On the Initialization of the Discrete Wavelet Transform Algorithm

Patrice Abry and Patrick Flandrin

Abstract—This letter shows that making use of the discrete wavelet transform to analyse data implies performing a preliminary initialization of the fast pyramidal algorithm. An approximation enabling easy performance of such an initialization is proposed.

I. MOTIVATION

TIME-SCALE methods have already proved relevant tools in signal processing. In most cases, for data *analysis*, one uses continuous wavelet transforms rather than a discrete wavelet transform (DWT), which is usually used in *coding*. However, we believe that, thanks to the fast pyramidal algorithm and the minimum redundancy dyadic sampling, the DWT is also an efficient tool when analysing experimental data.

In this letter, our aim is to stress, via two simple examples, the importance of performing a correct initialization of the DWT algorithm when examining the wavelet coefficients to analyze data. Theoretically speaking, initializing the algorithm means projecting data onto the space V_0: the first space of a sequence of successive approximation spaces V_j generating the chosen multiresolution analysis [1]. We propose an approximation that enables us to easily perform this initialization. We will then clarify the relations between the two possible interpretations of the detail $d_x(j, n)$ and approximation $a_x(j, n)$ coefficients: constant-Q band-pass filtered versions of the data versus outputs of the filter-bank pyramidal structure. These dual approaches are summarized through

$$a_x(j, n) = \langle x, \overset{o}{\phi}_{j,n} \rangle = \sum_k a_x(j - 1, k) \overset{o}{g}(k - 2n) \quad (1)$$

$$d_x(j, n) = \langle x, \overset{o}{\psi}_{j,n} \rangle = \sum_k a_x(j - 1, k) \overset{o}{h}(k - 2n) \quad (2)$$

where $\overset{o}{\phi}(t)$ (resp., $\overset{o}{\psi}(t)$) is the dual function of the scaling function ϕ (resp., of the wavelet ψ), and $\overset{o}{h}$ (resp., $\overset{o}{g}$) is its associated filter in the pyramidal algorithm (see [1] or [2] for complete definitions of dual functions. In the orthonormal case, one has $\overset{o}{\phi} \equiv \phi$ and $\overset{o}{\psi} \equiv \psi$). $\overset{o}{\phi}_{j,n}$ (resp., $\overset{o}{\psi}_{j,n}$) are dilated, time-shifted versions of $\overset{o}{\phi}$ (resp., $\overset{o}{\psi}$).

Last, we will illustrate (on a couple of examples) the importance of initializing the pyramidal algorithm. In spite of our using *an orthonormal Daubechies 6 wavelet* [1] in these examples, we do not restrict ourselves to the orthonormal case. In some applications, this initialization can be of greater importance when using nonorthogonal wavelets.

II. INITIALIZING THE PYRAMIDAL ALGORITHM

Experimental data very often consist of a collection of samples $\{x_n = x(nT_e), n \in \mathcal{Z}\}$, which are related to the physical signal $x(t)$ through the sampling theorem:

$$x(t) = \sum_x x_k \mathrm{sinc}\,(t - kT_e)$$

where T_e is the sampling period. Let us immediately remark that this sampling period gives a physical meaning to the approximation space V_0. Indeed, the discrete-time filtering of the fast algorithm sets $T_e = 1$ and thus selects the real-time spread of the father wavelet (the scaling function), which generates V_0 and of the mother wavelet. For instance, when the scaling function time support lies in $[-N, M]$, it actually means that the data will be projected onto a space V_0 generated by function $\phi(t)$ whose time support is $[-NT_e, MT_e]$. From now on, we will assume $T_e = 1$. Thus, the variable t will stand for a normalized time with respect to T_e.

