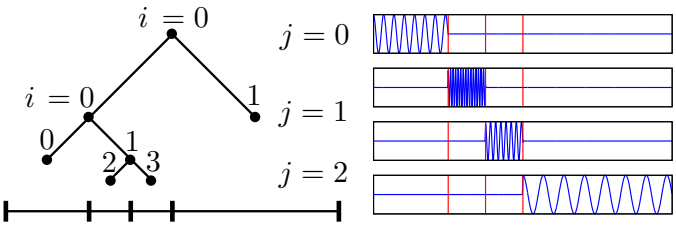

Fig. 2. *A dyadic tree λ defining a spatial segmentation (left) ; some local cosine basis functions $\psi_{i,s}^j$ of the basis $\mathcal{B}(\lambda)$ (right).*

B. Numerical Results

A synthetic sparse signal $f = (\Psi^\lambda)^{-1}h$ is generated using a random local cosine basis $\mathcal{B}(\lambda)$ and a random signal of spikes h with $\|h\|_{\ell^0} = \#\{k \setminus h(k) \neq 0\} = 30$, see figure 3, (a). The signal recovered by the non-adaptative algorithm of section III-C in an uniform cosine basis $\mathcal{B}(\lambda_0)$ is significantly different from the original, figure 3, (b). This is due to the fact that f is less sparse in $\mathcal{B}(\lambda_0)$, since $\|\Psi^{\lambda_0} f\|_{\ell^0} = 512$ and $\|\Psi^{\lambda_0} f\|_{\ell^1} \approx 2.8\|\Psi^\lambda f\|_{\ell^1}$. During the iterations of the algorithm presented in subsection III-B, the estimated best basis $\mathcal{B}(\lambda_s)$ evolves in order to match the best basis $\mathcal{B}(\lambda)$, see figure 3, (c1–c3). The recovered signal (c3) is nearly identical to f.

On figure 4 one can see a sound of a tiger howling, together with the signals recovered using a fixed fully subdivided local DCT basis and the best basis recovery algorithm of section III-B. Although the final adapted basis is not the same as the best basis of the original signal, it still provides an improvement of 2dB with respect to a fixed spatial subdivision. Figure 5 shows for various rates of sensing the recovery error, confirming that the iterative algorithm does not perform as good as the oracle best best basis computed from f.

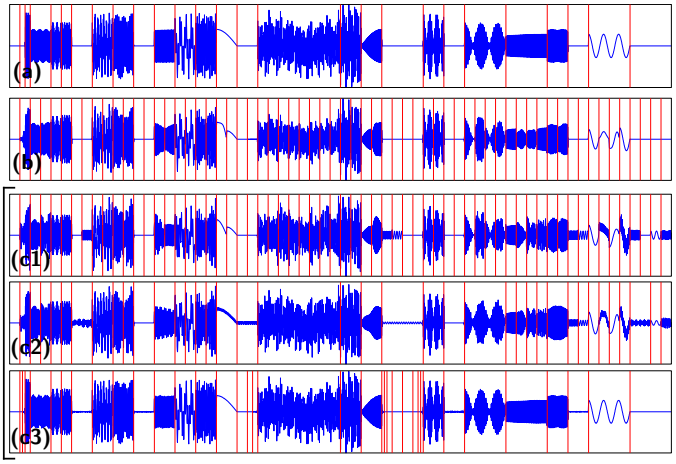

Fig. 3. *(a) synthetic sound signal with 30 random cosine atoms $N = 4096$; (b) recovery using a fixed cosine basis ; (c1) first iteration of the best basis recovery algorithm, $n = N/3$; (c2) iteration $s = 5$; (c3) iteration $s = 20$.*

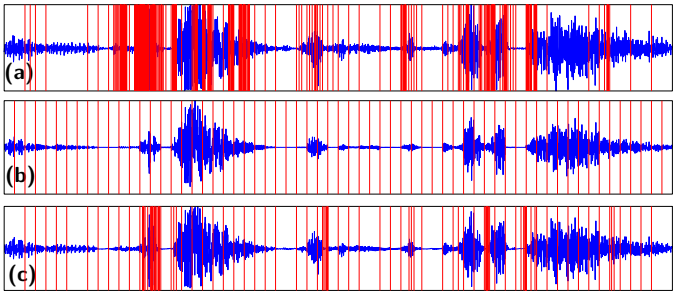

Fig. 4. *(a) sound signal of a tiger howling, together with the best spatial segmentation, $N = 32768$; (b) recovery using fixed local cosine basis, $n = N/3$ (PSNR=19.24dB) ; (c) recovery using best cosine basis, $n = N/3$ (PSNR=21.32dB)*

VI. Best Non-stationary Wavelet Packet Compressed Sensing

A. Adapted Non-stationary Wavelet Packet Transform

The non stationary wavelet transform and its extension to wavelet packets was introduced by Cohen et al. [12]. We give the definition of a non-stationary wavelet transform which corresponds to the decomposition in an orthogonal basis parameterized by a tagged binary tree λ. The best basis algorithm is detailed in [26].

