TIME INVARIANT ORTHONORMAL WAVELET REPRESENTATIONS

J.-C. Pesquet‡, H. Krim§ and H. Carfantan‡
‡Laboratoire des Signaux et Systèmes,
CNRS/UPS and GDR TdSI,
ESE, 91192 Gif sur Yvette Cedex, France
tel: (33) 1 69 41 80 40 Ex. 17-12, fax: (33) 1 69 41 30 60
e-mail: pesquet@lss.supelec.fr
§Stochastic Systems Group, LIDS, MIT,
Cambridge, MA 02139

Permission to publish this abstract is granted
Edics: SP2.4.4, SP2.4.5, SP2.5.3

Abstract

A simple construction of an orthonormal basis starting with a so called mother wavelet, together with an efficient implementation gained the wavelet decomposition easy acceptance and generated a great research interest in its applications. An orthonormal basis may not, however, always be a suitable representation of a signal, particularly when time (or space) invariance is a required property. The conventional way around this problem is to use a redundant decomposition.
In this paper, we address the time invariance problem for orthonormal wavelet transforms and propose an extension to wavelet packet decompositions. We show that it is possible to achieve time invariance and preserve the orthonormality. We subsequently propose an efficient approach to obtain such a decomposition. We demonstrate the importance of our method by considering some application examples in signal reconstruction and time delay estimation.

Keywords: wavelets, wavelet packets, time invariance, estimation, entropy.

Submitted to IEEE Trans. Signal Processing, Feb. 1994

§The work is partially supported by grants from ARO DAAL03-92-G-0115 (Center for Intelligent Control Systems), AFOSR F49620-92-J-0002, and NSF MIP-9015281.

1 Introduction

While the Fourier transform remains a fundamental building block in signal analysis, the increasing demands on signal processing techniques in a variety of complex areas have uncovered many problems for which the Fourier domain is not best adapted. One of the most challenging problems, researchers have to contend with, is the processing of nonstationary signals in general and that of transients in particular. For this reason, the research interest in the wavelet transform further grew over the last five years.

Wavelet transforms can be classified as either redundant or nonredundant (orthogonal). The continuous wavelet transform [6] and the frame decomposition [4] belong to the first class, whereas orthogonal [3, 9] and biorthogonal [1] wavelet decompositions are in the second class. Wavelet packets are a generalization of wavelets and allow one to optimize the representation of a signal. Wavelet packet transforms may also be defined in a redundant [16] or a non redundant form [19, 2].

The non redundant transforms are appealing for several reasons. Firstly, the compression ability of wavelet transforms is better preserved since no additional components are added. Another appealing feature of these transforms is the efficiency of implementation of the decomposition and the corresponding reconstruction through decimated filter banks. In a stochastic setting, a property which gives the orthogonal wavelet transform a useful characteristic is the statistical decorrelation of the wavelet coefficients of a white noise process representation.

The major drawback of non redundant transforms is their noninvariance in time (or space) (*i.e.* the coefficients of a delayed signal are not a time shifted version of those of the original signal). The time invariance property is particularly important in statistical signal processing applications, such as detection or parameter estimation of signals with unknown arrival time. This noninvariance implies that if a detector is designed in the wavelet coefficient domain, its performances will then depend on the arrival time of the signal. To overcome this difficulty, one has often preferred the use of redundant transforms in detection/estimation problems [11].[1] Other works have focussed on the design of alternative representations [15, 12].

In this paper, we show that it is possible to build different orthogonal wavelet representations of a signal while keeping the same analyzing

[1]Note that it allows also more flexibility in the choice of the analyzing wavelet.

wavelet. These decompositions differ in the way the time-scale plane is sampled. By choosing the decomposition which best fits the time localization of the signal, we obtain an improved representation which is time invariant (in a sense which is subsequently discussed). We also consider the extension of these properties to wavelet packets.

The paper is organized as follows. In Section 2, we give some background material together with the notational conventions. In Section 3, we discuss the reconstruction (or synthesis) problem starting with a redundant decomposition and describe the generalized class of orthogonal wavelet representations proposed in this paper. In Section 4, we develop an efficient algorithm for selecting the best wavelet decomposition and subsequently show that it is time-invariant. We extend all these results to wavelet packets, in Section 5. In Section 6, we show, by way of specific application examples, that our approach can achieve significant improvements over existing methods. We conclude with some remarks in Section 7.

2 Background

2.1 Multiresolution Analysis

An orthogonal wavelet decomposition of a signal $x(t) \in \mathrm{L}^2(\mathbb{R})$ leads to coefficients $\{\mathcal{W}_j^k(x)\}_{(k,j)\in\mathbb{Z}^2}$ such that

$$\mathcal{W}_j^k(x) \triangleq < x(t), \frac{1}{2^{j/2}}\psi(\frac{t}{2^j} - k) > \triangleq \frac{1}{2^{j/2}} \int_{-\infty}^{\infty} x(t)\, \psi^*(\frac{t}{2^j} - k)\, dt, \quad (1)$$

where the function $\psi(\cdot)$ is usually referred to as a mother wavelet and * stands for the complex conjugation. The orthonormal wavelet basis $\{2^{-j/2}\psi(t/2^j - k), (k,j) \in \mathbb{Z}^2\}$ may be built from a multiresolution analysis of $\mathrm{L}^2(\mathbb{R})$ [9]. In this case, the approximation of the signal at resolution 2^{-j} can be described by the coefficients

$$\mathcal{A}_j^k(x) \triangleq < x(t), \frac{1}{2^{j/2}}\phi(\frac{t}{2^j} - k) >, \quad k \in \mathbb{Z}, \quad (2)$$

where $\phi(\cdot)$ is the scaling function. The mother wavelet and the scaling functions then satisfy the so called two-scale equations:

$$2^{-\frac{1}{2}}\phi(\frac{t}{2} - k) = \sum_{l=-\infty}^{\infty} h_{l-2k}\, \phi(t-l), \quad (3)$$

$$2^{-\frac{1}{2}}\psi(\frac{t}{2} - k) = \sum_{l=-\infty}^{\infty} g_{l-2k}\, \phi(t-l), \quad (4)$$

where $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$ are respectively the impulse responses of lowpass and highpass paraunitary Quadrature Mirror Filters (QMF) [18]. If we consider the vector spaces $V_j \triangleq \mathrm{Span}\{\phi(t/2^j - k), k \in \mathbb{Z}\}$ and $O_j \triangleq \mathrm{Span}\{\psi(t/2^j - k), k \in \mathbb{Z}\}$, it results from Eqs. (3) and (4) that $V_{j+1} = V_j \overset{\perp}{\oplus} O_j$.[2] We then find that, for every $j_{\mathrm{m}} \in \mathbb{Z}$, $\{2^{-j/2}\psi(t/2^j - k), k \in \mathbb{Z}, j \leq j_{\mathrm{m}}\} \cup \{2^{-j_{\mathrm{m}}/2}\phi(t/2^{j_{\mathrm{m}}} - k), k \in \mathbb{Z}\}$ is an orthonormal basis of $\mathrm{L}^2(\mathbb{R})$. The interest in the QMF filters lies in the efficient computation of the orthogonal wavelet decomposition via a two-channel filter bank structure [10]. The decomposition which is useful in emphasizing the local features of a signal, presents however, a limitation, namely its noninvariance in time (or space). This implies that the wavelet coefficients of $\mathcal{T}_\tau[x(t)] \triangleq x(t-\tau)$, $\tau \in \mathbb{R}$, are generally not delayed versions of $\{\mathcal{W}_j^k(x)\}_{k\in\mathcal{Z}}$.

