Best Basis Compressed Sensing

Gabriel Peyré, *Member, IEEE*

Abstract—This paper proposes a best basis extension of compressed sensing recovery. Instead of regularizing the compressed sensing inverse problem with a sparsity prior in a fixed basis, our framework makes use of sparsity in a tree-structured dictionary of orthogonal bases. A new iterative thresholding algorithm performs both the recovery of the signal and the estimation of the best basis. The resulting reconstruction from compressive measurements optimizes the basis to the structure of the sensed signal. Adaptivity is crucial to capture the regularity of complex natural signals. Numerical experiments on sounds and geometrical images indeed show that this best basis search improves the recovery with respect to fixed sparsity priors.

Index Terms—Compressed sensing, best basis, sparsity, 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 compressive acquisition of data has been proposed jointly by Candès, Tao and Romberg [1], [2] and Donoho [3], [4].

This paper extends the compressed sensing recovery by switching from a fixed orthogonal basis to a tree structured dictionary of orthogonal bases. The adaptivity of a best basis representation increases the sparsity of sounds and geometrical images, which in turns makes the compressed sensing recovery more efficient. The tree structure of the dictionary is used in a fast iterative thresholding algorithm that estimates both the signal or the image to recover and the best basis that optimizes the sparsity of the representation.

A. Compressed Sensing

In a series of papers, Candès, Tao and Romberg [1], [2] and Donoho [3], [4] have proposed the idea of directly acquiring signal in a compressive form. Instead of performing the acquisition with a high sampling rate and then compressing the data in an orthogonal basis, the signal is rather projected on a reduced set of linear vectors. The compressibility of the signal is only exploited during the reconstruction phase, where one uses 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. Duarte et al. [5] detail a single pixel camera that acquires random projections from the visual scene through a digital micromirror array. A similar acquisition

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

strategy can be used in MRI imaging [6] to reduce the acquisition time and increase the spatial resolution.

B. Best Basis Representations

Sparse approximation in orthogonal bases is at the heart of many efficient compression and denoising algorithms. Fixed orthogonal bases are however not flexible enough to capture the complex regularity of sounds or natural images. For instance the orthogonal wavelet transform does not compress efficiently regular edges [7], [8] and a fixed local cosine basis fails to capture transient parts of musical sounds [9].

To improve the sparsity of complicated sounds or images, 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. To enable fast computation, this dictionary is required to have a tree structure, so that the best basis can be optimized using a fast dynamic programming algorithm.

Local cosine bases [9] 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 [10] and non stationary wavelet packets [11], [12].

The set of cartoon images is a simple model that captures the sketch content of natural images [13]. The curvelet frame of Candès and Donoho [7] 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 [8], [14], later refined by Mallat and Peyré [15], [16], provide such a geometric dictionary together with a fast optimization procedure to compute a basis adapted to a given image.

C. Previous Works

Compressed sensing recovery reconstructs a high resolution signal or image from low dimensional measurements. The resolution of this ill-posed linear inverse problem is regularized by introducing non-linear priors. Initial papers [1], [2], [3], [4] give theoritical results about the performance of compressed sensing using sparsity in an orthogonal basis. Increasing the efficiency of the priors to fit the features of sounds and natural images might improve the quality of the recovered signal.

Sparsity priors over redundant dictionaries. Increasing the redundancy of the frame used to sparsify the data might be useful to remove reconstruction artifacts. One can for instance use a frame of translation invariant wavelets which has proven useful for image denoising [9]. Increasing the redundancy can also improve the sparsity of the representation. The curvelet

frame [7] better represents edges in images than wavelets. The resulting lack of orthogonality might however deteriorate the coherence of the basis with the sensing vectors, see [17].

Advanced fixed priors. Other approaches to enhance compressed sensing reconstruction impose further constraints beyond sparsity. For instance, wavelets coefficients can be optimized in a scale-by-scale fashion [18] and positivity or a total variation constraint can be enforced [19].

One can use more advanced signals models not based on sparsity but rather on low dimensional smooth manifolds [20], [21], union of sub-spaces [22], or block-sparsity [23], [24].

Adaptive priors. A fixed prior does not take into account the time-frequency structures of natural sound or the geometry of edges that is different in each image. Adaptive reconstruction can be obtained using an iterative non-local regularization [25], [26]. It is possible to learn an adaptive sparsity prior together with the sensing matrix [27], which extends the initial idea of learning the sensing vectors alone [28].

In these adaptive methods, an iterative algorithm computes both the recovered signal and parameters that control the regularization prior to match the patterns of the signal. In this paper, we also propose to use an adaptive reconstruction method, but within the setting of sparsity in orthogonal bases. The sparsity is increased by selecting an optimized orthogonal basis in a tree structured dictionary.

D. Contributions and Outline of the Paper

Section II recalls the basics of compressed sensing and ℓ^1 reconstruction. Section III details the main contribution of this paper, which the best basis extension of compressed sensing, together with a fast algorithm to perform an approximate minimization. This new framework minimizes an energy on both 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. A fast iterative algorithm is thus derived from a series of surrogate functionals that progressively estimates the best basis. Section IV recalls best basis selection in a tagged tree structured dictionary, that is required to make our best basis compressed sensing fast. Sections V, VI and VII show numerical applications using 1D and 2D dictionaries which show the performance of our scheme on synthetic and natural data.

II. COMPRESSED SENSING

Compressed sensing acquisition and recovery. Compressed sensing acquisition computes a fixed set of n linear measurements of an unknown high resolution signal $f \in \mathbb{R}^N$ with $N \gg n$

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

The price to pay for this compressed sensing strategy is a non-linear reconstruction procedure to recover f from the compressed representation $y = \Phi f$.

The sparsity of f in a given orthogonal basis $\mathcal{B} = \{\psi_m\}_{m=0}^{N-1}$ of $\mathbb{R}^N$ is measured by using the ℓ^0 pseudo norm $\|\Psi f\|_0$, where

$$\Psi f = \{\langle f, \psi_m \rangle\}_m \quad \text{and} \quad \|x\|_0 = \#\{m \setminus x_m \neq 0\}.$$

This sparsity can be used to recover a signal that is a solution of the following minimization

$$\min_{g \in \mathbb{R}^N} \|\Psi g\|_0 \qquad \text{subject to} \qquad \Phi g = y. \qquad (1)$$

The minimization (1) is however combinatorial and thus intractable. It is relaxed by using the ℓ^1 norm $\|\Psi f\|_1$ of the coefficients of f in $\mathcal{B}$

$$\|x\|_1 = \sum_m |x_m|.$$

The recovered signal $f^\star$ is a solution of the following convex problem

$$f^\star \in \operatorname*{argmin}_{g \in \mathbb{R}^N} \|\Psi g\|_1 \qquad \text{subject to} \qquad \Phi g = y. \qquad (2)$$

Sparsity and incoherence. Compressed sensing theory requires two constraints for this ℓ^1 recovery to be efficient:

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

$$\|\Psi f\|_0 \leqslant s. \qquad (3)$$

The theory extends to signals that are well approximated with a signal that is s-sparse in $\mathcal{B}$.

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

$$(1 - \delta)\|x\|^2 \leqslant \|\Phi\Psi x\|^2 \leqslant (1 + \delta)\|x\|^2. \qquad (4)$$

for any s-sparse signal $\|x\|_0 \leqslant s$.

One should note that although the sparity assumption (3) constrains the ℓ^0 pseudo-norm of Ψf, the actual recovery process (2) 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 [29].