Quadrature mirror filters. The decomposition in a non-stationary (NS) wavelet packet basis is defined though a

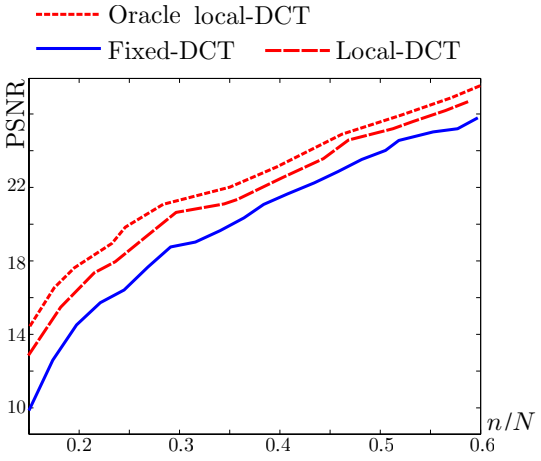

Fig. 5. *Recovery results for the signal of figure 4 for various rate of sensing n/N.*

cascade of filterings. In contrast to a traditional wavelet packet basis, these filterings might vary through the scales. A finite set $H = \{h_0, \ldots, h_{\kappa-1}\}$ of low pass filters is used together with the corresponding quadrature mirror filters

$$\forall\, \ell \in \{0, \ldots, \kappa-1\}, \quad g_\ell[k] = (-1)^k h_\ell[1-k].$$

In this article, we restrict ourselves to orthogonal bases and each filter h_ℓ thus satisfies

$$|\hat{h}_\ell(\omega)|^2 + |\hat{h}_\ell(\omega+\pi)|^2 = 2 \quad \text{and} \quad h_\ell(0) = \sqrt{2}. \tag{13}$$

where the discrete Fourier transform of a vector $f \in \mathbb{R}^n$ is defined as

$$\forall\, \omega \in \mathbb{R}/(2\pi\mathbb{Z}), \quad \hat{f}(\omega) \stackrel{\text{def.}}{=} \sum_k f[k] \exp(-\mathrm{i}k\pi\omega).$$

Discrete algorithm. The decomposition of a discrete signal $f \in \mathbb{R}^N$ in a NS wavelet packet basis $\mathcal{B}(\lambda)$ is defined through an iterative process that traverse the tree λ starting from the root p_0^0 of λ until reaching the set of leaves $\mathcal{L}(\lambda)$. This process is as follow.

- **Initialization.** Set $f_0^0 = f$ on the root node.
- **Filtering.** For each scale $j = 1, \ldots, J-1$, for each interior node $p_i^j \in \mathcal{I}(\lambda)$, the signals on the children nodes are defined as

$$f_{2i}^{j+1} = (f_i^j * h_\ell) \downarrow 2 \quad \text{and} \quad f_{2i+1}^{j+1} = (f_i^j * g_\ell) \downarrow 2 \tag{14}$$

  where $\ell \stackrel{\text{def.}}{=} \ell(p_i^j)$ and $*$ denote the convolution modulo $N/2^j$. The down sampling operator $g \downarrow 2 \in \mathbb{R}^{N/2^{j+1}}$ of a vector $g \in \mathbb{R}^{N/2^j}$ is defined as $(g \downarrow 2)[k] \stackrel{\text{def.}}{=} g(2k)$.
- **Output.** The transformed coefficients are gathered from the the leaves

$$\Psi^\lambda(f) \stackrel{\text{def.}}{=} \left\{ f_i^j \setminus p_i^j \in \mathcal{L}(\lambda) \right\} \in \mathbb{R}^N.$$

Thanks to the quadrature equation (13), the NS wavelet packet transform $f \mapsto \Psi^\lambda(f)$ can be inverted by traversing the tree from the leaf to the node. This leads to the iterations