Most of the time, the initialization of the DWT simply consists of applying the oversimplification:

$$a_x(j = 0, k) \equiv x_k$$

whereas, in fact, the $a_k(0, k)$ should result from the projection of $x(t)$ onto V_0:

$$a_x(0, k) = \int x(t) \overset{o}{\phi}(t - k)\, dt. \quad (3)$$

This result was already stated by Shensa, as reported in [3]. Recalling that $x(t)$ is only known through its samples $\{x_n\}_{n \in \mathcal{Z}}$, this inner product can hardly be computed but rewritten as

$$a_x(0, k) = \sum_k x_n \alpha_{k-n}$$

with $\alpha_p = \langle \mathrm{sinc}, \overset{o}{\phi}_{0,-p} \rangle = \int \mathrm{sinc}\,(t) \overset{o}{\phi}(t - p)\, dt$.

The general calculation of these inner products α_p is not easy, yet we propose a rough approximation that enables us

to estimate them. The examples of the next sections tend to prove that this approximation is valid. One can rewrite α_p as

$$\alpha_p = \int \overset{o}{\Phi}(\nu) \exp\left(-2i\pi\nu p\right)\Pi_1(\nu)\, d\nu$$

where $\overset{o}{\Phi}(\nu)$ and $\{\Pi_1(\nu) = 1$, if $0 \leq |\nu| \leq 1/2$, and 0 elsewhere$\}$ are the Fourier transforms of $\overset{o}{\phi}(t)$ and $\mathrm{sinc}\,(t)$, respectively. One then sees that neglecting the behavior of $\overset{o}{\Phi}(\nu)$ for $|\nu| \geq 1/2$ enables us to assimilate $\mathrm{sinc}\,(t)$ to a Dirac function within the integral. Thus, α_p simplifies to $\alpha_p \simeq \overset{o}{\phi}(-p)$. It follows that $a_x(0,k)$ reads

$$a_x(0,k) \simeq \sum_n x_n \overset{o}{\phi}(n-k). \tag{4}$$

The quality of this approximation can be examined in Fig. 1. A very good approximation of the time-shape of $\overset{o}{\phi}(t)$ can be obtained as the impulse response of the filter $\overset{o}{h}$ convolved with itself a large number of times. One then simply needs to collect the samples $\{\overset{o}{\phi}(p) = \overset{o}{\phi}(t = p)\}_{p\in\mathcal{Z}}$ to perform the approximate initialization of (4).

We will show, using examples, that initializing the algorithnm with (4) rather than with the samples $\{x_n\}$ makes an important difference. Let us note that these modifications would be useless in the case of scaling functions whose time support is $-1 \leq t \leq 1$, (e.g., the Haar wavelet or splines of order 0 or 1). Let us also remark that initializing the $a_x(0,k)$ is of minor importance when dealing with largely oversampled data $x(t)$. This is straightforward from the definition (see (3)) and reasoning among the same line.