The following recovery theorem ensures the perfect recovery using the ℓ^1 minimization.

Theorem 1. *([1], [2], [30]) If f is s-sparse in Ψ as defined in (3) and if the sensing matrix Φ satisfies $\delta_{2s} < \sqrt{2} - 1$, the solution of (2) satisfies $f^\star = f$.*

This theorem ensures that there is an unique solution to (1) and that it coincides with the solution of (2). It extends to the recovery of an approximately sparse signal, in which case $\|f - f^\star\|$ is comparable with the best s-terms approximation of f in the basis $\mathcal{B}$.

Random sensing matrices. The allowable sparsity s for which $\delta_{2s} < \sqrt{2} - 1$ needs to be large enough for Theorem 1 to be useful. For several random matrix ensembles, this is actually the case with high probability on the random matrix. For instance if the entries of Φ are drawn independently from a Gaussian distribution of variance $1/N$, $s \leqslant Cn/\log(N/n)$ ensures $\delta_{2s} < \sqrt{2} - 1$ with high probability. This result extends to other kinds of random distributions, see for instance [1], [2], [31], [3], [4], [32].

In this paper, following for instance [33], we consider a fast sampling operator

$$\Phi f = (P_1 H P_2 f) \downarrow_n,$$

where P_1 and P_2 are realizations of a random permutation of the N entries of a vector in $\mathbb{R}^N$, H is a 1D or 2D orthogonal Hadamard transform, and $\downarrow_n$ selects the n first entries of a vector, see [34] for a definition of the Hadamard transform and its fast implementation. Such a random sensing operator is computed in $O(N \log(N))$ operations, which is important to process high dimensional data.

Robust Compressed Sensing. To deal with noisy measurements $y = \Phi f + w$, where w is a bounded noise $\|w\| \leqslant \sigma$, one can turn the constrained formulation (2) into a penalized variational problem

$$f^\star = \underset{g \in \mathbb{R}^N}{\arg\min}\, E(g, \mathcal{B}, t) \quad \text{where} \tag{5}$$

$$E(g, \mathcal{B}, t) = \frac{1}{2}\|\Phi g - y\|^2 + t\|\Psi g\|_1. \tag{6}$$

where the Lagrange multiplier t should be set so that $\|\Phi f^\star - y\| \approx \sigma$. Compressed sensing theory extends to this noisy setting, since the recovery error $\|f - f^\star\|$ is of the order of the noise level σ, see [35], [4].

III. Best Basis Compressed Sensing

Fixed bases are not efficient enough to sparsify sounds and natural images, and more redundancy is required. This section extends the recovery process (5) to a dictionary 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 $\mathbb{R}^N$. Instead of using an a priori fixed basis such as the wavelet or Fourier basis, one chooses a parameter $\lambda^\star \in \Lambda$ adapted to the structures of the signal to process and use the optimized basis $\mathcal{B}^{\lambda^\star}$.

This dictionary $\mathcal{D}_\Lambda$ defines a highly redundant set of atoms

$$D_\Lambda = \left\{ \psi_m^\lambda \setminus \lambda \in \Lambda, m \right\}.$$

One could use directly D_Λ to reconstruct a signal from compressive measurements, but this is numerically intractable for large signals or images. Restricting the sparsity to orthogonal bases leads to fast algorithms for structured dictionaries such as those considered in Section IV.

Tree structure. To enable the fast optimization of a parameter $\lambda^\star$ adapted to a given signal or image f to process, we impose that each $\lambda \in \Lambda$ is a tree, that is a subset of the larger tree Λ. Typical examples use the tree λ to index a segmentation of either space (local cosine bases detailed in Section V) or scale (wavelet packets detailed in Section VI). Each node in a tree λ corresponds to a refinement in the segmentation.

Lagrangian and adapted basis. A best basis $\mathcal{B}^{\lambda^\star}$ adapted to a signal or an image $f \in \mathbb{R}^N$ is optimized to obtain the best possible approximation of f for a given sparsity, as measured by the ℓ^1 norm. This constrained formulation is turned into an Lagrangian unconstrained optimization problem by minimizing a Lagrangian $\mathcal{E}$

$$\lambda^\star = \underset{\lambda \in \Lambda}{\arg\min}\, \mathcal{E}(f, \mathcal{B}^\lambda, t) \tag{7}$$

$$\mathcal{E}(f, \mathcal{B}^\lambda, t) = \min_{g \in \mathbb{R}^N} \frac{1}{2}\|f - g\|^2 + t\|\Psi^\lambda g\|_1$$

and where $\Psi^\lambda g = \{\langle g, \psi_m^\lambda \rangle\}_{m=0}^{N-1}$. The variable t is a Lagrange multiplier that weights the quality of approximation in the chosen basis with the sparsity of the expansion.

The following lemma characterizes the best basis together with the best ℓ^1-penalized approximation in this basis. Its proof can be found for instance in [9].

Lemma 1. *The minimizer*

$$(f^\star, \lambda^\star) = \underset{(g, \lambda) \in \in \mathbb{R}^N \times \Lambda}{\arg\min} \frac{1}{2}\|f - g\|^2 + t\|\Psi^\lambda g\|_1$$

is given by
$$\begin{cases} \lambda^\star = \underset{\lambda \in \Lambda}{\arg\min}\, \mathcal{E}(f, \mathcal{B}^\lambda, t), \\ f^\star = S_t(f, \mathcal{B}^{\lambda^\star}), \end{cases} \tag{8}$$

where the soft thresholding operator is defined as

$$S_t(f, \mathcal{B}) = \sum_m s_t(\langle f, \psi_m \rangle)\, \psi_m \tag{9}$$

$$\text{where} \quad s_t(x) = \max(0, 1 - t/|x|)x$$

The soft thresholding S_t defined in (9) computes the best ℓ^1 approximation of a signal 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 $\mathcal{B}^\lambda$

$$\mathcal{E}(f, \mathcal{B}^\lambda, t) = \sum_m \gamma_t(|\langle f, \psi_m^\lambda \rangle|) \tag{10}$$

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

B. Best Basis Compressed Sensing Reconstruction

Best basis regularization. Compressed sensing 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 (5) is extended to this new setting by performing the minimization both on the signal to recover and on the basis

$$(f^\star, \lambda^\star) = \underset{(g, \lambda) \in \mathbb{R}^N \times \Lambda}{\arg\min} E(g, \mathcal{B}^\lambda, t), \tag{11}$$

where the energy E is defined in equation (5).

Informal discussion about best basis regularization. For several models of sounds and natural images, best basis methods have been shown either experimentally or theoretically to outperform fixed representations to sparsify the data. Although for the moment no theory is able to assess the performance of an adaptive minimization such as (11), Sections V, VI and VII show numerically that the increase of sparsity obtained with best basis computations translates to an increase of performances for the regularization of compressed sensing reconstruction.

We believe that the effectiveness of best basis regularization for compressed sensing comes in large part from the randomness of the measurements, which translates, as already noticed by Donoho et al. [33], on reconstruction artifacts being close to Gaussian white noise during an iterative thresholding reconstruction method. This key feature is very different from the kind of artifacts obtained for non-random sensing schemes, such as Radon transform in tomography or low pass filtering in optical imaging system. This results in a high incoherence between the residual noise with the sparsity vectors of $\mathcal{D}_\Lambda$ during our reconstruction scheme. This incoherence allows us to efficiently estimate an adapted basis with an iterative scheme without prior knowledge about the signal to reconstruct.