- **Initialization.** Let f_i^j be the entries of $\Psi^\lambda(f)$ for the leaves $p_i^j \in \mathcal{L}(\lambda)$.
- **Filtering.** For each scale $j = J-1, \ldots, 1$, for each interior node $p_i^j \in \mathcal{I}(\lambda)$, the signal is recovered using

$$f_i^j = (f_{2i}^{j+1} \uparrow 2) * \tilde{h}_\ell + (f_{2i+1}^{j+1} \uparrow 2) * \tilde{g}_\ell$$

  where $\ell \stackrel{\text{def.}}{=} \ell(p_i^j)$ and where $\tilde{h}_\ell[k] = h_\ell[-k]$. The up-sampling operator $g \uparrow 2 \in \mathbb{R}^{N/2^j}$ of a vector $g \in \mathbb{R}^{N/2^{j+1}}$ is defined as $(g \uparrow 2)[k] = g[k/2]$ for $k = 0 \mod 2$ and $(g \uparrow 2)[k] = 0$ otherwise.
- **Output.** The signal is recovered as $f = f_0^0$.

Both the forward transform $f \mapsto \Psi^\lambda(f)$ and the reverse transform are computed in $O(N)$ operations for a signal $f \in \mathbb{R}^N$.

NS wavelet packet basis. Due to the quadrature property (13), the NS wavelet packet transform Ψ^λ is an orthogonal transforms. The transform Ψ^λ is thus equivalent to the decomposition on set of orthogonal vectors

$$\forall\, p_i^j \in \mathcal{L}(\lambda),\ \forall\, 0 \leqslant s < n/2^j, \quad f_i^j[s] = \langle f,\, \psi_{i,s}^j \rangle.$$

These vectors defines an orthogonal NS wavelet packet basis $\mathcal{B}(\lambda)$ of $\mathbb{R}^N$

$$\mathcal{B}(\lambda) \stackrel{\text{def.}}{=} \left\{ \psi_{i,s}^j \setminus p_i^j \in \mathcal{L}(\lambda) \text{ and } 0 \leqslant s < n/2^j \right\}.$$

In order to compute the basis $\mathcal{B}(\lambda^\star)$ defined by (6) adapted to some function f, a fast best-basis search is introduced by Peyré and Ouarti [26]. The complexity algorithm is $O(N \log_2(N))$ for $\kappa = 1$ (which corresponds to the traditional wavelet packets) and $O(N^{\log_2(2\kappa)})$ for $\kappa > 1$.

B. Numerical Results

In our tests, the non-stationary wavelet packet dictionary is built using the family $H_D = \{h_i\}_{i=0}^{\kappa-1}$ where $\kappa = 5$ and where h_ℓ is the Daubechies orthogonal filter with $\ell+1$ vanishing moments. Figure 6 (c,d,e) shows a comparison of the recovery using the Daubechies wavelets corresponding to h_3, the wavelet packets dictionary which corresponds to $H = \{h_3\}$ and the non-stationary wavelet packets dictionary which corresponds to $H = H_D$. The signal $f(x) = f_0(x) + \varepsilon \sin(\omega x)$ is the superposition of a piecewise-regular signal and a sinusoid with high frequency ω. Figure 6 (a) shows the index $\ell(p_i^j)$ of the best NS wavelet packets basis, which is able to capture both the high frequency content of $\sin(\omega x)$ while minimizing the number of large coefficients created by the singularities of f_0.

Figure 7 shows another example of recovery using wavelets, the wavelet packets and the NS wavelet packets dictionaries. The signal $f(x) = \cos(\omega_1 x^2) + \cos(\omega_2 x)$ is the superposition of a quadratic chirp and a high frequency sinusoid.

VII. Best Bandlet Basis Compressed Sensing

A. Adapted Bandlet Transform

The bandelet bases dictionary was introduced by Le Pennec and Mallat [13], [29] to perform adaptive approximation of images with geometric singularities, such as the cartoon image in figure 9, (a). This transform has been refined by Mallat and Peyré [23] in order to have a dictionary of regular and orthogonal basis functions. We present a simple implementation of the bandelet transform inspired from [30]. This implementation results in a decomposition similar to the wedgelets of Donoho [31] but within the framework of a dictionary of orthogonal bases.

