Non Stationary Wavelets: A New Approach To Best Basis Selection

Khaled MELKEMI

Department of mathematics, university of Biskra BP.145 07000 Algeria
kmelkemi@yahoo.fr

Abstract: In image and signal processing, the algorithms of compression depend on wavelet bases and their approximations properties. It is a particular and suitable family which gives a sparse representation of piece-wise smooth functions (image or signal). By “suitable” we mean that the representation allows a simple identification of function's information (e.g. regularity).

In this work, we propose a new approach to select the “best” representation for a given function. First, we describe the construction of a wide family of non stationary wavelet bases with specific properties. Then, we give a criterion (and the proof) of best basis selection. Finally, we present some experimental results proving the effectiveness of our approach.

Key words: FWT algorithm, wavelets, image compression, filtering.

INTRODUCTION

The main defect of Fourier based compression methods (such as DCT) is due to the limitation put on the size of the blocks, and the inability to adjust the patterns to the nature of the signal (or image). An answer to this problem is provided by multiscale decomposition: low frequency trends occurring at the large scale in the signal can efficiently coded with very few coefficients.

Wavelets with many vanishing moments yield sparse decompositions of piecewise smooth curves and surfaces, and are very efficient for coding piece-wise smooth signals [MEL 99]. However, wavelets are ill suited to represent oscillatory patterns. Rapid variations of intensity can only be described by the small scale wavelet coefficients. Long oscillatory patterns thus require many of such fine scale coefficients. Unfortunately, those small scale coefficients carry very little energy, and are often quantized to zero, even at high bit rate. Much larger libraries called wavelet packet bases, introduced in [COI 93], have been constructed to address this problem. These bases can vary in scale, frequency and location. The idea is to decompose a discrete signal using all possible wavelet packet bases of a given wavelet kernel, and then to find the best wavelet packet basis [COI 93]. In general, a basis is a well suited basis if it can describe the target signal with a very small number of basis vectors. In [COI 93], one choice of the best basis is the one which results in fewest numbers of nonzero coefficients after threshold is performed. Other choices include the use of one sided metrics like entropy to assess the efficiency of a basis. However, these schemes are not optimal in rate-distortion sense, nor do they address the problem for arbitrary sets of quantization choices.

In other approach [RAM 93], the best basis and its optimal quantization choice are found using an R-D criterion of the smallest distortion for a given bit budget.

Many classes of signals have very diffuse representation in a standard wavelet basis. Fingerprints or seismic signals are few examples of non wavelet friendly signals. An adapted wavelet packets basis can often provide a very sparse representation of such signals. However, they are not sufficiently adaptive and computationally intensive. Emphasis in this paper has been placed on representation of signals with non-stationary wavelet bases. The original main contribution of this work:

(1) Proposing a new model for oscillatory signal: they are considered as a linear combination of cosine and sine functions, with some specified frequencies. Mathematically speaking, signal is a chebyshevian function.

(2) Giving a link between nonstationary wavelet bases and chebyshevian functions and their properties.

(3) Introducing a new criterion for the best basis. A theoretical existence and uniqueness results are proved. Hence a computational algorithm is proposed.
The organization of the paper is as follows. In section 2 and 3, we present a motivation of our work and the non stationary wavelet bases. We discuss the conditions under which they provide an exact approximation of oscillatory signals. In section 4 and 5, we formulate the problem of selecting the best basis. We then present a series of results that simplify and solve this optimization problem.

1. Motivation and problem statement

1.1. A mathematical formulation

A useful orthonormal basis permits to approximate efficiently a function (in our interesting class) with only a few numbers of vectors. Compression and denoising are two examples of applications. A signal in a separable Hilbert space \( H \) is approximated by

\[
\mathbf{f}_m = \sum_{k=0}^{m-1} \left< f, \mathbf{e}_{n_k} \right> \mathbf{e}_{n_k},
\]

where the \( m \) vectors \( \{ \mathbf{e}_{n_k} \}_{k=0,\ldots,m-1} \) are selected adaptively in an orthonormal basis \( \{ \mathbf{e}_k \}_{k \in \mathbb{N}} \) of \( H \).

The approximation error is given by

\[
\epsilon_m(f) = \left\| f - \mathbf{f}_m \right\|_2^2 = \sum_{k=2m}^{\infty} \left| \left< f, \mathbf{e}_{n_k} \right> \right|^2.
\]

The optimal choice \( \mathbf{f}_m \) (which minimizes \( \epsilon_m(f) \)) corresponds to the increasing sequence \( \{ \mathbf{e}_{n_k} \}_{k=0,\ldots,m-1} \) such that

\[
\left| \left< f, \mathbf{e}_{n_k} \right> \right| \geq \left| \left< f, \mathbf{e}_{n_{k+1}} \right> \right|, \quad k \in \mathbb{N}.
\]

Assume that, for a constants \( C > 0 \) and \( r > 1/2 \):

\[
\left| \left< f, \mathbf{e}_{n_k} \right> \right| \leq C k^{-r}, \quad k \geq m.
\]

Then, with some calculations, we have

\[
\epsilon_m(f) \leq \frac{C^2}{2r-1} m^{1-2r}.
\]

We wish a smallest possible \( m \) and a nice approximation (smallest approximation error).