C. Surrogate Functionals

Searching in the whole dictionary $\mathcal{D}_\Lambda$ for the best basis parameter $\lambda^\star \in \Lambda$ that minimizes (11) 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 this issue, we relax the energy minimization (11) and define 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

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

Such a surrogate functional is introduced by Daubechies et al. [36] to derive an iterative thresholding algorithm that solves the ℓ^1 regularization of inverse problems.

As long as $\mu > \|\Phi^*\Phi\|$, the surrogate energy E_h is a smooth, strictly convex, modification of the original energy E. Furthermore,

$$E_h \geqslant E \quad \text{and} \quad E_h(h, \mathcal{B}^\lambda, t) = E(h, \mathcal{B}^\lambda, t).$$

It is thus reasonable to use E_h as a proxy for the original minimization (11), which leads us to consider, for a fixed h,

$$(f^\star(h), \lambda^\star(h)) = \operatorname*{argmin}_{(g,\lambda) \in \mathbb{R}^N \times \Lambda} E_h(g, \mathcal{B}^\lambda, t). \quad (12)$$

The following proposition shows that this surrogate minimization diminishes the energy E.

Proposition 1. If $\mu > \|\Phi^*\Phi\|$, one has for all λ

$$E(f^\star(h), \mathcal{B}^{\lambda^\star(h)}, t) \leqslant E(h, \mathcal{B}^\lambda, t).$$

Proof: Condition $\mu > \|\Phi^*\Phi\|$ ensures that

$$\frac{\mu}{2}\|g - h\|^2 - \frac{1}{2}\|\Phi g - \Phi h\|^2 \geqslant 0$$

so that

$$E_h(g, \mathcal{B}^\lambda, t) \geqslant E(g, \mathcal{B}^\lambda, t).$$

One thus has

$$E(f^\star(h), \mathcal{B}^{\lambda^\star(h)}, t) \leqslant E_h(f^\star(h), \mathcal{B}^{\lambda^\star(h)}, t)$$
$$\leqslant E_h(h, \mathcal{B}^\lambda, t) = E(h, \mathcal{B}^\lambda, t).$$

∎

The following theorem ensures that the minimization (12) is easily solved using a best basis search and a soft thresholding.

Theorem 2. *The minimization* (12) *has a global minimum which is given by*

$$\begin{cases} \lambda^\star(h) = \operatorname*{argmin}_{\lambda \in \Lambda} \mathcal{E}(\tilde{h}, \mathcal{B}^\lambda, t), \\ f^\star(h) = S_{t/\mu}(\tilde{h}, \mathcal{B}^{\lambda^\star(h)}), \end{cases} \quad (13)$$

$$\text{where} \quad \tilde{h} = h + \frac{1}{\mu}\Phi^{\mathrm{T}}(y - \Phi h),$$

where the Lagrangian $\mathcal{E}$ is defined in equation (7) *and the soft thresholding operator $S_{t/\mu}$ is defined in equation* (9).

Proof: The energy E_h is expanded as follow

$$E_h(g, \mathcal{B}^\lambda, t) = \frac{\mu}{2}\|g\|^2 - \langle \Phi g, y \rangle + \langle \Phi g, \Phi h \rangle - \mu\langle g, h \rangle$$
$$+ t\|\Psi^\lambda g\|_1 + C_{h,y}$$
$$= \frac{\mu}{2}\|g\|^2 - \langle g, \mu h + \Phi^{\mathrm{T}}(y - \Phi h)\rangle$$
$$+ t\|\Psi^\lambda g\|_1 + C_{h,y}$$

where $C_{h,y}$ is independent of λ and g. Up to multiplicative and additive constants, one has

$$E_h(g, \mathcal{B}^\lambda, t) \propto \frac{1}{2}\|h + \frac{1}{\mu}\Phi^{\mathrm{T}}(y - \Phi h) - g\|^2 + t\|\Psi^\lambda g\|_1.$$

The result of the theorem follows from lemma 1. ∎

Equation (13) shows that the parameter $\lambda^\star(h)$ is the best basis parameter of the modified guess $\tilde{h}$. It also shows that $f^\star(h)$ is obtained by thresholding $\tilde{h}$ in that basis. The best parameter $\lambda^\star(h)$ is thus found by optimizing a Lagrangian $\mathcal{E}$, which can be achieved with a fast algorithm, as explained in Section IV.

D. Best Basis Compressed Sensing Algorithm

The minimization of $E(g, \mathcal{B}^\lambda, t)$ over (g, λ) is replaced by the minimization of a set of surrogate functionals $E_{f_k}(g, \mathcal{B}^\lambda, t)$ where f_k is the current estimate of the solution at iteration k. Starting from an initial pair (f_0, λ_0), this defines

$$(f_{k+1}, \lambda_{k+1}) = \operatorname*{argmin}_{(g,\lambda) \in \mathbb{R}^N \times \Lambda} E_{f_k}(g, \mathcal{B}^\lambda, t). \quad (14)$$

Algorithm 1 details the step by step implementation of (14), that is derived from (14) using Theorem 2. The steps of the algorithm are repeated until a user defined tolerance η is reached. In the noiseless setting $y = \Phi f$, one can still use this algorithm and decay the value of $t = t_k$ toward zero during the iterations, in a fashion similar to the Morphological Components Algorithm (MCA) algorithm [37].

This algorithm extends to a best basis setting the iterative thresholding algorithm that has been proposed by several authors to solve inverse problems such as (5), see for instance [36], [38] and the references therein.

Convergence of the the algorithm. For a single basis $\mathcal{B}$, condition $\mu > \|\Phi^*\Phi\|$ ensures that this iterative thresholding converges to a solution of (5), see [36]. This condition can in fact be relaxed to $\mu > \|\Phi^*\Phi\|/2$, see [38]. In contrast to the

---

Algorithm 1: Best-basis compressed sensing algorithm.

Initialization.: Set $k = 0$, $f_0 = 0$.

repeat

> **Update the estimate:** Set
>
> $$\tilde{f}_k = f_k + \frac{1}{\mu}\Phi^{\mathrm{T}}(y - \Phi f_k). \qquad (15)$$
>
> **Update best basis:** Compute the best basis parameter
>
> $$\lambda_{k+1} = \operatorname*{argmin}_{\lambda \in \Lambda} \mathcal{E}(\tilde{f}_k, \mathcal{B}^\lambda, t). \qquad (16)$$
>
> This minimization is carried out with a fast procedure, as detailed in section IV.
>
> **Denoise the estimate:** Compute
>
> $$f_{k+1} = S_{t/\mu}(\tilde{f}_k, \mathcal{B}^{\lambda_{k+1}}),$$
>
> where S_t is the threshold operator defined in (9).

until $\|f_{k+1} - f_k\| < \eta$;

---

fixed basis setting, the iterations of (14) that require a best basis search are difficult to analyze. Proposition 1 shows that the energy $E(f_k, \mathcal{B}^{\lambda_k}, t)$ is decaying through the iterations and thus converges to some limit value. Since this energy is non-convex in (f, λ), nothing more can however be said about the convergence of the iterates f_k.