To circumvent this problem, one can resort to a redundant decomposition of the signal $x(t)$ effected as,

$$\widetilde{\mathcal{W}}_{2^j}^{\theta}(x) \triangleq < x(t), \frac{1}{2^{j/2}}\psi(\frac{t-\theta}{2^j}) >, \tag{5}$$

$$\widetilde{\mathcal{A}}_{2^j}^{\theta}(x) \triangleq < x(t), \frac{1}{2^{j/2}}\phi(\frac{t-\theta}{2^j}) >, \quad \theta \in \mathbb{R}, j \in \mathbb{Z}, \tag{6}$$

This representation is time-invariant since the redundant wavelet and approximation coefficients of $\mathcal{T}_\tau[x(t)]$ are respectively $\mathcal{T}_\tau[\widetilde{\mathcal{W}}_{2^j}^{\theta}(x)]$ and $\mathcal{T}_\tau[\widetilde{\mathcal{A}}_{2^j}^{\theta}(x)]$, $j \in \mathbb{Z}$. Throughout the paper, we will consider redundant wavelet decompositions using wavelets built from a multiresolution analysis.

2.2 Wavelet Packet Decomposition

The wavelet packet decomposition [19] is an extension of the wavelet representation, which allows the best matched analysis to a signal. To define wavelet packets, we first need to introduce functions of $\mathrm{L}^2(\mathbb{R})$, $W_m(t)$, $m \in \mathbb{N}$, such that

$$\int_{-\infty}^{\infty} W_0(t) = 1, \tag{7}$$

and, for all $k \in \mathbb{Z}$,

$$2^{-\frac{1}{2}} W_{2m}(\frac{t}{2} - k) = \sum_{l=-\infty}^{\infty} h_{l-2k}\, W_m(t-k), \tag{8}$$

$$2^{-\frac{1}{2}} W_{2m+1}(\frac{t}{2} - k) = \sum_{l=-\infty}^{\infty} g_{l-2k}\, W_m(t-k), \tag{9}$$

[2] The symbol $\overset{\perp}{\oplus}$ stands for the orthogonal sum of vector spaces.

where $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$ are the previously defined impulse responses of the QMF filters. If, for every $j \in \mathbb{Z}$, we define the vector space $\Omega_{j,m} \triangleq \text{Span}\{W_m(t/2^j - k), k \in \mathbb{Z}\}$, we can then show that

$$\Omega_{j,m} = \Omega_{j+1,2m} \overset{\perp}{\oplus} \Omega_{j+1,2m+1}. \tag{10}$$

As a result, if we denote by $\mathcal{P}$ a partition[3] of $\mathbb{R}^+$ into intervals $I_{j,m} = [2^{-j}m, \ldots, 2^{-j}(m+1)[$, $j \in \mathbb{Z}$ and $m \in \{0, \ldots, 2^j - 1\}$, then

$$\mathrm{L}^2(\mathbb{R}) = \overset{\perp}{\bigoplus_{(j,m)/I_{j,m}\in\mathcal{P}}} \Omega_{j,m}. \tag{11}$$

In an equivalent way, $\{2^{-j/2}W_m(2^{-j}t - k), k \in \mathbb{Z}, (j,m)/I_{j,m} \in \mathcal{P}\}$ is an orthonormal basis of $\mathrm{L}^2(\mathbb{R})$. Such a basis is called a wavelet packet. The coefficients resulting from the decomposition of a signal $x(t)$ in this basis are

$$C^k_{j,m}(x) \triangleq < x(t), \frac{1}{2^{j/2}} W_m(\frac{t}{2^j} - k) > . \tag{12}$$

By varying the partition $\mathcal{P}$, different choices of wavelet packets are possible. For instance, a special wavelet packet is the orthonormal wavelet basis such that $\phi(t) = W_0(t)$ and $\psi(t) = W_1(t)$. We have then $V_j = \Omega_{j,0}$ and $O_j = \Omega_{j,1}$. Another particular case is the equal subband analysis which is defined, at a given resolution level $j_{\mathrm{m}} \in \mathbb{Z}$, by $\mathcal{P} = \{I_{j_{\mathrm{m}},m}, m \in \mathbb{N}\}$. Each possible choice corresponds to a different structure of the filter bank used to implement the related wavelet packet decomposition. This structure may also be described by a binary tree whose nodes are indexed by (j,m) and whose leaves correspond to the indices (j,m) such that $I_{j,m} \in \mathcal{P}$. Such a tree will subsequently be referred to as a *frequency tree*. Fig. 1 shows the frequency trees corresponding to an equal subband analysis.

Generally, a decomposition onto a basis is evaluated by its ability to compress and provide a compact description of the useful information in a signal. It is thus of interest to select the partition $\mathcal{P}$ for which an optimized representation of the analyzed signal is obtained. Several criteria have been proposed to evaluate the compactness of a representation [19, 2]. One of the best known measures is the entropy, which is defined as

$$\bar{\mathcal{H}}_{\mathrm{e}}(\{\alpha_k\}_{k\in\mathbb{Z}}) \triangleq -\sum_k P_k \ln(P_k), \tag{13}$$

where

$$P_k = \frac{|\alpha_k|^2}{\sum_l |\alpha_l|^2}, \tag{14}$$

[3]Recall that a partition $\mathcal{P}$ of a set $\mathcal{B}$ is a set of nonempty disjoints subsets whose union is $\mathcal{B}$.

and $\{\alpha_k\}_{k\in\mathbb{Z}}$ is the sequence of coefficients of the decomposition in a given basis. A binary tree search method was developed by Wickerhauser and Coifman [2] to find the wavelet packet which minimizes a given criterion $\mathcal{H}(\cdot)$. This algorithm requires the criterion to be additive in the sense that

$$\mathcal{H}(\{\alpha_k\}_{k\in\mathbb{Z}} \cup \{\beta_k\}_{k\in\mathbb{Z}}) = \mathcal{H}(\{\alpha_k\}_{k\in\mathbb{Z}}) + \mathcal{H}(\{\beta_k\}_{k\in\mathbb{Z}}). \qquad (15)$$

Note that the entropy criterion is not additive but, due to the orthonormality of the considered decompositions, it can be shown to be tantamount to using the additive criterion

$$\mathcal{H}_e(\{\alpha_k\}_{k\in\mathbb{Z}}) \triangleq -\sum_k |\alpha_k|^2 \ln(|\alpha_k|^2). \qquad (16)$$

The time noninvariance problem of orthonormal wavelet decompositions is also present in wavelet packet representations. We can similarly obtain a redundant wavelet packet representation which is time-invariant, as follows:

$$\widetilde{\mathcal{C}}^{\theta}_{2^j,m} \triangleq < x(t), \frac{1}{2^j} W_m(\frac{t-\theta}{2^j}) >, \quad \theta \in \mathbb{R}, j \in \mathbb{Z}, m \in \mathbb{N}. \qquad (17)$$

For ease of notation, we will omit the variable "(x)" in $\mathcal{C}^k_{j,m}(x)$, $\widetilde{\mathcal{C}}^{\theta}_{2^j,m}(x)$, $\mathcal{W}^k_j(x)$, $\widetilde{\mathcal{W}}^{\theta}_{2^j}(x)$, $\mathcal{A}^k_j(x)$ and $\widetilde{\mathcal{A}}^k_{2^j}(x)$ whenever there is no ambiguity.

3 Reconstruction from Redundant Wavelet Coefficients

Time invariance is important in many applications and may, as previously mentioned, be achieved by way of a redundant wavelet decomposition. It is often of interest in signal processing applications to reconstruct/retrieve a signal from its perturbed[4] wavelet representation. An obvious way to do so would be to select the subset of coefficients $\{\mathcal{W}^k_j, (k,j) \in \mathbb{Z}^2\}$ from the set of redundant wavelet coefficients and reconstruct the signal from its orthonormal wavelet representation. There exist, however, many different ways to achieve this reconstruction. In particular, we will see that we can extract different orthonormal bases from the wavelet family $\{2^{-j/2}\psi[(t-\theta)/2^j], \theta \in \mathbb{R}, j \in \mathbb{Z}\}$.