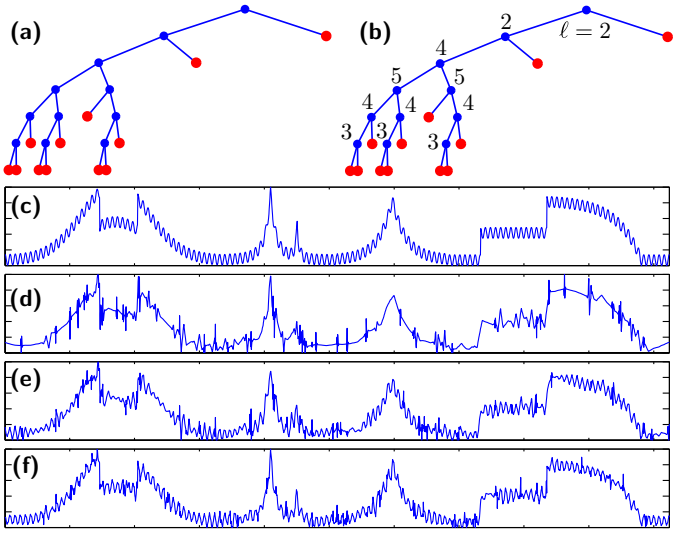

Fig. 6. *(c) original signal, $N = 4096$; (d) recovery using orthogonal wavelets, $n = N/3$ (PSNR=23.77dB) ; (e) recovery using best wavelet packets basis, $n = N/3$ (PSNR=26.50dB), the corresponding dictionary tree is shown in (a) ; (e) recovery using best non-stationary wavelet packets basis, $n = N/3$ (PSNR=29.42dB), the corresponding dictionary tree is shown in (b).*

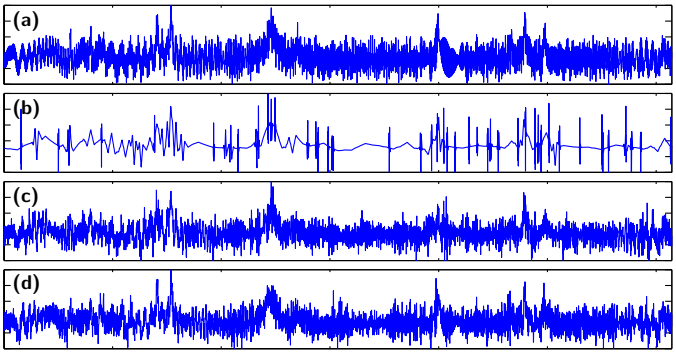

Fig. 7. *(a) original signal, $N = 4096$; (b) recovery using orthogonal wavelets, $n = N/3$ (PSNR=17.59dB) ; (c) recovery using best wavelet packets basis, $n = N/3$ (PSNR=20.46dB) ; (d) recovery using best non-stationary wavelet packets basis, $n = N/3$ (PSNR=21.66dB).*

Quad-tree dyadic subdivision. A bandelet basis $\mathcal{B}(\lambda)$ is parameterized by a quadtree λ that defines a segmentation of the image plane $[0,1]^2 = \bigcup \left\{ S_i^j \setminus p_i^j \in \mathcal{L}(\lambda) \right\}$ where $S_i^j = [i_1 2^{-j}, (i_1+1)2^{-j}] \times [i_2 2^{-j}, (i_2+1)2^{-j}]$ for any index $i = i_1 + 2^j i_2$ with $0 \leqslant i_1, i_2 < 2^j$. Each leave is also tagged with a token $\ell(p_i^j) \in \{0, \ldots, \kappa-1\} \cup \emptyset$. A token $\ell(p_i^j) \neq \emptyset$ defines a direction $\theta(p_i^j) = \pi \ell(p_i^j)/\kappa \in [0, \pi)$ that parameterizes locally the direction of the bandlet basis vectors in S_i^j. Figure 9, (a) shows an example of such a tagged quad-tree. The bandelet transform corresponding to this basis applies independently over each square S_i^j of the image either

– if $\ell(p_i^j) = \emptyset$: a 2D isotropic wavelet transform,
– if $\ell(p_i^j) \neq \emptyset$: a 1D wavelet transform along the direction defined by the angle $\ell(p_i^j)$.

Figure 8 (a,b) shows an example of quad-tree λ together with the corresponding subdivision of $[0,1]^2$.

Fast bandlet transform. We now detail the forward transform algorithm, that applies a local orthogonal transform on each square of the segmentation defined by λ.

– **Segmentation of the pixels.** For each leaf $p_i^j \in \mathcal{L}(\lambda)$, the corresponding square in the segmentation is denoted as S_i^j. The pixel values of f located in S_i^j are stored in f_i^j

$$\forall\, 0 \leqslant k_1, k_2 < 2^j,\ f_i^j[k_1, k_2] = f[k_1 + i_1, k_2 + i_2].$$

for any index $i = i_1 + 2^j i_2$ with $0 \leqslant i_1, i_2 < 2^j$

– **Squares without geometry.** If $\ell(p_i^j) = \emptyset$, a 2D isotropic wavelet transform is applied to $f_i^j \in \mathbb{R}^{N/2^{2j}}$ in order to get the bandlet transform coefficients $\Psi^\lambda(f)_i^j \in \mathbb{R}^{N/2^{2j}}$ over S_i^j.