In theory, nothing prevents the best basis parameter λ_k to cycle through a series of limit bases, but this was never the case in the numerical experiments performed in Sections V-B, VI-B and VII-B. We always observed that λ_k converged to its final value $\lambda^\star$ after a small number of iterations, that is of the order of 10 to 20 for natural sounds and images. Once this final regime is attained, one can use classical results about the convergence of iterative thresholding algorithms, that show for instance that the energy $E(f_k, \mathcal{B}^{\lambda^\star}, t)$ is decaying at speed $O(1/k)$ through the iterations, see [38].

We note that since the objective to minimize is non-convex and there is no convergence guarantee, a different initialization f_0 might lead to a different sequence of bases $\mathcal{B}^{\lambda_k}$, and hence a different final result. Although this is indeed the case, we did not observed in our numerical tests significant improvement of using alternate initialization strategies with respect to setting $f_0 = 0$.

IV. BEST BASIS COMPUTATION

This section reviews a best basis framework common to the dictionaries used in the numerical experiments of this paper. These dictionaries enjoy a hierarchical tree structure, which makes the computation of the best basis in (16) fast. This framework does not cover the special case of non-stationary wavelet packets, detailed in Section VI that requires a more advanced procedure, see [12].

A. Tree structured dictionaries

Structure of the atoms. This paper focusses on dictionaries $\mathcal{D}_\Lambda$ having a multiscale tree structure. The atoms of $\mathcal{D}_\Lambda$ span sub-spaces $V_{j,i}$ of $\mathbb{R}^N$ for scales $0 \leqslant j \leqslant J = \log_2(N)/d$

and position $0 \leqslant i < 2^{dj}$, that obey a refinement relationship

$$\forall j < J, \quad V_{j,i} = \bigoplus_{\varepsilon=0}^{2^d - 1} V_{j+1, 2^d i + \varepsilon},$$

the sum being orthogonal.

Each sub-space $V_{j,i}$ has dimension $N/2^{dj}$ and is equipped with one or several orthogonal bases indexed by a token $\ell \in \Omega$

$$\mathcal{B}_{j,i}^\ell = \left\{ \psi_{j,i,s}^\ell \setminus \forall 0 \leqslant s < N/2^{dj} \right\}. \qquad (17)$$

Structure of an orthogonal basis. The parameter λ that indexes a basis $\mathcal{B}^\lambda \in \mathcal{D}_\Lambda$ is a binary tree (for 1D signals, where $d = 1$) or a quad-tree (for 2D, where $d = 2$). The set of nodes of λ is denoted as $\mathcal{N}(\lambda)$ and each node $(j, i) \in \mathcal{N}(\lambda)$ is located at some level $0 \leqslant j \leqslant J$ and position $0 \leqslant i < 2^{dj}$. A node (j, i) is thus located in the jth row and the ith column of the tree. Each interior node $(j, i) \in \mathcal{I}(\lambda) \subset \mathcal{N}(\lambda)$ has 2^d children $\{(j+1, 2^d i), \ldots, (j+1, 2^d(i+1) - 1)\}$. The leaves nodes $(j, i) \in \mathcal{L}(\lambda)$ have no child.

A basis $\mathcal{B}^\lambda$ is obtained by aggregating bases $\mathcal{B}_{j,i}^\ell$ defined in (17) for (j, i) that are leaves of λ, and for a specific choice $\ell = \ell_{j,i} \in \Omega$ of token at each node of the tree

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

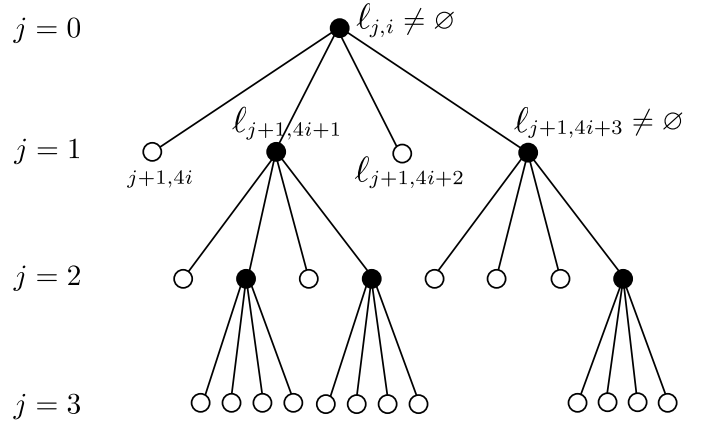Figure 1 shows an example of such a tagged quad-tree λ.

Fig. 1. *Examples of quad-tree λ.*

When one does not care about the location of the basis elements in the tree, the basis is written as $\mathcal{B}^\lambda = \{\psi_m^\lambda\}_m$ where the index is $m = (j, i, s, \ell_{j,i})$ with $(j, i) \in \mathcal{L}(\lambda)$ and $0 \leqslant s < N/2^{dj}$.

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

− **The local cosine dictionary** [10], [9] 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 $|\Omega| = 1$ since there is no need for additional information beside the spatial segmentation.

− **The non-stationary wavelet packets dictionary** [11], [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, $\ell_{j,i} \in \Omega = \{0, \dots, \kappa - 1\}$ indicates the index of a wavelet filter that is used to subdivide the frequency axis. The particular case of wavelet packets [10] is obtained for $\kappa = 1$.

− **The bandlet dictionary** [8], [15] 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 $S_{j,i} \subset [0,1]^2$ of size $2^{-j} \times 2^{-j}$. A leaf node $(j, i) \in \mathcal{L}(\lambda)$ defines the local bandlet transform and caries a geometrical information token $\ell_{j,i} \in \Omega = \{0, \dots, \kappa - 1\} \cup \emptyset$. A token $\ell_{j,i} \neq \emptyset$ is defined over a geometrical square $S_{j,i}$ in which the bandlet vectors are elongated and follow the angle $\pi \ell_{j,i} / \kappa$ that approximates the orientation of the closest edge. A token $\ell_{j,i} = \emptyset$ is defined in isotropic square $S_{j,i}$ that corresponds to regions where no edge is present.

Other classes of dictionaries include bases composed of atoms with rapidly varying oscillations such as the modulated bases of Coifman et al. [39] and the chirplets dictionary of Candès [40].

B. Dynamic programming for Best-basis Computation

A fast best basis search algorithm makes use of the tree structure of $\mathcal{D}_\Lambda$ and the fact that the Lagrangian $\mathcal{E}(f, \mathcal{B}^\lambda, t)$ to minimize is split as a sum over the coefficients of the decomposition (10). It was originally presented by Coifman et al. [10] and is a particular instance of the Classification and Regression Tree (CART) algorithm of Breidman et al. [41] as explained by Donoho [42].

Algorithm 2 details the implementation of this fast best basis search. This algorithm does not work for non-stationary wavelet packets, which require a more complex optimization procedure described in [12].

This algorithm requires the decomposition of f onto each atom $\psi_{j,i,s}^\ell$ for all values of (i, j, s, ℓ), and the computation of Lagrangians $\mathcal{E}_{j,i}$ that depends on these inner products. For several multiscale dictionaries, such as those considered in this paper, fast algorithms perform this computation $O(P)$ operations, where P is the total number of atoms in $\mathcal{D}_\Lambda$. The resulting complexity is thus $O(|\Omega| N \log_2(N))$, where $|\Omega| = 1$ for the local cosine and stationary wavelet packets, and $|\Omega|$ is the number of orientations in the bandlet 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 signal f is 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 (2).