[4]This pertubation is generally caused by some addition of noise or some coarse quantization process.

Proposition 1 *Let two vector spaces be defined as*

$$V_{j,p} \triangleq \text{Span}\{\phi(\frac{t-p}{2^j}-k), \quad k \in \mathbb{Z}\}, \tag{18}$$

$$O_{j,p} \triangleq \text{Span}\{\psi(\frac{t-p}{2^j}-k), \quad k \in \mathbb{Z}\}, \tag{19}$$

for $j \in \mathbb{N}$ and $p \in \{0,\ldots,2^j-1\}$. It follows that

$$V_{j,p} = V_{j+1,p} \overset{\perp}{\oplus} O_{j+1,p} = V_{j+1,p+2^j} \overset{\perp}{\oplus} O_{j+1,p+2^j}, \tag{20}$$

$\{2^{-j/2}\phi[(t-p)/2^j-k], k \in \mathbb{Z}\}$ and $\{2^{-j/2}\psi[(t-p)/2^j-k], k \in \mathbb{Z}\}$ being respectively orthonormal bases of $V_{j,p}$ and $O_{j,p}$.

Proof: By using (3)-(4), we can write

$$2^{-\frac{j+1}{2}}\phi(\frac{t-p}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} h_{l-2k}\, 2^{-\frac{j}{2}}\phi(\frac{t-p}{2^j}-l), \tag{21}$$

$$2^{-\frac{j+1}{2}}\psi(\frac{t-p}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} g_{l-2k}\, 2^{-\frac{j}{2}}\phi(\frac{t-p}{2^j}-l). \tag{22}$$

and thereby establish the same relationships between $\{\phi[(t-p)/2^j - k], k \in \mathbb{Z}\}$, $\{\phi[(t-p)/2^{j+1}-k], k \in \mathbb{Z}\}$ and $\{\psi[(t-p)/2^{j+1}-k], k \in \mathbb{Z}\}$ as those between $\{\phi(t/2^j-k), k \in \mathbb{Z}\}$, $\{\phi(t/2^{j+1}-k), k \in \mathbb{Z}\}$ and $\{\psi(t/2^{j+1}-k), k \in \mathbb{Z}\}$. The property is therefore satisfied for the index p. Similarly, Eqs. (3)-(4) straightforwardly lead to

$$2^{-\frac{j+1}{2}}\phi(\frac{t-p-2^j}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} h'_{l-2k}\, 2^{-\frac{j}{2}}\phi(\frac{t-p}{2^j}-l), \tag{23}$$

$$2^{-\frac{j+1}{2}}\psi(\frac{t-p-2^j}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} g'_{l-2k}\, 2^{-\frac{j}{2}}\phi(\frac{t-p}{2^j}-l), \tag{24}$$

where $h'_k \triangleq h_{k-1}$ and $g'_k \triangleq g_{k-1}$ satisfy exactly the same paraunitary conditions as the filters with impulse responses h_k and g_k. The desired property thus holds for the index $p+2^j$. ∎

The previous proposition states that two different orthonormal bases are possible for decomposing the space $V_{j,p}$ at the next lower resolution 2^{-j-1}. These two decompositions differ in the time-localization of the basis functions. A binary tree can be used to describe the different possible choices at each resolution level j (see Fig. 2). Each node of this tree is indexed by parameters (j,p). The redundant wavelet coefficients $\{\widetilde{\mathcal{W}}_{2^j}^k\}_{k\in\mathbb{Z}}$, $j \geq 1$, may be structured according to this tree by associating to the node (j,p), $p \in \{0,\ldots,2^j-1\}$, the set $\{\widetilde{\mathcal{W}}_{2^j}^{2^jk+p}\}_{k\in\mathbb{Z}}$. If we assume that the multiscale decomposition is performed on j_{m} levels, it is easy to check by Relation (20) that the set of functions $\{2^{-j/2}\psi[(t-p_j)/2^j - k], k \in \mathbb{Z}, 1 \le j \le j_{\mathrm{m}}\} \cup \{2^{-j_{\mathrm{m}}/2}\phi[(t-p_{j_{\mathrm{m}}})/2^{j_{\mathrm{m}}} - k], k \in \mathbb{Z}\}$ is an orthonormal basis of V_0, where for each $p_{j_{\mathrm{m}}} \in \{0,\ldots,2^{j_{\mathrm{m}}}-1\}$, p_j is the number corresponding to the j least significant bits (LSBs) in the binary representation of $p_{j_{\mathrm{m}}}$. It is clear that $2^{j_{\mathrm{m}}}$ different bases can be generated, each one being graphically represented by a path from the root to a leaf of the tree.

The above results show that there exist (at least) $2^{j_{\mathrm{m}}}$ different ways of reconstructing a given signal. According to Eqs. (21) and (22), the coefficients $\{\widetilde{\mathcal{A}}_{2^j}^{2^jk+p}\}_{k\in\mathbb{Z}}$, $p \in \{0,\ldots,2^j-1\}$, may be calculated from coefficients $\{\widetilde{\mathcal{A}}_{2^{j+1}}^{2^{j+1}k+p}\}_{k\in\mathbb{Z}}$ and $\{\widetilde{\mathcal{W}}_{2^{j+1}}^{2^{j+1}k+p}\}_{k\in\mathbb{Z}}$, in the same way as coefficients $\{\mathcal{A}_j^k\}_{k\in\mathbb{Z}}$ are obtained from coefficients $\{\mathcal{A}_{j+1}^k\}_{k\in\mathbb{Z}}$ and $\{\mathcal{W}_{j+1}^k\}_{k\in\mathbb{Z}}$. Namely, this reconstruction may be recursively achieved by using the following relation:

$$\widetilde{\mathcal{A}}_{2^j}^{2^jk+p} = \sum_{l=-\infty}^{\infty} h_{k-2l}\,\widetilde{\mathcal{A}}_{2^{j+1}}^{2^{j+1}l+p} + \sum_{l=-\infty}^{\infty} g_{k-2l}\,\widetilde{\mathcal{W}}_{2^{j+1}}^{2^{j+1}l+p}. \tag{25}$$

The well-known corresponding synthesis filter bank is given by Fig. 3. In this figure, the operator $2\uparrow$ is an interpolator by a factor 2, *i.e.* its inputs $\{e_k\}_{k\in\mathbb{Z}}$ and its output $\{s_k\}_{k\in\mathbb{Z}}$ are such that

$$s_k = \begin{cases} e_{\frac{k}{2}} & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd.} \end{cases} \tag{26}$$

According to Eqs. (23) and (24), $\{\widetilde{\mathcal{A}}_{2^j}^{2^jk+p}\}_{k\in\mathbb{Z}}$, $p \in \{0,\ldots,2^j-1\}$, may just as well be obtained via $\{\widetilde{\mathcal{A}}_{2^{j+1}}^{2^{j+1}k+p+2^j}\}_{k\in\mathbb{Z}}$ and $\{\widetilde{\mathcal{W}}_{2^{j+1}}^{2^{j+1}k+p+2^j}\}_{k\in\mathbb{Z}}$. This is achieved by carrying out the following recursion,

$$\widetilde{\mathcal{A}}_{2^j}^{2^jk+p} = \sum_{l=-\infty}^{\infty} h'_{k-2l}\,\widetilde{\mathcal{A}}_{2^{j+1}}^{2^{j+1}l+p+2^j} + \sum_{l=-\infty}^{\infty} g'_{k-2l}\,\widetilde{\mathcal{W}}_{2^{j+1}}^{2^{j+1}l+p+2^j}. \tag{27}$$