1.2. Why non stationary bases?

Orthogonal wavelet bases \( \{ \psi_{jk} \}_{j,k \in \mathbb{Z}} = \{ 2^{j/2} \psi(2^j \cdot - k) \}_{j,k \in \mathbb{Z}} \) (with compact support) are a particular family which represent efficiently piece-wise smooth signals. We consider the following example:

\[
Z(t) = p(t) \chi_{[0,1]}(t), \quad t \in \mathbb{R},
\]

where \( p \) is a polynomial of degree \( \leq N-1 \). We decompose the signal in an orthonormal wavelet bases with \( N \) vanishing moments:

\[
\int_{\mathbb{R}} t^k \psi(t) dt = 0, \quad k = 0,\ldots,N-1.
\]

The wavelet decomposition of the signal \( Z \) is given by

\[
Z = \sum_{j,k \in \mathbb{Z}} \langle Z, \psi_{jk} \rangle \psi_{jk}.
\]

The vanishing moments of the functions \( \psi_{jk} \) show that

\[
\langle Z, \psi_{jk} \rangle = 0, \quad \text{if } \text{int}(I_{jk}) \cap [0,1] = \emptyset,
\]

where \( I_{jk} = \text{supp} \psi_{jk} \).

It means that all wavelet coefficients are localized around 0 and 1 (discontinuities of \( Z \)). So, we have a compact representation of \( Z \). This property may be generalized to piece-wise smooth signals \( f \) by using the vanishing moment properties and Taylor formula. We obtain for large scales \( 2^j \)

\[
\langle f, \psi_{jk} \rangle = \int_{I_{jk}} \int_{-2^{-j} k}^{2^{-j} k} (t - 2^{-j} k)^{N-1} \psi_{jk}(x) f^{(N)}(t) dt dx.
\]

More precisely, one can prove that

\[
\left| \left< f, \psi_{jk} \right> \right| \leq \gamma_{N,p} (2^{-j})^{N-1} \left\| f^{(N)} \right\|_{L_p(I_{jk})},
\]

where \( \gamma_{N,p} \) is a constant which is independent of \( f \). This inequality shows that if \( f \) is piece-wise smooth, one can rearrange the sequence \( \{ \left< f, \psi_{jk} \right> \}_{j,k \in \mathbb{Z}} \) such that the condition (1) (and by the way (2)) holds with a constant \( r \) linked to the smoothness of \( f \). This remark shows us that the wavelet coefficients \( \left< f, \psi_{jk} \right> \) depend on the function \( f^{(N)} \).

In order to illustrate this fact, let us consider the following example:

\[
S(t) = \cos(\omega t) \chi_{[0,1]}(t), \quad t \in \mathbb{R},
\]

where \( \omega \) is a positive number. We have

\[
S^{(N)}(t) = \omega^N S_N(t),
\]

where \( S_N \) is a bounded function defined by

\[
S_{2^k}(t) = (-1)^k \cos(\omega t) \chi_{[0,1]}(t), \quad t \in \mathbb{R},
\]

and

\[
S^{(N)}(t) = \omega^N \sum_{k=0}^{N-1} (-1)^k \cos(\omega t) \chi_{[0,1]}(t),
\]

and...
The function $S$ is very oscillatory if the frequency $\omega$ is large. Then, the amplitude of wavelet coefficients $\|(S, \psi_{jk})\|$ becomes (very) large. Indeed, the coefficient $C$ in (2) depends on the function $S^{(N)}$ and becomes very large when $S$ has a very oscillatory patterns (i.e., $\omega$ is very large). More precisely, we have

$$C = \gamma_{N, p} \omega^N \|S_N\|_{L^p([0,1])}.$$ 

In this case, the inequality (2) does not give any guarantee on the quality of approximation for “small” numbers $m$.

The equalities (3) and (4) show us that the function $S$ is sensitive to the oscillations of $S$. One can consider the question: “can we find a basis $\psi_{jk}$ satisfying similar conditions of vanishing moments:

$$\int_{\mathbb{R}} \cos(\omega x) \psi_{jk}(x) dx = 0, \quad \text{for all } j, k \quad (5)$$

To satisfy these conditions, the functions $\psi_{jk}$ are generated by dilation and translation of a single function $\psi_{k} = 2^{j/2} \psi(2^{j}x - k)$, if and only if $\omega = 0$.