Algorithm 2: Best-basis selection algorithm.

for $0 \leqslant j \leqslant J,\ 0 \leqslant i < 2^{jd} - 1,\ \ell \in \Omega$ **do**
  Compute $\langle f, \psi_{j,i,s}^\ell \rangle$.

for $0 \leqslant j \leqslant J,\ 0 \leqslant i < 2^{jd} - 1$ **do**
  Compute $\ell_{j,i} = \underset{\ell \in \Omega}{\operatorname{argmin}} \sum_s \gamma_t(|\langle f, \psi_{j,i,s}^\ell \rangle|)$.

for $j = J, \dots, 0$ **do**
  for $i = 0, \dots, 2^{jd} - 1$ **do**
    Compute

$$\mathcal{E}_{j,i} = \sum_s \gamma_t(|\langle f, \psi_{j,i,s}^{\ell_{j,i}} \rangle|),$$

$$\tilde{\mathcal{E}}_{j,i} = \sum_{\varepsilon=0}^{2^d - 1} \tilde{\mathcal{E}}_{j+1, 2^d i + \varepsilon},$$

with the convention that $\tilde{\mathcal{E}}_{j,i} = +\infty$ if $j = J$.
    if $\tilde{\mathcal{E}}_{j,i} \leqslant \mathcal{E}_{j,i}$ **then**
      Declare (j, i) as interior $(j, i) \in \mathcal{I}(\lambda)$.
    if $\tilde{\mathcal{E}}_{j,i} > \mathcal{E}_{j,i}$ **then**
      Declare (j, i) as leaf $(j, i) \in \mathcal{L}(\lambda)$.
    Update the cumulative Lagrangian

$$\tilde{\mathcal{E}}_{j,i} \leftarrow \min(\tilde{\mathcal{E}}_{j,i}, \mathcal{E}_{j,i}).$$

− 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 the knowledge of this basis is not available in practice.

− Recovery using the algorithm of section III-D, that estimates iteratively the best basis.

Experiments are performed on noisy measurements $y = \Phi f_0 + w$ where w is a Gaussian white noise of variance σ^2. The noise level σ is supposed to be known, and the regularization parameter t is adjusted so that the residual final error satisfies $\|y - \Phi f^\star\| = \sqrt{n}\sigma$. For noiseless experiments where $\sigma = 0$, we still use our iterative thresholding algorithms, but decay linearly the threshold $t = t_k$ to 0 during the iterations, which is similar to the approach used in the Morphological Component Algorithm of Starck et al. [37]. We used a relaxation parameter $\mu = 1$ for the iteration (15).

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 [10], [9]. Each leaf node $(j, i) \in \mathcal{L}(\lambda)$ corresponds to a selected interval $[x_{j,i}, x_{j,i+1}]$, where $x_{j,i} = 2^{-j} N i - 1/2$. For each of these leave node, the local cosine basis vectors are defined as

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

where b is a smooth windowing function that satisfies some compatibility conditions [9].

The decomposition of a signal f on the vectors of some basis $\mathcal{B}^\lambda$ is computed in $O(N\log(N))$ using fast Fourier transforms. A best basis $\mathcal{B}^{\lambda^\star}$ that minimizes (7) is computed in $O(N\log(N)^2)$ operations using Algorithm 2. Figure 2 shows some examples of basis vectors.
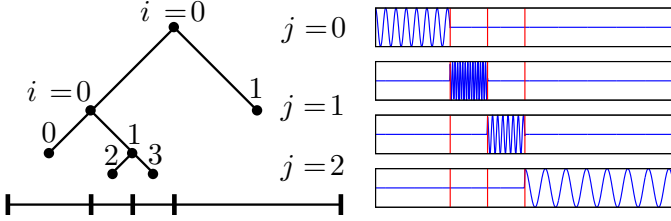

Fig. 2. *A dyadic tree λ defining a spatial segmentation (left) ; some local cosine basis functions $\psi_{j,i,s}$ of the basis $\mathcal{B}^\lambda$ (right).*

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 $\|h\|_0 = \#\{k \setminus h(k) \neq 0\} = 30$, see figure 3, (a). We use here noiseless measurements $y = \Phi f$. The signal recovered by the non-adaptative algorithm of section III-D 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 $\|\Psi^{\lambda_0}f\|_0 = 512$ and $\|\Psi^{\lambda_0}f\|_1 \approx 2.8\|\Psi^\lambda f\|_1$. During the iterations of the algorithm presented in subsection III-C, the estimated best basis $\mathcal{B}^{\lambda_k}$ evolves in order to match the best basis $\mathcal{B}^\lambda$, see figure 3, (c1–c3). The recovered signal $f^\star$ in (c3) is nearly identical to f. We note however that the segmentation $\lambda^\star$ optimized by our algorithm differs slightly from the original segmentation λ_0. In particular one can notice an over-segmentation of the leftmost interval, where the recovered signal $f^\star$ is small but not vanishing, on the contrary to the original signal 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-C. We use here noisy measurements $y = \Phi f + w$, with $\|w\| = 0.03\|\Phi f\|$. 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.

VI. BEST NON-STATIONARY WAVELET PACKET COMPRESSED SENSING

A. Adapted Non-stationary Wavelet Packet Transform

The non stationary (NS) wavelet transform and its extension to wavelet packets was introduced by Cohen et al. [11]. 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. This tree structured dictionary is more general than the trees considered in Section IV-A because a token $\ell_{j,i}$ is associated to each node $(j,i) \in \mathcal{I}(\lambda) \cup \mathcal{L}(\lambda)$ and not only to leaves $(j,i) \in \mathcal{L}(\lambda)$. The
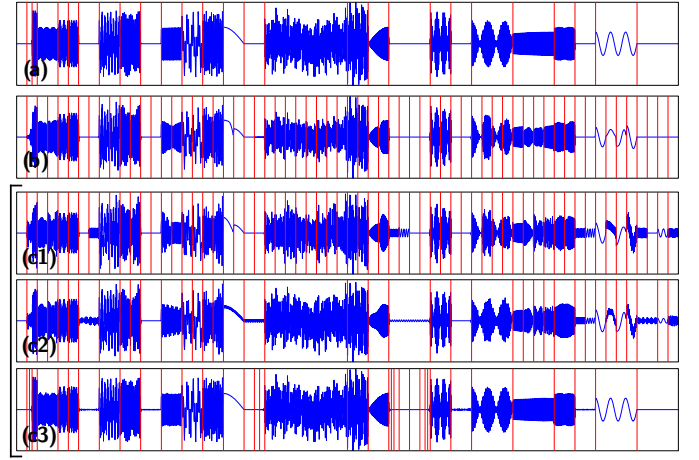

Fig. 3. *(a) synthetic sound signal with 30 random cosine atoms $N = 4096$; (b) recovery using a fixed cosine basis ; (c1) first iteration of the best basis recovery algorithm, $n = N/3$; (c2) iteration $s = 5$; (c3) iteration $s = 20$.*
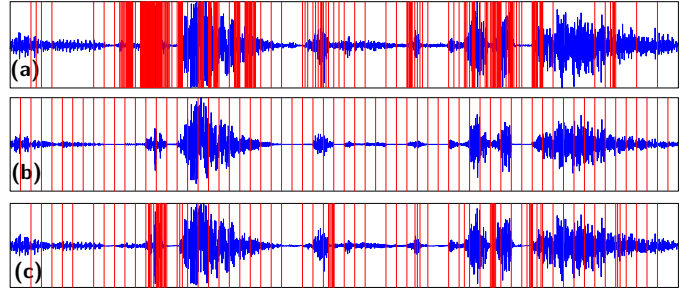