Due to the simple relation between $\{h'_k\}_{k\in\mathbb{Z}}$, $\{g'_k\}_{k\in\mathbb{Z}}$, and $\{h_k\}_{k\in\mathbb{Z}}$, $\{g_k\}_{k\in\mathbb{Z}}$, we obtain the synthesis filter bank of Fig. 4. The only difference with Fig. 3 is that the operator $2\uparrow$ has been replaced by the operator $2\uparrow'$ whose input $\{e_k\}_{k\in\mathbb{Z}}$ and output $\{s_k\}_{k\in\mathbb{Z}}$ are such that

$$s_k = \begin{cases} e_{\frac{k-1}{2}} & \text{if } k \text{ is odd} \\ 0 & \text{if } k \text{ is even} \end{cases}. \tag{28}$$

Note that the filter banks of Figs. 3 and 4 may be associated to dual analysis filter banks. The corresponding decimator by a factor 2, $2\downarrow$

(resp. $2\downarrow'$), is such that its output $\{s_k\}_{k\in\mathbb{Z}}$ is obtained from its input $\{e_k\}_{k\in\mathbb{Z}}$ by

$$s_k = e_{2k} \quad (\text{resp. } s_k = e_{2k+1}). \tag{29}$$

Retaining the even samples amounts to the decimation commonly used in any orthonormal wavelet decomposition. The above discussion, however, shows that it is always possible to keep the odd samples and still satisfy the perfect reconstruction property, by a proper modification of the interpolation scheme. This remark allows us to give Relation (20), a simple digital filtering interpretation. Having two possible choices at each resolution level (even or odd decimation), particularly clarifies the total number of orthonormal bases, 2^{j_m}.

4 Best Non Redundant Set of Wavelet Coefficients

4.1 Algorithm

In the previous section, it was proved that the original signal may be reconstructed from its redundant wavelet decomposition by selecting different sets of orthonormal coefficients. The question which quite naturally arises is how to carry out the selection. This necessitates choosing a criterion such as those discussed in Section 2.2 (*e.g.* the entropy), which would reflect the matching properties of a given representation to a signal. Upon selecting a criterion, the solution lies in devising an efficient implementation of its optimization (minimization with our conventions). To reduce the complexity of the procedure, we further impose that the criterion satisfy the additivity property given by Eq. (15).

For the sake of efficiency, we recursively evaluate the criterion $\mathcal{H}(\cdot)$ for each sequence of coefficients at a given resolution. By associating a variable $\bar{\mathcal{H}}_{j,p} \triangleq \mathcal{H}(\{\widetilde{\mathcal{W}}_{2^l}^{2^l k+p_l}\}_{k\in\mathbb{Z},1\leq l\leq j})$ to each node (j,p) of the tree of Fig. 2, where p_l is the number corresponding to the l LSBs of p, and using the additivity of $\mathcal{H}(\cdot)$, the following can be deduced:

$$\bar{\mathcal{H}}_{j+1,p} = \bar{\mathcal{H}}_{j,p} + \mathcal{H}(\{\widetilde{\mathcal{W}}_{2^{j+1}}^{2^{j+1}k+p}\}_{k\in\mathbb{Z}}), \tag{30}$$

$$\bar{\mathcal{H}}_{j+1,p+2^j} = \bar{\mathcal{H}}_{j,p} + \mathcal{H}(\{\widetilde{\mathcal{W}}_{2^{j+1}}^{2^{j+1}k+p+2^j}\}_{k\in\mathbb{Z}}), \tag{31}$$

for $j \geq 1$ and $p \in \{0,\ldots,2^j-1\}$ (with $\bar{\mathcal{H}}_{0,0} \triangleq 0$). If the number of operations in computing this criterion is assumed proportional to the data length K, the complexity of the direct approach is of the order $2^{j_m}K$, for a coarsest resolution level j_m, while for the recursive technique, it is proportional to $2K/2 + 4K/4 + \cdots + 2^{j_m}(K/2^{j_m} + K/2^{j_m}) = (j_m +$

$1)K$.[5] The recursive solution thus prevents an exponential growth in the computational cost of $\mathcal{H}(\cdot)$.

The 2^{j_m} comparisons of all $\mathcal{H}_{j,p}$ account for the rest of the computational burden. A reasonable choice of j_m thus results in a rather limited complexity.

4.2 Time-Invariance Properties

The selection of the best representation, as described in the previous section, results in a time-invariance. As will be shown below, the time-invariance property of the redundant wavelet transform (5) can thus be preserved for the orthonormal one, if one adequately chooses the basis.

Let $x(t) \in V_0$ be a signal analyzed on j_m resolution levels and let $\{\widetilde{\mathcal{W}}_{2^j}^{2^j k+p_j}(x)\}_{k\in\mathbb{Z},1\leq j\leq j_m} \cup \{\widetilde{\mathcal{A}}_{2^{j_m}}^{2^{j_m}k+p_{j_m}}(x)\}_{k\in\mathbb{Z}}$ be one of its orthonormal representation, where p_j is the number corresponding to the j LSBs of p_{j_m}. If another signal $y(t)$ is such that $y(t) = \mathcal{T}_\tau[x(t)]$, $\tau \in \mathbb{Z}$,[6] it follows from the time-invariance of the redundant wavelet decomposition that

$$\widetilde{\mathcal{W}}_{2^j}^{2^j k+p_j}(y) = \widetilde{\mathcal{W}}_{2^j}^{2^j k+p_j-\tau}(x), \quad j \in \{1,\ldots,j_m\}, \tag{32}$$

$$\widetilde{\mathcal{A}}_{2^{j_m}}^{2^{j_m}k+p_{j_m}}(y) = \widetilde{\mathcal{A}}_{2^{j_m}}^{2^{j_m}k+p_{j_m}-\tau}(x). \tag{33}$$

Proposition 2 *By writing*

$$p_j - \tau = -2^j r_j + q_j, \quad r_j \in \mathbb{Z}, q_j \in \{0,\ldots,2^j-1\}, \tag{34}$$

the wavelet decomposition of $y(t) = \mathcal{T}_\tau[x(t)]$, $\tau \in \mathbb{Z}$, results in

$$\{\widetilde{\mathcal{W}}_{2^j}^{2^j k+p_j}(y)\}_{k\in\mathbb{Z}} = \mathcal{T}_{r_j}[\{\widetilde{\mathcal{W}}_{2^j}^{2^j k+q_j}(x)\}_{k\in\mathbb{Z}}], \quad j \in \{1,\ldots,j_m\} \tag{35}$$

$$\{\widetilde{\mathcal{A}}_{2^{j_m}}^{2^{j_m}k+p_{j_m}}(y)\}_{k\in\mathbb{Z}} = \mathcal{T}_{r_{j_m}}[\{\widetilde{\mathcal{A}}_{2^{j_m}}^{2^{j_m}k+q_{j_m}}(x)\}_{k\in\mathbb{Z}}], \tag{36}$$

where q_j corresponds to the j LSBs of q_{j_m}.

Proof: We use a downward induction to prove the proposition. We assume that for index $j+1$, we have

$$q_{j+1} = \sum_{l=0}^{j} \epsilon_l 2^l, \tag{37}$$

[5] For simplicity, it is assumed that the number of samples at resolution 2^{-j} is exactly $K/2^j$, without taking into account the boundary effects of the wavelet decomposition.