III. EXAMPLE 1: ANALYZING A DIRAC FUNCTION

Let us imagine that the data to be analyzed reads $x(t) = \delta(t - t_0)$. Then, one has: $d_\delta(j,k) = \overset{o}{\psi}_{j,k}(t_0)$. The output coefficients of the pyramidal algorithm are plotted in Fig. 2. When no initialization was made (top), one can see that the *detail* signals do not correspond to samples of any dilated version of $\overset{o}{\psi}$, which can be checked by direct numerical comparison. On the contrary, when the initialization has been performed (bottom), one perfectly recognizes the samples of dilated versions of $\overset{o}{\psi}$, which is the theoretical result expected from the inner product in (2). Therefore, this example shows that the two approaches of (1) and (2) are well meeting only when the proposed initialization of the $a_x(0,k)$ is performed. Furthermore, if one is interested in locating, in the data, transient events, (modeled by a Dirac impulse) to be detected from the $d_x(j,k)$, it is clear from Fig. 2 that an incorrect initialization is responsible for an incorrect location estimation. Indeed, the maximum of $|d_x(j,k)|$ at scale j neither coincide in time with that of another scale or with the location of the Dirac impulse. This lack of coincidence will therefore be a serious drawback when detecting transients from time-aligned maxima, whereas initializing the DWT will ensure a time coincidence of the impulse and the maxima of the $|d_x(j,k)|$ through scales.
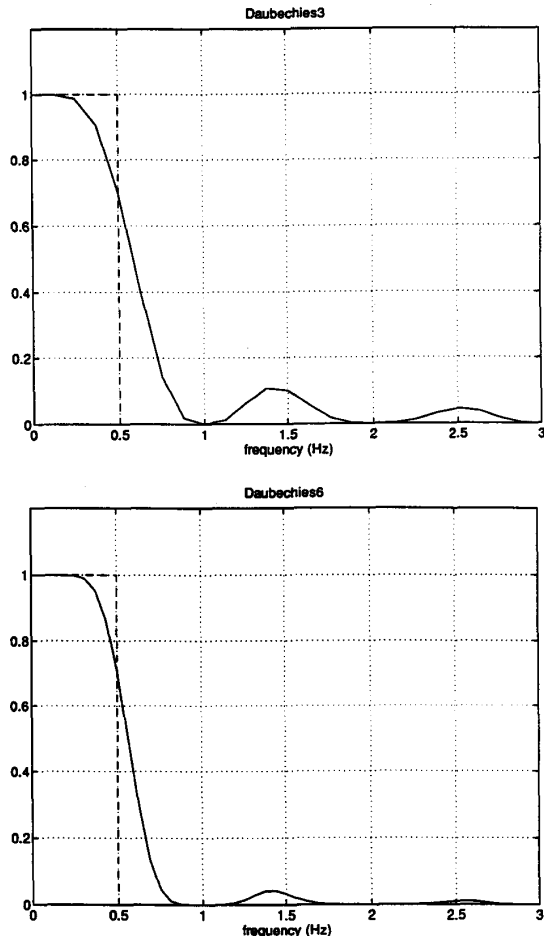

Fig. 1. Fourier transforms of the scaling functions associated with Daubechies3 wavelet (top) and Daubechies6 wavelet (bottom). The shape of the ideal low-pass filter has been superimposed. The higher the degree of regularity, the better the quality of the approximation (most of the energy lies within $|\nu| \leq 1/2$).

In Fig. 2, one also sees that the values taken by the $d_x(j,k)$ remain the same from scale to scale, but one should note that this is so only because the Dirac function was located at *a good position* on the dyadic grid. If it had not been so well located, one would still have had samples of a dilated wavelet, but the reference position of the sampling would have changed with scale. Yet, initializing the pyramid still ensures the correspondence of (1) and (2), and this is the crucial point.