Fig. 4. *(a) sound signal of a tiger howling, together with the best spatial segmentation, $N = 32768$; (b) recovery using fixed local cosine basis, $n = N/3$ (PSNR=19.24dB) ; (c) recovery using best cosine basis, $n = N/3$ (PSNR=21.32dB)*
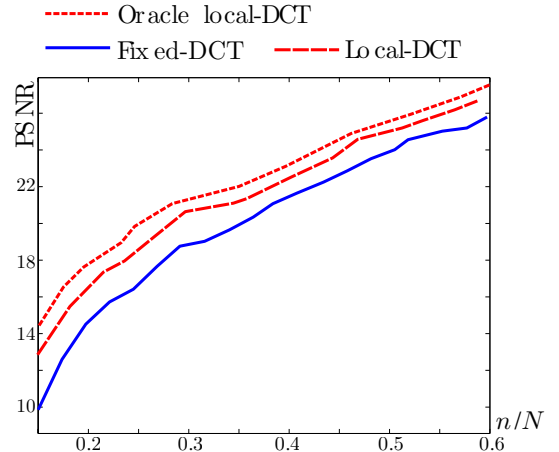

Fig. 5. *Recovery results for the signal of figure 4 for various rate of sensing n/N, in the noiseless setting $y = \Phi f$.*

corresponding best basis algorithm is thus a generalization of Algorithm 2 that is detailed in [12].

The NS wavelet packet dictionary depends on the choice of a set $H = \{h_0, \ldots, h_{\kappa-1}\}$ of low pass quadrature mirror filters, to which is associated the high pass filters $g_\ell[k] = (-1)^k h_\ell[1-k]$.

NS Wavelet Packet Transform. The decomposition of $f = f_{0,0} \in \mathbb{R}^N$ in a NS wavelet packet basis $\mathcal{B}^\lambda$ iteratively computes $f_{j,i} \in \mathbb{R}^{N/2^j}$ by traversing the nodes (j,i) of the

tree λ from top to bottom and computing

$$\begin{cases} f_{j+1,2i} = (f_{j,i} * h_{\ell_{j,i}}) \downarrow 2, \\ f_{j+1,2i+1} = (f_{j,i} * g_{\ell_{j,i}}) \downarrow 2, \end{cases} \quad (18)$$

where $\downarrow 2$ is the sub-sampling by two operator. The coefficients computed on the leaves correspond to the projection of f on the vectors of $\mathcal{B}^\lambda$

$$\forall (j,i) \in \mathcal{L}(\lambda), \quad f_{j,i}[s] = \langle f, \psi^\lambda_{j,i,s} \rangle.$$

This forward NS wavelet packet transform is computed in at most $O(N \log_2(N))$ operations for a signal $f \in \mathbb{R}^N$. A best basis $\mathcal{B}^{\lambda^*}$ defined by (7) adapted to some signal f is computed using a fast best-basis search introduced by Peyré and Ouarti [12]. The complexity algorithm is $O(N \log_2(N))$ for $\kappa = 1$ (which corresponds to the traditional wavelet packets) and $O(N^{1+\log_2(\kappa)})$ for $\kappa > 1$.

B. Numerical Results

In our tests, the non-stationary wavelet packet dictionary is built using a family $H_D = \{h_\ell\}_{\ell=0}^{\kappa-1}$ 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. We use here noiseless measurements $y = \Phi f$. Figure 6 (a) shows the index $\ell_{j,i}$ 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.
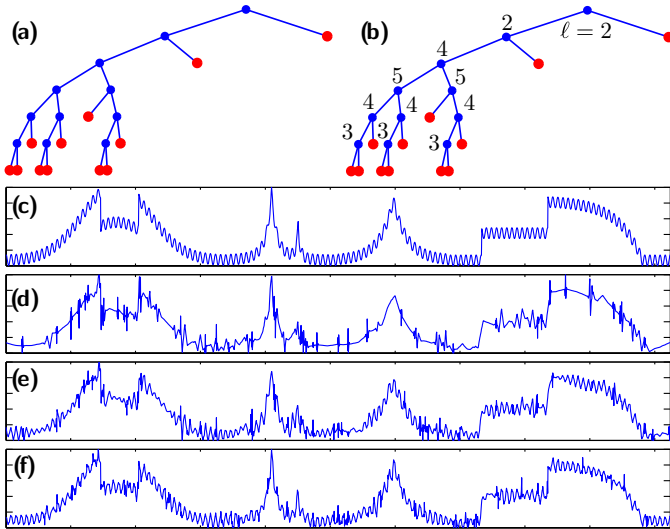

Fig. 6. *(c) original signal, $N = 4096$; (d) recovery using orthogonal wavelets, $n = N/3$ (PSNR=23.77dB) ; (e) recovery using best wavelet packets basis, $n = N/3$ (PSNR=26.50dB), the corresponding dictionary tree is shown in (a) ; (f) recovery using best non-stationary wavelet packets basis, $n =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

Figure 7 shows another example of recovery using wavelets, the wavelet packets and the NS wavelet packets dictionaries.

The signal f has dimension $N = 4096$ and is the superposition of chirps and spikes

$$f(x) = \sum_{k=1}^{3} \cos(\omega_k (x/N)^k) + \sum_{k=0}^{10} \mu_k (1 + |(x/N) - x_k|/\rho_k)^{-4}$$

for $0 \leqslant x < N$, where $\omega_1 = \pi/2N, \omega_2 = \pi N$ and $\omega_3 = 0.7\pi N$, where the $x_k \in [0,1]$ are random position, $\mu_k \in [2,5]$ are random weights and $\rho_k \in [0, 0.03]$ are random widths. We use here noisy measurements $y = \Phi f + w$ with $\|w\| = 0.03\|\Phi f\|$. The best NS wavelet packets basis is able to better recover both the high frequency oscillations of the chirps and the localized spikes than fixed representations.
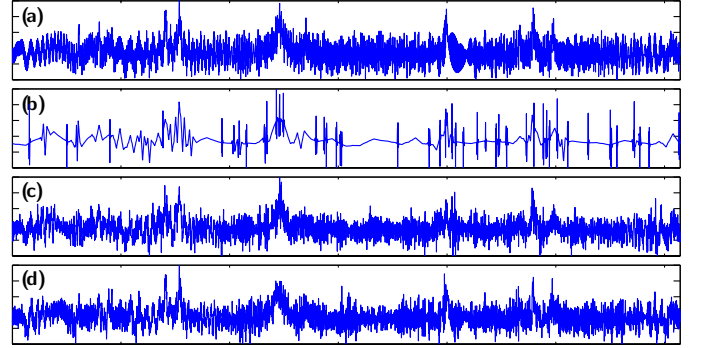

Fig. 7. *(a) original signal, $N = 4096$; (b) recovery using orthogonal wavelets, $n = N/3$ (PSNR=17.59dB) ; (c) recovery using best wavelet packets basis, $n = N/3$ (PSNR=20.46dB) ; (d) recovery using best non-stationary wavelet packets basis, $n = N/3$ (PSNR=21.66dB).*

VII. BEST BANDLET BASIS COMPRESSED SENSING

A. Adapted Bandlet Transform

The bandlet bases dictionary was introduced by Le Pennec and Mallat [8], [14]. Bandlets perform an efficient adaptive approximation of images with geometric singularities, such as the cartoon image in Figure 10. This transform has been refined by Mallat and Peyré [15], [16] to obtain a dictionary of regular and orthogonal basis functions. We present a simplified bandlet transform inspired from [16]. This implementation results in a decomposition similar to the wedgelets of Donoho [13] but within the framework of a dictionary of orthogonal bases.