[6] The restriction to integer values of the time delay is not a problem in practice as it is due to the arbitrary choice of the resolution level 0 as the highest resolution level.

where $\epsilon_j \in \{0,1\}$, $j \in \{0,\ldots,j_{\rm m}\}$, is the $(j+1)^{\rm st}$ LSB in the binary representation of $q_{j_{\rm m}}$. By using Eq. (34), we then obtain, for $j \in \{1,\ldots,j_{\rm m}-1\}$,

$$\begin{aligned} p_j - \tau &= p_j - p_{j+1} - 2^{j+1} r_{j+1} + q_{j+1} \\ &= -2^j(2r_{j+1} + \eta_j - \epsilon_j) + \sum_{l=0}^{j-1} \epsilon_l 2^l, \end{aligned} \quad (38)$$

where $\eta_j = (p_{j+1} - p_j)/2^j$ is the $(j+1)^{\rm st}$ LSB in the binary representation of $p_{j_{\rm m}}$. Since r_j and q_j are defined in a unique way by Eq. (34), we can conclude that

$$q_j = \sum_{l=0}^{j-1} \epsilon_l 2^l, \quad (39)$$

which ends the proof. ■

In light of the above result, we see that the ${p_{j_{\rm m}}}^{\rm th}$ orthonormal representation of a translated signal $y(t)$ is the ${q_{j_{\rm m}}}^{\rm th}$ representation of the original signal up to some shifts r_j of the wavelet coefficients, at each resolution level j. It is clear that, by using a time-invariant optimization criterion[7], the best representation for $y(t)$ is obtained by some shift (at each scale) of the coefficients of the best representation for $x(t)$.

5 Extension to Wavelet Packets

5.1 A Class of Orthonormal Representations

Given the importance of the time invariance in signal processing problems, together with the fact that wavelet packet bases are a generalization of wavelet bases, it is natural to explore the extendibility of the results in the previous sections.

Our approach here, is similar to that for wavelets, in the sense that we will proceed to show that there exist many possible orthonormal wavelet packet representations of a signal which can be extracted from its redundant wavelet packet decomposition. These representations are characterized by different time-localizations of the functions which form the corresponding wavelet packet basis.

Proposition 3 *Let a vector space be defined as*

$$\Omega_{j,m,p} \triangleq \mathrm{Span}\{W_m(\frac{t-p}{2^j} - k), \quad k \in \mathbb{Z}\}, \quad (40)$$

[7] A criterion is said to be time-invariant when it is not sensitive to any translation of the coefficients. Note that the entropy is a time-invariant criterion.

for $j \in \mathbb{N}$, $(m,p) \in \{0,\ldots,2^j-1\}^2$, *we have*

$$\begin{aligned}\Omega_{j,m,p} &= \Omega_{j+1,2m,p} \overset{\perp}{\oplus} \Omega_{j+1,2m+1,p} \\ &= \Omega_{j+1,2m,p+2^j} \overset{\perp}{\oplus} \Omega_{j+1,2m+1,p+2^j}, \end{aligned} \quad (41)$$

and $\{2^{-j/2}W_m[(t-p)/2^j-k], k \in \mathbb{Z}\}$ *is an orthonormal basis of* $\Omega_{j,m,p}$.

Proof : The proposition follows from Eqs. (8)-(9), which lead to

$$2^{-\frac{j+1}{2}}W_{2m}(\frac{t-p}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} h_{l-2k}\, 2^{-\frac{j}{2}}W_m(\frac{t-p}{2^j}-l), \quad (42)$$

$$2^{-\frac{j+1}{2}}W_{2m+1}(\frac{t-p}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} g_{l-2k}\, 2^{-\frac{j}{2}}W_m(\frac{t-p}{2^j}-l), \quad (43)$$

and

$$2^{-\frac{j+1}{2}}W_{2m}(\frac{t-p-2^j}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} h'_{l-2k}\, 2^{-\frac{j}{2}}W_m(\frac{t-p}{2^j}-l) \quad (44)$$

$$2^{-\frac{j+1}{2}}W_{2m+1}(\frac{t-p-2^j}{2^{j+1}}-k) = \sum_{l=-\infty}^{\infty} g'_{l-2k}\, 2^{-\frac{j}{2}}W_m(\frac{t-p}{2^j}-l) \quad (45)$$

■

As a consequence, let $\mathcal{P}$ be a partition of $[0,1[$ in intervals $I_{j,m}$, $\{2^{-j/2}W_m[(t-p_{j,m})/2^j-k], k \in \mathbb{Z}, (j,m)/I_{j,m} \in \mathcal{P}\}$ is an orthonormal basis of $\Omega_{0,0}$ if

$$p_{0,0} = 0, \quad (46)$$

$$p_{j,m} = \sum_{l=0}^{j-1} \eta_{l,\lfloor 2^{l-j}m \rfloor} 2^l, \quad \eta_{l,\lfloor 2^{l-j}m \rfloor} \in \{0,1\}, \quad (47)$$

where $\lfloor\cdot\rfloor$ denotes the greatest integer lower than its argument.[8] The previous condition is easily obtained, by recalling that at any node (j,m), for $j \geq 0$ and $m \in \{0,\ldots,2^{j+1}-1\}$, we have

$$p_{j+1,m} = p_{j,\lfloor m/2 \rfloor} + \eta_{j+1,\lfloor m/2 \rfloor} 2^j. \quad (48)$$

The latter result may also be interpreted as the possibility of choosing either even or odd decimations in the filter bank implementation of the usual wavelet packet decomposition. It is clear that for a given filter bank structure, we can generate 2^κ different orthonormal bases, κ being the number of nodes in the frequency tree characterizing the

[8]The above equation implies that $p_{j,2m} = p_{j,2m+1}$, for $j \geq 1$ and $m \in \{0,\ldots,2^{j-1}-1\}$.

wavelet packet decomposition considered. It follows that the maximal number of possibilities is $2^{2^{j_m}-1}$, which corresponds to an equal subband analysis. Note that, unlike the wavelet case, the different orthonormal bases cannot generally be represented by a binary tree, because of the additional flexibility in the analysis provided by the parameter m.

5.2 Optimization of the Representation

As in the case of orthonormal wavelet decompositions, the goal is to reconstruct in the best way possible a signal from its complete decomposition (*i.e.* find the optimal (j, m, p) triplets). A direct approach would be prohibitive and result in an impractical solution. In what follows we proceed by (i) determining the best filter bank structure for a given signal, then (ii) obtaining the optimal time-localization parameters.

In the first step of the algorithm, we proceed very similarly to Wickerhauser's method to find the wavelet packet best matched to the analyzed signal. Let $\mathcal{H}(\cdot)$ be an additive criterion to be minimized. If $\mathcal{H}_{j,m} \triangleq \mathcal{H}(\{\widetilde{C}^k_{2^j,m}\}_{k\in\mathbb{Z}})$ and $\mathcal{P}$ denotes the optimal partition of $[0,1[$, the algorithm may be summarized as follows:

- $\forall\, m \in \{0,\ldots,2^{j_m}-1\}$, $\widetilde{\mathcal{H}}_{j_m,m} \triangleq \mathcal{H}_{j_m,m}$;
- $\forall\, j \in \{j_m - 1,\ldots,0\}$,

  $\forall\, m \in \{0,\ldots,2^j-1\}$,

  if $\mathcal{H}_{j,m} \leq \frac{1}{2}(\widetilde{\mathcal{H}}_{j+1,2m} + \widetilde{\mathcal{H}}_{j+1,2m+1})$

  $\widetilde{\mathcal{P}}_{j,m} \triangleq \{I_{j,m}\}$,

  $\widetilde{\mathcal{H}}_{j,m} \triangleq \mathcal{H}_{j,m}$,

  otherwise

  $\widetilde{\mathcal{P}}_{j,m} \triangleq \widetilde{\mathcal{P}}_{j+1,2m} \cup \widetilde{\mathcal{P}}_{j+1,2m+1}$,

  $\widetilde{\mathcal{H}}_{j,m} \triangleq \frac{1}{2}(\widetilde{\mathcal{H}}_{j+1,2m} + \widetilde{\mathcal{H}}_{j+1,2m+1})$;
- $\mathcal{P} = \widetilde{\mathcal{P}}_{0,0}$.