IV. EXAMPLE 2: ANALYZING A WAVELET

Let us now use another striking example: $x(t) = \psi_{j0,k0}(t)$. Let us assume orthonormality of the $\{\psi_{j,k}(t)\}$; then, one has $d_{\psi_{j0,k0}}(j,k) = 1$ when $(j,k) = (j_0,k_0)$, and $d_{\psi_{j0,k0}}(j,k) = 0$ elsewhere. The wavelet coefficients computed by the pyramidal algorithm are displayed in Fig. 3. When no initialization is made, one gets $d_{\psi_{j0,k0}}(j,k)$, which are significantly different from 0 for scales $j \neq j_0$ and locations $k \neq k_0$. On the contrary, initializing the $a_x(0,k)$ results in one large coefficient at the
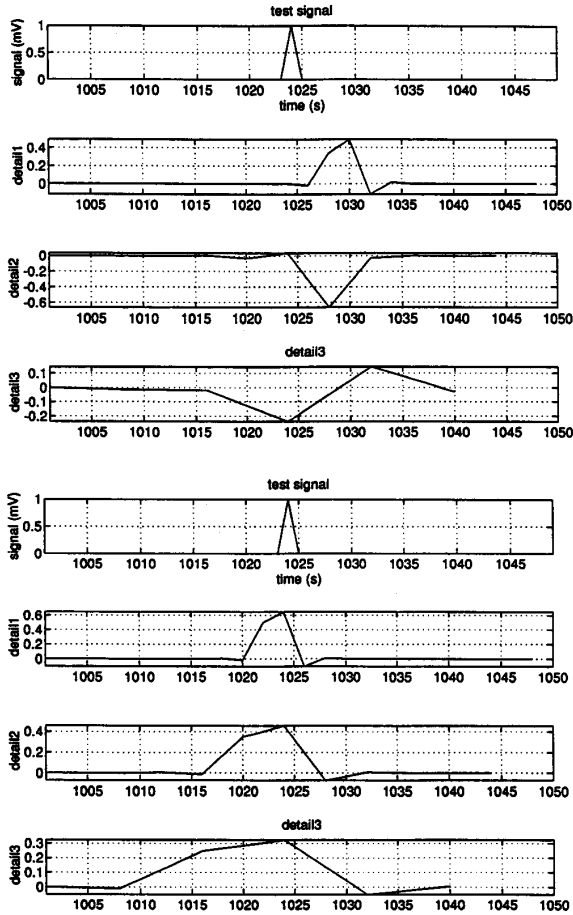
Fig. 2. DWT of a Dirac function (least asymmetric Dirac Daubechies6 wavelet) without initialization (top) and with a correct initialization of the pyramidal algorithm (bottom). In this latter case, one recognizes exactly sampled shaped of dilated versions of the mother wavelet.
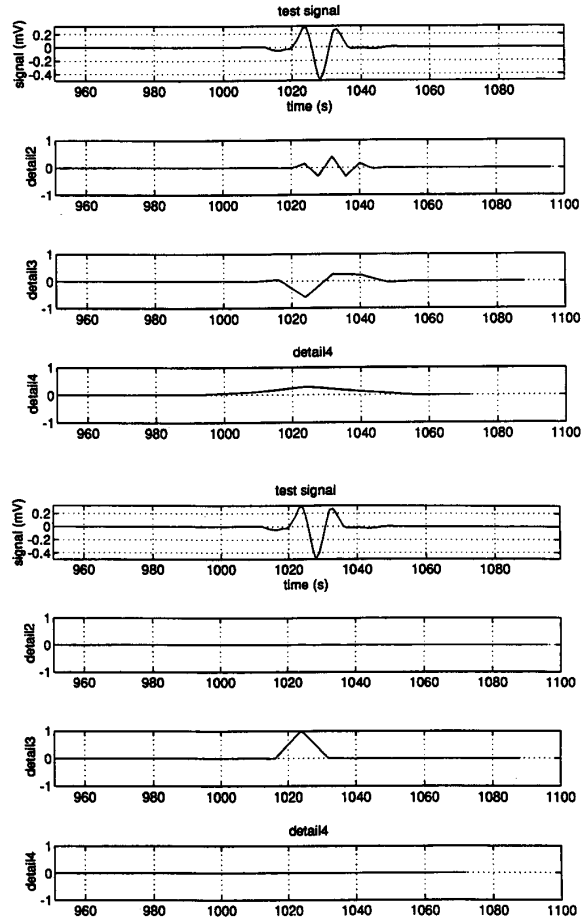

Fig. 3. DWT of a $\psi_{3,128}(t)$ signal (least asymmetric (Daubechies6 wavelet) without initialization (top) and with a correct initialization of the pyramidal algorithm (bottom). In this latter case only, one finds a single very strong coefficient (on scale 3), betraying that $x(t)$ is a dilated version of the mother wavelet.

correct time- scale location, corresponding to the theoretical result. This is, of course, of importance since it is clearly telling us that the analyzed data consists of one of the atoms involved in the decomposition. Of course, the remark of the previous section concerning a *good location* on the dyadic grid still holds.

V. CONCLUSION

In this letter, we have shown that a relevant interpretation of the information conveyed by data can be performed through the DWT on condition that a correct initialization of the pyramidal algorithm is performed. We have also proposed a simple yet efficient, way of calculating the $a_x(0, k)$, which, in some sense, should always be the preliminary step when running a DWT.

REFERENCES

[1] I. Daubechies, *Ten Lectures on Wavelets.* New York: SIAM, 1992.
[2] Y. Meyer, *Ondelettes et algorithmes concurrents.* Hermann, 1992, 1st ed.
[3] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Trans. Inform. Theory,* vol. 38, no. 2, pp. 569–586, 1992.