In order to satisfy the conditions (5), we must consider the non stationary wavelet bases of $L^2(\mathbb{R})$, namely

$$\left\{\phi^\lambda_{jk}\right\}_{j,k \in \mathbb{Z}} \cup \left\{\psi^\lambda_{jk}(x) = 2^{j/2} \psi(2^{j}x - k) \right\}_{j \geq j_0, k \in \mathbb{Z}}$$

for some $j_0 \in \mathbb{Z}$.

2. The algorithm of best basis selection

We propose here a new adaptive approach. It is optimal in sense that the best-basis is determined by the minimization of some norm. Our selection criterion is based on the theory of Chebyshevian spaces. It is a generalization of polynomials which conserves a large number of their properties (Taylor formula, splines construction, optimal approximation, . . ) [MEL 99][SCH 81]. In order to obtain a fast algorithm, we consider the translation invariant Chebyshev spaces. This space $\mathcal{S}_\lambda$ is defined as a kernel of differential operator with real and constants coefficients:

$$L_\lambda = \prod_{k=1}^{N} \left( \frac{d}{dx} - \lambda_k \right),$$

the polynomial case corresponds to $\lambda_k = 0$, for all $k$. One defines the spline functions by replacing polynomials with elements of the chebyshev space, and preserving a large number of their important properties [MEL 99]. This generalization is useful since the parameters $\lambda_k$ offer a number of freedom degree. We consider in this paper the vectors $\hat{\lambda} \in \mathbb{C}^N$ satisfying the following conditions:

(A1) $\hat{\lambda} = \overline{\lambda}$.

(A2) $\lambda_k - \lambda_\ell \notin 2^{j_0 + 1} i \pi \mathbb{Z} \setminus \{0\}$ for all $j \geq j_0$, $k, \ell = 1, \ldots, N$ with some $j_0 \in \mathbb{Z}$.

(A3): It exists a vector $\mu \in \mathbb{R}^N$ such that $\lambda_k = i \mu_k$ for all $k = 1, \ldots, N$.

The first condition ensures that the operator $L$ has a real coefficients. The condition (A2) ensures the stability in $L^2$-sense of the basis

$$\left\{\phi^\lambda_{jk}\right\}_{j,k \in \mathbb{Z}} \cup \left\{\psi^\lambda_{jk}(x) = 2^{j/2} \psi(2^{j}x - k) \right\}_{j \geq j_0, k \in \mathbb{Z}}$$

The last condition is a restriction to trigonometric spaces. It enables us to adapt to the different structures (oscillations, smoothness) of the analyzed function. We quote the generalized Taylor formula from [MEL 99]:

**Theorem 1.** For each function $f \in C^N([a,b])$, there exists a function $u_f \in \mathcal{S}_\lambda$ such that

$$f(x) = u_f(x) + \int_a^b G_{\lambda}(x-t)L_\lambda f(t)dt,$$

Where $G_{\lambda}$ is the Green’s function associated with the operator $L_\lambda$.

The Green function satisfy the following inequality (see [SCH 81]):

$$G_{\lambda}(x-y) \leq \kappa \frac{(x-y)^{N-1}}{(N-1)!}, \quad \forall x, y \in [a,b], \quad (8)$$

where $\kappa$ is nonnegative constant. Thanks to the condition (A3), the number $\kappa$ does not depend on $\hat{\lambda}$.

Using nonstationary wavelet bases with the property (5), we obtain (for large $j$)
\[ \langle f, \psi_{jk} \rangle = \int_G G_k(x - 2^{-j} k) \psi_{jk}(x) f(t) \, dt \, dx. \]

From the relation (8), we show that
\[ \| f, \psi_{jk} \| \leq C (2^{-j})^N \| L_\lambda f \| \| \lambda \| . \]

The constant \( C \) is a positive and independent of \( \lambda \).

As examples, the well-adapted basis to represent the piece-wise polynomials functions is the wavelet associated to the operator \( \frac{\partial^N}{\partial x^N} \). On the other hand, the basis associated to the operator \( \frac{\partial^2}{\partial x^2} + \omega^2 I \) is the most appropriate to represent the signal \( S \). Indeed, in this case, all coefficients \( \langle S, \psi_{jk} \rangle \) are vanish if \( \{0,1 \} \cap \text{int}(I_{jk}) = \emptyset \).

We introduce the following definition:

**Definition 2.** The best nonstationary wavelet basis to represent a function \( f \) relative to the norm \( \| \cdot \| \) is the nonstationary basis associated to the operator \( L_\lambda \) minimizing the norm \( \| L_\lambda f \| \).

The following observation is very important to describe an algorithm of best-basis selection relative to a given signal.