The only difference with Wickerhauser's algorithm lies in the use of a redundant wavelet packet decomposition which is also reflected by the 1/2 scaling factor. This variation on the algorithm can be seen to be equivalent to averaging $\mathcal{H}(\{\widetilde{C}^{2^jk+p}_{2^j,m}\}_{k\in\mathbb{Z}})$ over $p \in \{0,\ldots,2^j-1\}$, for each $j \in \mathbb{N}$ and $m \in \{0,\ldots,2^j-1\}$, Following the optimization in the structure, we proceed to obtain the time-localization parameters of the basis functions. As a result we have to compare the 2^κ different orthonormal representations, (κ being the number of nodes in the frequency tree corresponding to the partition $\mathcal{P}$). This entails computing

$$\sum_{(j,m)/I_{j,m}\in\mathcal{P}} \mathcal{H}\left(\{\tilde{\mathcal{C}}_{2^j,m}^{2^j k+p_{j,m}}\}_{k\in\mathbb{Z}}\right),$$

for all sets of integers $p_{j,m}$ which can be expressed in the the form of Eqs. (46)-(47). Because of the generality of the tree structure, it is unfortunately difficult to obtain a recursion to reduce the computational complexity. This says that for a data length K, we have $2^\kappa K$ operations to compute the criterion and $2^\kappa - 1$ comparisons to make. The computational burden may rise quite rapidly unless j_{m} is reasonable.

5.3 Time-Invariance

Much like orthonormal wavelet decompositions, a time-invariance property may be derived for wavelet packet decompositions.

Proposition 4 *Let $x(t)$ and $y(t)$ be two continuous time signals such that*

$$\mathcal{C}_{0,0}^k(y) = \mathcal{T}_\tau[\mathcal{C}_{0,0}^k(x)], \quad k\in\mathbb{Z}, \tag{49}$$

and $\tau\in\mathbb{Z}$, we then have, for all $j\in\mathbb{N}$ and $m\in\{0,\ldots,2^j-1\}$,

$$\tilde{\mathcal{C}}_{2^j,m}^{2^j k+p_{j,m}}(y) = \mathcal{T}_{r_{j,m}}[\tilde{\mathcal{C}}_{2^j,m}^{2^j k+q_{j,m}}(x)], \quad k\in\mathbb{Z}, \tag{50}$$

where the integers $p_{j,m}$ satisfy Eqs. (46)-(47) and the integers $q_{j,m}$ are defined as follows:

$$q_{0,0} \triangleq 0, \tag{51}$$

and

$$q_{j,m} \triangleq \sum_{l=0}^{j-1} \epsilon_{l,\lfloor 2^{l-j}m\rfloor} 2^l, \quad \epsilon_{l,\lfloor 2^{l-j}m\rfloor}\in\{0,1\}, \tag{52}$$

when $j\geq 1$, whereas the integers $r_{j,m}$ are such that

$$p_{j,m} - \tau = -2^j r_{j,m} + q_{j,m}. \tag{53}$$

Proof : First, note that there exist unique $r_{j,m}$ and $q_{j,m}\in\{0,\ldots,2^j-1\}$ satisfying Eq. (53). One can further note that Eq. (50) is equivalent to

$$\tilde{\mathcal{C}}_{2^j,m}^{2^j k+p_{j,m}}(y) = \tilde{\mathcal{C}}_{2^j,m}^{2^j k+p_{j,m}-\tau}(x), \tag{54}$$

by recalling the time-invariance of the redundant wavelet packet decomposition. To complete the proof, we will show by induction that $q_{j,m}$

may be expressed as in Eqs. (51)-(52). The property is obviously satisfied when $j = m = 0$. We proceed to prove that the property being satisfied for the indices (j, m), implies that it is also satisfied for the indices $(j+1, 2m)$ and $(j+1, 2m+1)$. We use Eq. (53) to write

$$\begin{aligned} p_{j+1,2m} - \tau &= 2^j(\eta_{j,m} - r_{j,m}) + q_{j,m} \\ &= -2^{j+1} r'_{j+1,2m} + \sum_{l=0}^{j} \epsilon_{l,\lfloor 2^{l-j} m \rfloor} 2^l, \end{aligned} \tag{55}$$

where the integers $r'_{j+1,2m}$ and $\epsilon_{j,m}$ are defined by

$$\eta_{j,m} - r_{j,m} \triangleq -2r'_{j+1,2m} + \epsilon_{j,m}, \quad \epsilon_{j,m} \in \{0, 1\}. \tag{56}$$

Due to the uniqueness of representation (53), we can conclude that

$$q_{j+1,2m} = \sum_{l=0}^{j} \epsilon_{l,\lfloor 2^{l-j} m \rfloor} 2^l. \tag{57}$$

The property is also satisfied for indices $(j+1, 2m+1)$ as a result of the equality of $p_{j+1,2m}$ and $p_{j+1,2m+1}$ which implies that $r_{j+1,2m+1} = r_{j+1,2m}$ and $q_{j+1,2m+1} = q_{j+1,2m}$. ■

The above results clearly show that, by using a time-invariant criterion (*e.g.* entropy), one can generate a time-invariant wavelet packet representation. This is to say that the optimal representation corresponding to a time-shifted signal is derived from the optimal representation of the original signal by translating each set of wavelet packet coefficients corresponding to the indices (j, m) by $r_{j,m}$.

6 Applications

6.1 Estimation of Noisy Transients

In the same way as signal subspace approximation is used in spectral estimation [17], thresholding of the wavelet coefficients has been proposed in the literature [5, 13, 8, 14, 7] as a means to enhance estimation of an unknown signal in noise. The key idea in fact lies in the ability of wavelet transforms to compress most of the useful information of the signal and spread that of the noise. The compact representation of the signal information then plays a key role in discriminating against the noise, so long as the energy of the latter does not overwhelm the signal components.

In this section, we demonstrate the importance of the time-invariance property of a multiscale representation and its effect on estimation problems. We assume an observed signal $y(t) \in \mathbb{R}$ as the sum of a signal $x(t) \in \mathbb{R}$ to be estimated and of a stationary, zero-mean normal noise $b(t)$. Without loss of generality, we will assume that the power spectral density (PSD) of the noise is 1. Because of the linearity of the considered transforms, we also have an additive noise model for the wavelet packet coefficients:

$$\tilde{C}_{2^j,m}^{2^j k+p_{j,m}}(y) = \tilde{C}_{2^j,m}^{2^j k+p_{j,m}}(x) + \tilde{C}_{2^j,m}^{2^j k+p_{j,m}}(b), \quad 0 \leq j \leq j_{\mathrm{m}}, 0 \leq m \leq 2^j - 1, \tag{58}$$

where j_{m} designates the coarsest resolution level of decomposition and the $p_{j,m}$'s are integers satisfying Eqs. (46)-(47).[9] Since the representations of interest are furthermore orthonormal, the random variables $\{\tilde{C}_{2^j,m}^{2^j k+p_{j,m}}(b)\}_{k\in\mathbb{Z},(j,m)/I_{j,m}\in\mathcal{P}}$ are i.i.d. N(0,1), for any partition $\mathcal{P}$ of $[0,1[$ in intervals $I_{j,m}$. Note that the mean square estimation error of $x(t)$ integrated over time, is equal to the sum of the mean square estimation errors of all its wavelet packet components, because of the orthonormality.

The coefficients of the signal $x(t)$ are nonlinearly estimated by using a thresholding technique, as follows:

$$\hat{\tilde{C}}_{2^j,m}^{2^j k+p_{j,m}}(x) \triangleq \begin{cases} \tilde{C}_{2^j,m}^{2^j k+p_{j,m}}(y) & \text{if } |\, \tilde{C}_{2^j,m}^{2^j k+p_{j,m}}(y) \,| \geq \xi \\ 0 & \text{otherwise} \end{cases}, \tag{59}$$

where $\xi \geq 0$ denotes the threshold level. Note that the method may appear as an extension of the matched filtering approach allowing one to cope with the case where the signal of interest is unknown. The difficulty is then to select an adequate threshold ξ. This choice is the result of a trade-off between

- a good reduction of the noise,
- the preservation of the main signal components.