The quadtree λ that parametrizes the bandlet basis $\mathcal{B}^\lambda$ defines a segmentation

$$[0,1]^2 = \bigcup_{(j,i) \in \mathcal{L}(\lambda)} S_{j,i}$$

where each $S_{j,i}$ is a square of size $2^{-j} \times 2^{-j}$. Figure 8 shows an example of such a dyadic subdivision.

The sub-image $\tilde{f}_{j,i} = \{f(x)\}_{x \in S_{j,i}} \in \mathbb{R}^{N/2^{2j}}$ extracted from the square $S_{j,i}$ is retransformed according to the value of the token $\ell_{j,i}$. If $\ell_{j,i} = \emptyset$, it means that no edge is present in $S_{j,i}$ and an orthogonal 2D wavelet transform is applied to $\tilde{f}_{j,i}$ to obtain $f_{j,i}$. If $\ell_{j,i} \neq \emptyset$, an edge passes through $S_{j,i}$ with a direction close to the angle $\theta_{j,i} = \pi \ell_{j,i}/\kappa \in [0, \pi)$. Then a 1D directional wavelet transform is applied to $\tilde{f}_{j,i}$ to obtain $f_{j,i}$. This transform is computed by ordering the
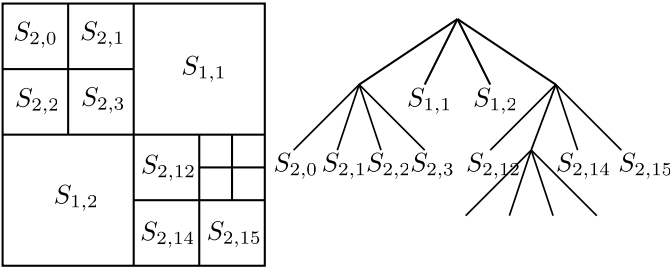
Fig. 8. *Left: example of dyadic subdivision of $[0,1]^2$ in squares $S_{j,i}$; right: corresponding quad-tree λ.*

pixels of $\tilde{f}_{j,i}$ along the direction $\theta_{j,i}$ and then applying a 1D Haar transform, as described in [16]. Figure 9 (c,d) shows this directional wavelet transform. The resulting coefficients $f_{j,i}$ correspond to projections of f on the bandlet vectors of the basis $\mathcal{B}^\lambda$

$$\forall\,(j,i) \in \mathcal{L}(\lambda), \quad f_{j,i}[s] = \langle f, \psi_{j,i,s}^{\ell_{j,i}}\rangle.$$

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

A best basis $\mathcal{B}^{\lambda^*}$ defined by (7) adapted to some function f is computed using a fast best-basis search introduced by Mallat and Peyré [15]. Figure 10 shows an example of such a segmentation adapted to a geometric image.
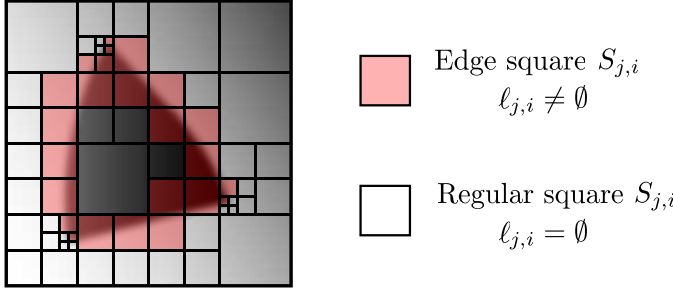

Fig. 10. *Example of subdivision λ^* adapted to a geometrically regular function f. The basis $\mathcal{B}^{\lambda^*}$ minimizes $\mathcal{E}(f, \mathcal{B}^\lambda, t)$.*

B. Numerical Results

The geometric image f depicted in figure 12 (a), (a) is used to compare the performance of the original compressive sensing algorithm in a wavelet basis to the adaptative algorithm in a best bandlet basis. We use here noisy measurements $y = \Phi f + w$ with $\|w\| = 0.03\|\Phi f\|$. We use a translation invariant 7/9 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 $\theta_{j,i}$ that closely match the direction of the edges of f. Figure 12, (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.
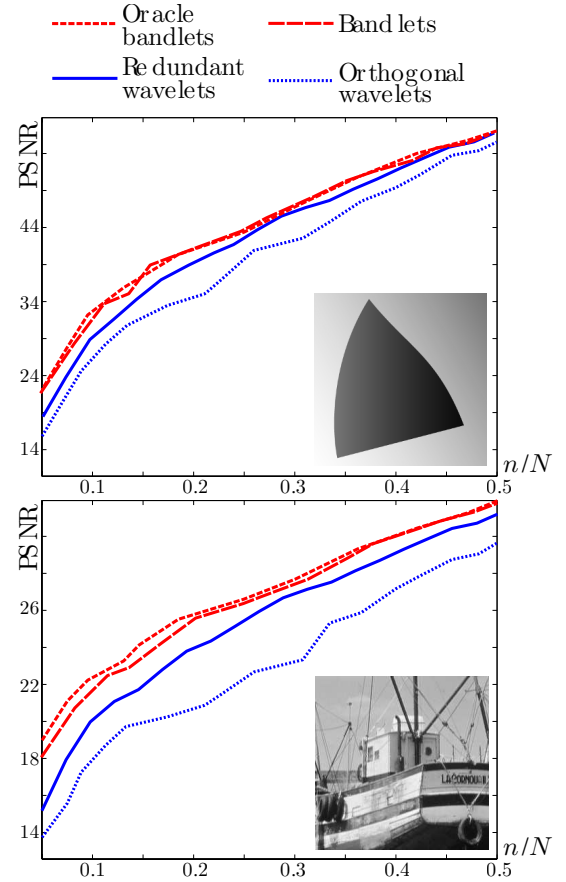

Fig. 11. *Recovery results for various rate of sensing n/N, in the noiseless setting $y = \Phi f$.*

Figure 11 shows the recovery error for various sensing rates, in the noiseless setting $y = \Phi f$. 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 underlying geometry efficiently and the algorithm performs as good as the oracle basis.

VIII. CONCLUSION

This paper tackles the problem of improving the reconstruction from compressed sensing measurements. The proposed method extends compressed sensing recovery to the setting where the sparsity is imposed in a tree structured dictionary of orthogonal bases. This best basis adaptive prior enhances the sparsity of the signal to recover with respect to fixed priors in an orthogonal basis. An iterative algorithm minimizes a variational energy over both the signal to reconstruct and the best basis parameter. Numerical results show that this approach is successful for natural sounds and geometric images that contain a broad range of sharp transitions. Best basis regularization extends to more general linear inverse problems, and could potentially improves the reconstruction beyond randomized measurements.