– **Square with geometry.** If $\ell(p_i^j) \neq \emptyset$, the points $k = (k_1, k_2)$ are projected along the direction $\theta(p_i^j) = \pi \ell(p_i^j)/\kappa$, $\forall\, 0 \leqslant k_1, k_2 < N/2^j$,

$$\pi(k) \overset{\text{def.}}{=} \sin(\theta(p_i^j))k_1 - \cos(\theta(p_i^j))k_2 \in \mathbb{R}.$$

These projected values are sorted

$$\pi(\varphi(0)) \leqslant \pi(\varphi(1)) \leqslant \ldots \leqslant \pi(\varphi(N/2^{2j} - 1)).$$

The one-to-one mapping

$$\varphi : \{0, \ldots, N/2^{2j} - 1\} \to \{0, \ldots, N/2^j - 1\}^2$$

orders the discrete 2D points (k_1, k_2) along a 1D discrete axis. This allows to define a 1D vector $\tilde{f}_i^j$ obtained by reordering the values of f_i^j as

$$\forall\, k \in \{0, \ldots, N/2^{2j} - 1\}, \quad \tilde{f}_i^j[k] \overset{\text{def.}}{=} f_i^j[\varphi(k)].$$

Figure 9 (c) shows an example of such a 1D reordering when the direction $\ell(p_i^j)$ follow closely an edge. In this case the resulting vector $\tilde{f}_i^j$ is regular. The bandlet transform coefficients $\Psi^\lambda(f)_i^j \in \mathbb{R}^{N/2^{2j}}$ over S_i^j are the 1D wavelet transform coefficients of $\tilde{f}_i^j$ in a 1D Haar basis

$$\Psi^\lambda(f)_i^j[s] = \langle \tilde{f}_i^j,\, \beta_s \rangle$$

where $\{\beta_s\}_{s=0}^{N/2^{2j}-1}$ is the 1D orthogonal Haar basis.

– **Output.** The transformed coefficients are

$$\Psi^\lambda(f) \overset{\text{def.}}{=} \left\{ \Psi^\lambda(f)_i^j[s] \setminus p_i^j \in \mathcal{L}(\lambda), 0 \leqslant s < 2^{2j} \right\}.$$

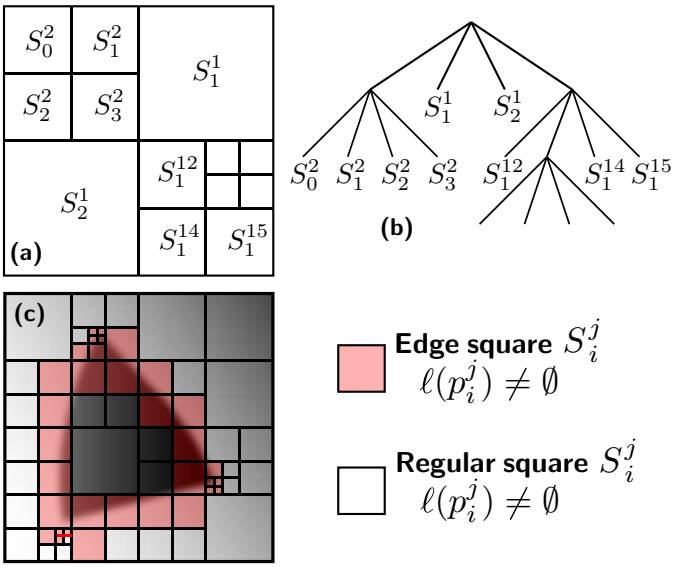

Fig. 8. *(a) example of dyadic subdivision of $[0,1]^2$ in squares S_i^j, (b) corresponding quad-tree λ, (c) example of subdivision $\lambda^\star$ adapted to a geometrically regular function f. The basis $\mathcal{B}(\lambda^\star)$ minimizes $\mathcal{E}(f, \mathcal{B}(\lambda), t)$.*