A problem which also arizes when dealing with noisy signals, is the evaluation of the criterion used to find the best basis. As described in the Appendix, it is possible to modify the entropy criterion to avoid the overwhelming effect of the noise.

[9] The wavelet formulation results straightforwardly as it is a special case of that of the wavelet packet.

We provide two simulation examples of noisy signals to substantiate the foredescribed nonlinear estimation procedure. In both cases, the signal to noise ratio and the normalized mean square estimation error are defined respectively as

$$\mathrm{SNR} = \frac{\sum_{k=0}^{K-1} \mathcal{A}_0^k(x)^2}{K \mathrm{Var}\{\mathcal{A}_0^k(\mathrm{b})\}} = \frac{1}{K} \sum_{k=0}^{K-1} \mathcal{A}_0^k(x)^2 \tag{60}$$

$$\mathrm{NMSE} = \frac{\sum_{k=0}^{K-1} [\mathcal{A}_0^k(x) - \hat{\mathcal{A}}_0^k(x)]^2}{\sum_{k=0}^{K-1} \mathcal{A}_0^k(x)^2}, \tag{61}$$

where $\{\hat{\mathcal{A}}_0^k(x)\}_{0 \leq k < K}$ designate the estimates of $\{\mathcal{A}_0^k(x)\}_{0 \leq k < K}$ and K is the number of samples.[10] In our simulations, $K = 256$ and the threshold level ξ is equal to 4. Daubechies filters [3] with 8 coefficients (4 vanishing moments) are used to construct the wavelet and wavelet packet decompositions. The signal is analyzed on 4 resolution levels ($j_{\mathrm{m}} = 4$).
Example 1 The signal of interest is the sum of two wavelet packets:

$$x(t) = \frac{15}{2}[\frac{1}{2}\psi(\frac{t-63}{2}) + \frac{1}{8}W_2(\frac{t-191}{8})]. \tag{62}$$

This signal corresponds to a somewhat ideal case for the wavelet packet representation.[11] Fig. 5 shows the original signal and the noisy signal with an SNR $= -3.571$ dB as well as the temporal estimate which corresponds to the direct thresholding of the signal samples. Fig. 6 shows the estimates obtained by using the usual wavelet and wavelet packet decompositions and the time-invariant wavelet and wavelet packet representations. The gain in performance which appears in this example is confirmed by a Monte Carlo study involving 500 different realizations of the noise. The results are summarized in Table 1.
Example 2 The signal to be estimated is such that

$$\mathcal{A}_0^k(x) = 8.5 \exp(-0.072 \mid k - 123 \mid) \cos(\frac{2\pi}{5} \mid k - 123 \mid), \tag{63}$$

and the SNR is now 3.537 dB. Table 1 illustrates the improvement resulting from the use of the time invariant representations.

6.2 Time Delay Estimation

The time-invariance property may also be useful in time delay estimation problems. We address the problem of estimating a time delay $\tau \in \mathbb{Z}$ from

[10] As is common in the pratice of wavelet decompositions, the signal $x(t)$ is assumed to belong to V_0.

[11] Such a transient is clearly unrealistic in most applications unless we have the ability to choose the signal.

two observed processes, which may be assumed to be measurements of two sensors:

$$\mathrm{y}_n(t) \triangleq x_n(t) + \mathrm{b}_n(t), \quad n \in \{1, 2\}, \tag{64}$$

where $x_1(t)$ is an unknow signal, $x_2(t) \triangleq x_1(t - \tau)$ is its delayed version and $\mathrm{b}_1(t)$ and $\mathrm{b}_2(t)$ are two uncorrelated stationary, zero-mean Gaussian noise processes.

In this case, we use a wavelet decomposition to illustrate this application. As was explained in the previous section, estimates of the wavelet coefficients of $x_1(t)$ and $x_2(t)$ are obtained, by finding the best wavelet representations of these signals (from $\mathrm{y}_1(t)$ and $\mathrm{y}_2(t)$) and using a thresholding technique. It is then possible to estimate τ by making use of Relation (34).[12] In light of (34), the parameters r_j corresponds to the shift between the wavelet coefficients of $x_1(t)$ and $x_2(t)$, for a given resolution level j, and may therefore be determined by a classical correlation method (by finding the maximum absolute value of the correlation sequence). In this way, Eq. (34) allows us to calculate an estimate $\hat{\tau}_j$ of τ, for each resolution level. We then proceed to compute a global estimate of τ as a weighted average of these estimates. The weighting factors are defined as the sum of the energies of the estimated wavelet coefficients of $x_1(t)$ and $x_2(t)$ at each resolution level.[13] It is worth noting that this approach does not require the reconstruction of the signals from their wavelet coefficients. So, the computational load remains low.

Example 3 The reference signal is (in V_0) such that

$$\mathcal{A}_0^k(x_1) = 4.3 \exp[-0.019(k - 128)^2] \cos[\frac{\pi}{5}(k - 128)], \tag{65}$$

and the delay is $\tau = 10$. The PSDs of the noises are normalized to 1. Daubechies filter with 8 coefficients are used to carry out wavelet decompositions on 3 resolution levels. The estimates of the delay obtained by the proposed method are graphically represented in Fig. 7. The mean square estimation error evaluated with 150 realizations is $\mathrm{E}\{(\hat{\tau} - \tau)^2\} = 1.207$. If the delay is estimated by simply picking the maximum value of the correlation function of $\mathrm{y}_1(t)$ and $\mathrm{y}_2(t)$, poorer results are obtained as illustrated in Fig. 7. We then have $\mathrm{E}\{(\hat{\tau} - \tau)^2\} = 4.947$.

[12] The last statement holds provided that a unique best representation exists for both $x_1(t)$ and $x_2(t)$.

[13] When the energy of either $x_1(t)$ or $x_2(t)$ vanishes, the corresponding resolution level is not taken into account.

7 Conclusion

In this paper, we have shown that a class of orthonormal wavelet/wavelet packet decompositions may be obtained by varying the time localizations of the basis functions. By selecting the set which minimizes a proper energy concentration criterion (*e. g.* the entropy), we obtain an optimized representation of the analyzed signal. The appeal of this approach lies in the resulting time (space) invariance of the orthonormal decomposition of a given process. As illustrated in the examples, a substantial performance improvement in estimation/detection problems involving noisy transients may be achieved. This is obtained at the expense of an increased computational complexity.

Appendix: Optimal Representation in the Presence of Noise

To simplify the discussion of the thresholding procedure, we reformulate the problem of signal estimation as follows: we want to estimate an unknown deterministic sequence $\{\alpha_k\}_{1\leq k\leq K}$ from observations $\{\gamma_k\}_{1\leq k\leq K}$ such that,

$$\gamma_k \triangleq \alpha_k + \beta_k, \tag{66}$$

where $\{\beta_k\}_{1\leq k\leq \mathbb{Z}}$ are i.i.d. N(0,1) random variables.[14] We further have

$$\alpha_k \neq 0 \quad \text{iff } k \in \{\tilde{k}_1, \ldots, \tilde{k}_L\}, \tag{67}$$

where $\{\tilde{k}_l\}_{1\leq l\leq K}$ is a sequence obtained by reindexing $\{1,\ldots,K\}$ and $L \in \{1,\ldots,K\}$. In our case, these last quantities are also unknown. When dealing with noisy signals, it must be kept in mind that our goal is to optimize the representation of *the signal to be estimated.* Considering the (non normalized) entropy criterion, directly calculated from the observed signal, we obtain

$$\mathcal{H}_{\mathrm{e}}(\{\gamma_k\}_{1\leq k\leq K}) = \sum_{l=1}^{L} \mathcal{H}_{\mathrm{e}}(\alpha_{\tilde{k}_l} + \beta_{\tilde{k}_l}) + \sum_{l=L+1}^{K} \mathcal{H}_{\mathrm{e}}(\beta_{\tilde{k}_l}). \tag{68}$$

If the noise tends to be dominant (*i.e.* $K \gg L$), it is clear that the second element in the above equation may swamp out the first and thereby perturb the optimal search. Note that the definition of the entropy in this context is a random variable in contrast to that given in an information theoretic setting. Expressions of the mean and the variance of the noise term of the entropy may be found in [7].