**Observation 3.** A differential operator \( L \) satisfying the conditions (A1)-(A3) if and only if there exist a positive numbers \( 0 < \mu_1 < \mu_2 < \cdots < \mu_d \) such that

\[ L = \left( \frac{\partial}{\partial x} \right)^{r_0} \prod_{k=1}^{d} \left( \frac{\partial^2}{\partial x^2} + \mu_k^2 \right)^{r_k}, \]

where the integers \( \{ r_k \}_{k=0,\ldots,d} \) satisfy \( r_0 + 2 \sum_{k=1}^{d} r_k = N \).

In addition, for each function \( f \) in Sobolev space \( H^N \) we have

\[ \widehat{L_f}(\xi) = \left[ (i\xi)^{r_0} \prod_{k=1}^{d} (\mu_k^2 - \xi^2)^{r_k} \right] \hat{f}(\xi), \]

where \( \hat{f} \) denote the Fourier transform of \( f \).

We propose in our approach to consider the standard \( L^2 \) norm for many reasons. First, it is the unique hertibertian norm in \( L^p \) spaces and the wavelet analysis use often this norm. The other major reason is the isometric property of this norm with the Fourier transform. We obtain to minimization problem of the function

\[ \int_{\mathbb{R}} \left| \widehat{L_f}(\xi) \right|^2 d\xi = \int_{\mathbb{R}} \left[ \xi^{2r_0} \prod_{k=1}^{d} (\mu_k^2 - \xi^2)^{2r_k} \right] \left| \hat{f}(\xi) \right|^2 d\xi, \]

subject to the coefficients \( \mu \in \mathbb{R}^d \) and \( \{ r_k \}_{k=0,\ldots,d} \).

In order to solve this optimization problem we state the problem in its general form.

### 3. Classification of images

Let \( D_f(\mathbb{R}) \) be a set of real, nonnegative and measurable functions \( u \) such that

\[ \int_{\mathbb{R}} (1 + |t|^2) u(t) \, dt < +\infty, \quad k = 0,\ldots,2N. \]

For a function \( u \in D_2(\mathbb{R}) \), we define the function

\[ F_u(x) = \int_{\mathbb{R}} u(t) \left( \prod_{k=1}^{d} (t-x_k)^2 \right) \, dt, \]

For \( x = (x_1,x_2,\ldots,x_N)^T \in \mathbb{R}^N \), we consider the minimization problem:

\[ \text{Find } m \in \mathbb{R}^N, \text{ such that } F_u(m) = \min_{m \in \mathbb{R}^N} F_u(x). \]  

**Theorem 4.** (Generalized expectation). The problem (9) has a unique solution \( m \) up to a permutation of the sequence \( m_1, m_2,\ldots, m_N \).

We introduce the following definition.

**Definition 5.** We call \( m \) the generalized expectation (or mean) of order \( N \) (\( N \)-expectation for short).

The classical expectation is a particular case of our definition where \( N = 1 \).

We present here an idea of the proof, for more details; we refer the reader to [MEL 08a].

**Sketch proof of Theorem 4.** First we solve the equation \( \nabla F_u(x) = 0 \), and we prove that the Hessian \( \nabla^2 F_u(m) \) is a positive definite matrix. We note that our calculus show that

\[ m_i \neq m_j, \quad \forall i \neq j. \]

and this property does not depends on \( u \).
4. Experimental examples

We present here some experimental examples of a specific images. We have choused a piece wise image: Lena, an image with smooth region and some texture (no isolated singularities): Mandrill. The third image is a generic texture where there is no smooth region in the mathematical sense.

Figure 1. Original image 1: Lena

Figure 2. Reconstructed image 1 with 5.7% of the information (standard wavelet)

Figure 3. Original image 2: Mandril

Figure 4. Reconstructed image 2 with 15% of information (standard wavelet)

Figure 5. Original image 3: Texture

Figure 6. Reconstructed image 3 with 20% of the information (standard wavelet)

The third example (Figure 5 and 6) shows that the standard wavelet basis is very poor. Then we have to perform non stationary algorithms with the parameters $\lambda = (-900; 900)$. The choice is optimal since the texture can be considered as combination of cosine and sine function with frequencies near to 900.
Figure 7. Reconstructed image with parameters (-900i; 900i), and 20% of information

We can detect these frequencies by the algorithm of Newton applied on the proposed function.

4. Conclusion

In this paper we have presented a new approach to represent images in order to perform many procedures as compression and image denoising. Our approach is adaptive. We described a simple algorithm with a mathematical justification which finds the best representation in a specified sense. Finally, we have presented some experimental results proving the effectiveness of our approach.

ACKNOWLEDGMENT

The authors wish to thank Prof. Karoui and Prof Melgani for their fructuous comments. This work was supported in part by a grant from MESRS Num B701/50/06.

REFERENCES