Both the 2D wavelet transform applied to f_j^i if $\ell(p_i^j) = \emptyset$ and the 1D wavelet transform applied to $\tilde{f}_i^j$ if $\ell(p_i^j) \neq \emptyset$ are orthogonal transforms. The resulting bandlet transform $f \mapsto \Psi^\lambda f$ is thus an orthogonal transform and the coefficients are the decomposition of f on a set of orthogonal bandlet basis vectors that compose the bandlet basis $\mathcal{B}(\lambda)$

$$\forall\, p_i^j \in \mathcal{L}(\lambda),\ \forall\, s < N/2^{2j},\ \Psi^\lambda(f)_i^j[s] = \langle f,\, \psi_{i,s}^j \rangle.$$

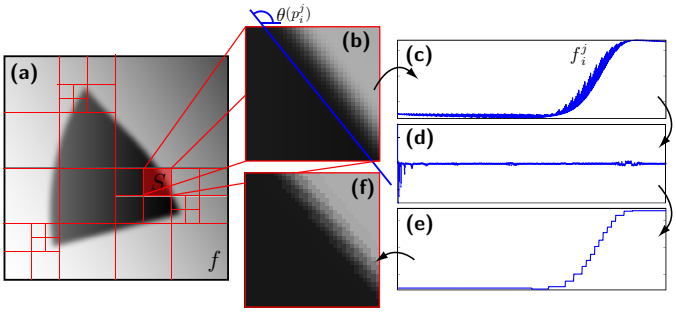

Fig. 9. *(a) a geometric image together with some adapted dyadic segmentation Q ; (b) a square S_i^j together with its adapted direction $\theta(p_i^j)$; (c) the 1D signal f_i^j obtained by mapping the pixels values $f(x^{(i)})$ on a 1D axis ; (d) the 1D Haar coefficients of f_i^j ; (e) the 1D approximation obtained by reconstruction from the 20 largest Haar coefficients ; (f) the corresponding square approximated in bandelet.*

Fast inverse bandlet transform. The backward reconstruction algorithm is obtained by reversing the steps of the bandlet decomposition process.

- **Retrieve the segmented coefficients.** Each dyadic square S_i^j for $p_i^j \in \mathcal{L}(\lambda)$ in the segmentation is treated independently. A set of bandlet coefficients $\Psi^\lambda(f)_i^j$ is associated to each of those leaves.
- **Squares without geometry.** If $\ell(p_i^j) = \emptyset$, a 2D inverse isotropic wavelet transform is applied to $\Psi^\lambda(f)_i^j$ in order to recover f_i^j.
- **Squares with geometry.** If $\ell(p_i^j) = \emptyset$, a 1D inverse Haar wavelet transform is applied to $\Psi^\lambda(f)_i^j[s]$ in order to recover the 1D vector $\tilde{f}_i^j \in \mathbb{R}^{N/2^{2j}}$

$$\forall\, k < N/2^{2j},\ \tilde{f}_i^j[k] = \sum_{s=0}^{N2^{2j}-1} \Psi^\lambda(f)_i^j[s]\, \beta_s[k].$$

The original coefficients are obtained by reordering this 1D vector

$$\forall\, k_1, k_2 < N/2^j,\ f_i^j[k_1, k_2] = \tilde{f}_i^j[\varphi^{-1}(k_1, k_2)].$$

where φ^{-1} is the reverse mapping of φ.

Keeping only a few bandelet coefficients and setting the others to zero performs an approximation of the original image that follows the local direction $\ell(p_i^j)$, see figure 9 (f).

In order to find the best basis $\mathcal{B}(\lambda^\star)$ adapted to some function f as defined in (6), a fast algorithm is introduced in [23]. This algorithm is a particular instance of the dynamic programming algorithm exposed in section IV-B. This algorithm tests, for all the dyadic squares S_i^j, a finite number κ of evenly sampled directions $\theta(p_i^j) \in [0, \pi)$. This fast best basis search defines the quad-tree segmentation λ and the set of directions $\{\ell(p_i^j)\}_{p_i^j \in \mathcal{L}(\lambda)}$ adapted to a given image f that minimizes (6). This process segments the image f into squares S_i^j over which f is smooth, thus setting $\ell(p_i^j) = \emptyset$ and squares containing an edge, where $\ell(p_i^j)$ closely matches the direction of this singularity. Figure 8 (c) shows an example of such a segmentation adapted to a geometric image.