REFERENCES

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

[2] ——, "Decoding by linear programming," *IEEE Trans. Info. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
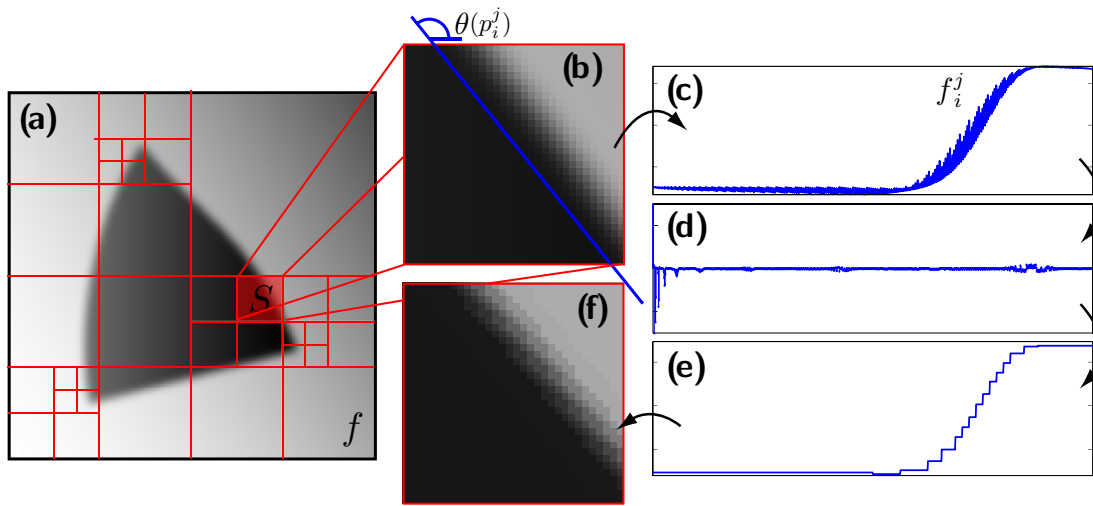
Fig. 9. *(a) a geometric image together with an adapted dyadic segmentation ; (b) a square $S_{j,i}$ together with its adapted direction $\theta_{j,i}$; (c) the 1D signal obtained by re-ordering the pixels of $\tilde{f}_{j,i}$ along direction $\theta_{j,i}$; (d) the 1D Haar coefficients $f_{j,i}$ of $\tilde{f}_{j,i}$; (e) the 1D approximation obtained by reconstruction from the 20 largest Haar coefficients ; (f) the corresponding square approximated in bandelet.*
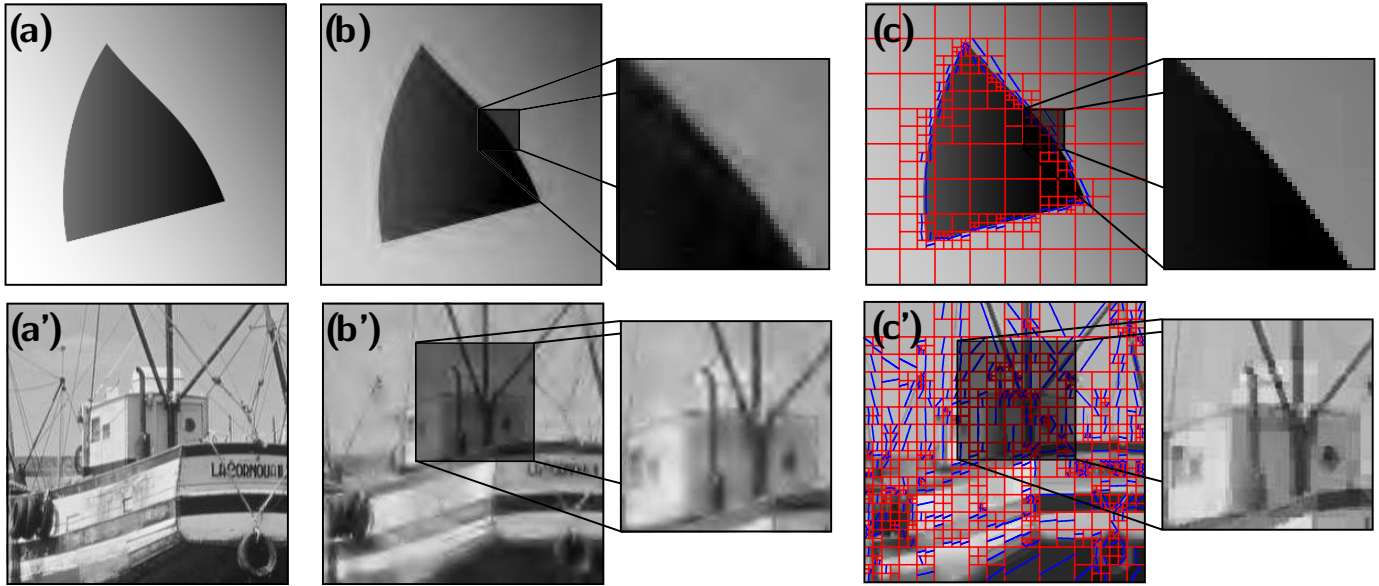

Fig. 12. *(a,a') original image ; (b) compressed sensing reconstruction using the translation invariant wavelet frame, $n = N/6$ (PSNR=37.1dB) ; (c) reconstruction using iteration in a best bandlet basis (PSNR=39.3dB). (b') wavelet frame, PSNR=22.1dB, (c') bandlet basis, PSNR=23.9dB.*

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Info. 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," *Commun. on Pure and Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.

[5] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 2, no. 25, pp. 83–91, 2008.

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

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

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

[9] S. Mallat, *A Wavelet Tour of Signal Processing, 3rd edition*. Elsevier, 2009.

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

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

[12] N. Ouarti and G. Peyré, "Best basis denoising with non-stationary wavelet packets," *Proc. of ICIP 2009*, 2009.

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

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

[15] S. Mallat and G. Peyré, "Orthogonal bandlets bases for geometric image approximation," *Commun. on Pure and Appl. Math.*, vol. 61, no. 9, pp. 1173–1212, 2008.

[16] G. Peyré and S. Mallat, "Surface compression with geometric bandelets," *ACM Transactions on Graphics, (SIGGRAPH'05)*, vol. 24, no. 3, pp. 601–608, Aug. 2005.

[17] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Info. Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.

[18] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, 2006.

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

[20] R. Baraniuk and M. Wakin, "Random projections of smooth manifolds,"

Foundations of Computational Mathematics, vol. 9, no. 1, pp. 51–77, 2009.

[21] G. Peyré, "Manifold models for signals and images," *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 249–260, Feb. 2009.

[22] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," *IEEE Trans. Info. Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[23] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *Preprint*, 2008.

[24] M. Duarte, M. Wakin, and R. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," *Proc. of SPARS 2005*, 2005.

[25] G. Peyré, S. Bougleux, and L. D. Cohen, "Non-local regularization of inverse problems," in *ECCV'08*, ser. Lecture Notes in Computer Science, D. A. Forsyth, P. H. S. Torr, and A. Zisserman, Eds., vol. 5304. Springer, 2008, pp. 57–68.

[26] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *CAM Report 09-03*, 2009.

[27] J. M. D. Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Proc.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[28] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Proc.*, vol. 5, no. 12, pp. 5695–5702, 2007.

[29] 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.

[30] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris, Série I*, no. 346, pp. 589–592, 2008.

[31] ——, "Compressive sampling," *Proceedings of the International Congress of Mathematicians, Madrid, Spain*, 2006.

[32] M. Rudelson and R.Vershynin, "On sparse reconstruction from fourier and gaussian measurements," *Commun. on Pure and Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.

[33] D. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *Preprint*, 2006.

[34] J. Shanks, "Computation of the fast walsh-fourier transform," *IEEE Transactions on Computers*, vol. C, no. 18, pp. 457–459, 1969.

[35] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. on Pure and Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

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

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

[38] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

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

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

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

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