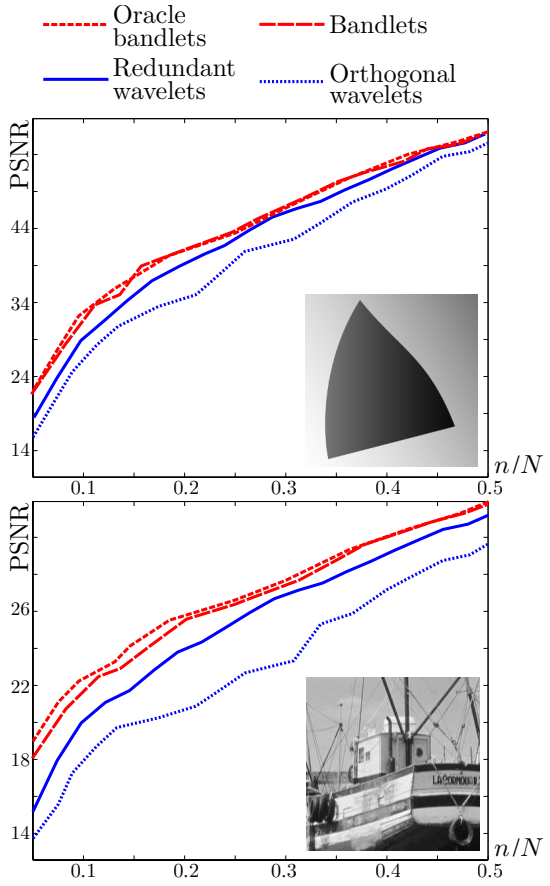

Fig. 10. *Recovery results for various rate of sensing n/N.*

B. Numerical Results

The geometric image f depicted in figure 11 (a), (a) is used to compare the performance of the original compressed sensing algorithm in a wavelet basis to the adaptative algorithm in a best bandlet basis. We use a translation invariant wavelet tight frame, which is more efficient for inverse problems than orthogonal wavelets. Since the wavelet basis is not adapted to the geometric singularities of such an image, reconstruction (b) has ringing artifacts. The adapted reconstruction (c) exhibits fewer such artifacts since the bandlet basis functions are elongated and follow the geometry. The segmentation is depicted after the last iteration, together with the chosen directions $\theta(p_i^j)$ that closely match the direction of the edges of f. Figure 11, (a',b',c') shows the recovery results for a natural image containing complex geometric structures such as edges, junctions and sharp line features. The best bandlet recovery is able to resolve these features efficiently.

Figure 10 shows the recovery error for various sensing rate. For low rate (n/N close to 0), the basis estimated by the iterative algorithm is not as good as the oracle basis estimated from f. For higher rates, the algorithm is able to find the

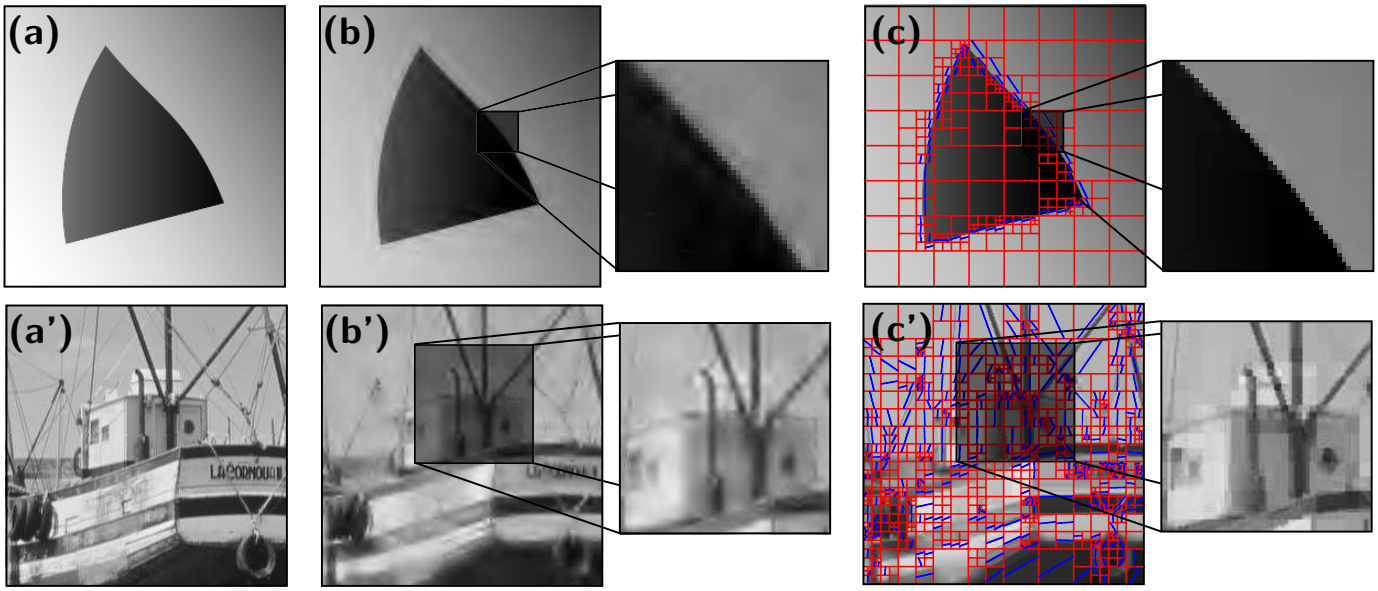

Fig. 11. *(a,a') original image ; (b) compressed sensing reconstruction using the translation invariant wavelet frame, $n = N/6$ (PSNR=37.1dB) ; (c) reconstruction using iteration in a best bandlet basis (PSNR=39.3dB). (b') wavelet frame, PSNR=22.1dB, (c') bandlet basis, PSNR=23.9dB.*

underlying geometry efficiently and the algorithm performs as good as the oracle basis.

VIII. Conclusion

This paper has developed an adaptive reconstruction scheme for compressed sensing. This best basis framework allows to recover signals with complex structures from compressed measurements. This approach is successful for natural sounds and geometric images that contain a broad range of sharp transitions.

References

[1] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[2] ——, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] ——, "For most large underdetermined systems of linear equations, the minimal ℓ_1 norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[5] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," *Proc. of Computational Imaging IV at SPIE Electronic Imaging*, 2006.

[6] D. D. M. Lustig and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1998.

[8] E. Candès and D. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Comm. Pure Appl. Math.*, vol. 57, no. 2, pp. 219–266, 2004.

[9] D. Donoho and M. Elad, "Maximal sparsity representation via l1 minimization," *Proc. Nat. Aca. Sci.*, vol. 100, pp. 2197–2202, 2003.

[10] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *to appear in IEEE Transactions on Information Theory*, 2007.

[11] R. Coifman and V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. IT–38, no. 2, pp. 713–718, Mar. 1992.

[12] A. Cohen and N. Dyn, "Nonstationary subdivision schemes and multiresolution analysis," *SIAM J. Math. Anal.*, vol. 27, no. 6, pp. 1745–1769, 1996.

[13] E. Le Pennec and S. Mallat, "Bandelet Image Approximation and Compression," *SIAM Multiscale Modeling and Simulation*, vol. 4, no. 3, pp. 992–1039, 2005.

[14] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Preprint*, 2004.

[15] E. Candès and J. Romberg, "Practical signal recovery from random projections," *Wavelet Applications in Signal and Image Processing XI, Proc. SPIE Conf. 5914*, 2004.

[16] R. Baraniuk and M. Wakin, "Random projections of smooth manifolds," *to appear in Foundations of Computational Mathematics*, 2007.

[17] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1999.

[18] M. Rudelson and R.Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *to appear in Communications on Pure and Applied Mathematics*, 2007.

[19] E. Candès, "Compressive sampling," *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, 2006.

[20] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, 2005, submitted.

[21] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. on Image Proc.*, vol. 7, pp. 319–355, July 1998.

[22] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math*, vol. 57, pp. 1413–1541, 2004.

[23] S. Mallat and G. Peyré, "Orthogonal bandlets for geometric image approximation," *Preprint*, 2006.

[24] R. R. Coifman, G. Matviyenko, and Y. Meyer, "Modulated Malvar-Wilson bases," *Appl. Comput. Harmon. Anal.*, vol. 4, no. 1, pp. 58–61, 1997.

[25] E. Candès, "Multiscale chirplets and near-optimal recovery of chirps," *Technical Report, Stanford University*, 2002.

[26] G. Peyré and N. Ouarti, "Best basis search in a non-stationary wavelet packets dictionary," *Preprint*, 2007.

[27] L. Breiman, J. H. Friedman, R. A. Olshen, , and C. J. Stone, *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.

[28] D. L. Donoho, "Cart and best-ortho-basis: A connection," *Ann. Stat.*, vol. 25, no. 5, pp. 1870–1911, 1997.

[29] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 423–438, 2005.

[30] G. Peyré and S. Mallat, "Surface compression with geometric bandelets," *ACM Transactions on Graphics, (SIGGRAPH'05)*, vol. 24, no. 3, Aug. 2005.

[31] D. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Annals of Statistics*, vol. 27, no. 3, pp. 859–897, 1999.