It is therefore useful to find a criterion which is less sensitive to noise. This can be achieved by first noting that thresholding the observed data by $\xi > 1$ is in fact equivalent to minimizing the criterion

$$\mathcal{J}(\emptyset) \triangleq 0, \tag{69}$$

$$\mathcal{J}(\{k_l\}_{1\leq l\leq M}) \triangleq \sum_{l=1}^{M} \mathcal{H}_{\mathrm{e}}(\gamma_{k_l}) + M\xi^2 \ln(\xi^2), \tag{70}$$

defined for every $M \in \{1,\ldots,K\}$ and every sequence $\{k_l\}_{1\leq l\leq M}$ extracted from $\{1,\ldots,K\}$. The minimal value of $\mathcal{J}(\cdot)$ is indeed, obtained for the set of integers $\mathcal{E}_\xi \triangleq \{k \in \{1,\ldots,K\}/ \mid \gamma_k \mid > \xi\}$.[15] We then

[14] A finite number K of variables was obtained, by assuming that the signal is observed during a finite time.

[15] The property follows from the fact that $\mathcal{H}_{\mathrm{e}}(\gamma_k) + \xi^2 \ln(\xi^2) < 0$ iff $\mid \gamma_k \mid > \xi$, provided that $\xi > 1$.

propose to use the minimal value of $\mathcal{J}(\cdot)$ as a criterion to compare the different representations of the noisy signal:

$$\mathcal{H}_{\mathrm{n}}(\{\gamma_k\}_{1\leq k\leq K}) \triangleq \mathcal{J}(\mathcal{E}_\xi) = \sum_{k/|\gamma_k|>\xi} \mathcal{H}_{\mathrm{e}}(\gamma_k) + \mu(\mathcal{E}_\xi)\xi^2 \ln(\xi^2), \qquad (71)$$

where $\mu(\mathcal{E}_\xi)$ denotes the cardinality of $\mathcal{E}_\xi$. The first term in the right member of Eq. (70) is the entropy of the thresholded sequence, whereas the second term may be interpreted as a regularization term favoring a small number of selected components. It is also useful to note that the criterion $\mathcal{H}_{\mathrm{n}}(\cdot)$ is additive and time-invariant.

References

[1] A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. to appear in Com. Pure and Appl. Math., 1990.

[2] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory*, IT-38:713–718, Mar. 1992.

[3] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Com. Pure and Appl. Math.*, XLI:909–996, 1988.

[4] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory*, IT-36:961–1005, Sep. Sep. 1990.

[5] D. L. Donoho and I. M Johnstone. Ideal spatial adaptation by wavelet shrinkage. preprint Dept. Stat., Stanford Univ., Jun. 1992.

[6] A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal*, 15:723–736, 1984.

[7] H. Krim and J.-C. Pesquet. Robust multiscale representation of processes and optimal signal reconstruction. in preparation.

[8] B. Lumeau, J.-C. Pesquet, J.-F. Bercher, and L. Louveau. Optimisation of bias-variance tradeoff in non parametric spectral analysis by decomposition into wavelet packet. In Y. Meyer and S. Roques, editors, *Progress in wavelet analysis and applications*, pages 285–290. Editions Frontieres, Gif sur Yvette, 1993.

[9] S. Mallat. Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315:69–87, Sep. 1989.

[10] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, PAMI-11:674–693, Jul. 1989.

[11] S. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theory*, IT-38:617–643, Mar. 1992.

[12] S. Mallat and Z. Zhang. Adaptive time-frequency decomposition with matching pursuits. In *Proceedings of IEEE Symposium on Time-Frequency and Time-Scale Analysis*, pages 7–10, Victoria, Oct. 4-6 1992.

[13] P. Moulin. Wavelet thresholding techniques for power spectrum estimation. submitted to IEEE Trans. Signal Processing, Aug. 1992.

[14] P. Moulin. A wavelet regularization method for diffuse radar target imaging and speckle noise reduction. *Journ. Math. Imaging and Vision*, 3:123–134, 1993.

[15] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multiscale transforms. *IEEE Trans. Inform. Theory*, IT-38:587–607, Mar. 1992.

[16] B. Torrésani. Time-frequency representations: wavelet packets and optimal decomposition. preprint, 1990.

[17] D. W. Tufts and R. Kumaresan. Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood. *IEEE Proceedings*, Sep. 1982. Spectrum Estimation Issue.

[18] P. P. Vaidyanathan. *Multirate systems and filter banks.* Prentice Hall, New Jersey, 1992.

[19] M. V. Wickerhauser. INRIA lectures on wavelet packet algorithms. In *Ondelettes et paquets d'ondelettes*, pages 31–99, Roquencourt, Jun. 17-21 1991.

List of Tables

List of Figures

	temporal	wavelets	time inv. wavelets	wav. pack.	time inv. wav. pack.
example 1	0.6037	0.6887	0.1720	0.6096	0.0875
example 2	0.3346	0.2854	0.2404	0.2150	0.1501

Table 1: NMSE in the estimation of noisy transients averaged on 500 realizations.

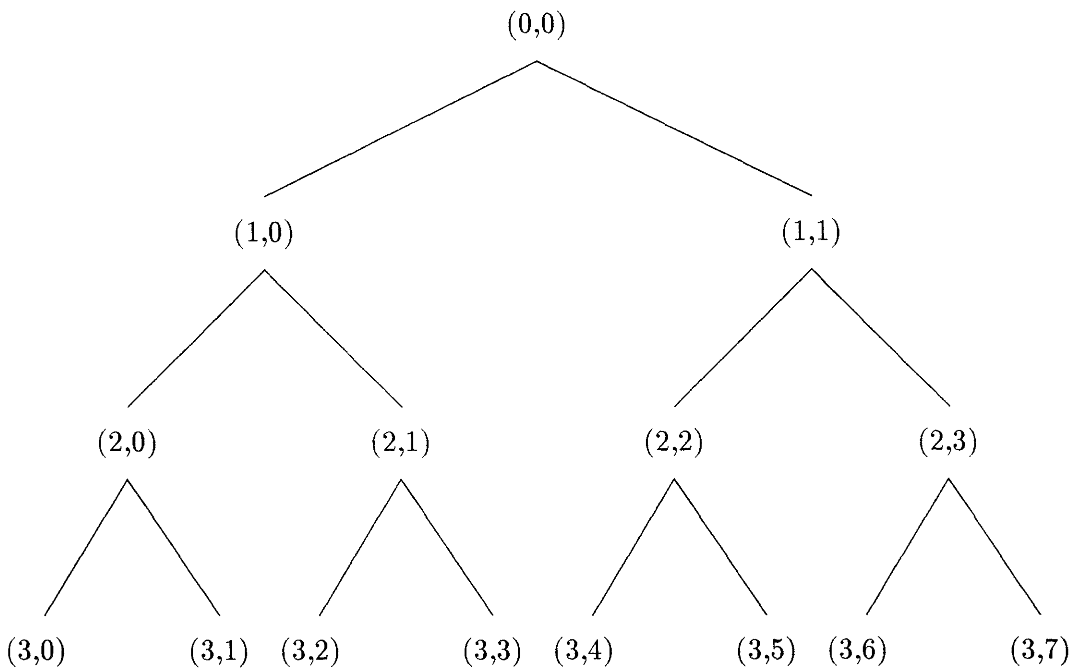

Figure 1: Frequency tree of a 3 resolution level equal subband decomposition. (Nodes are indexed by parameters (j, m).)

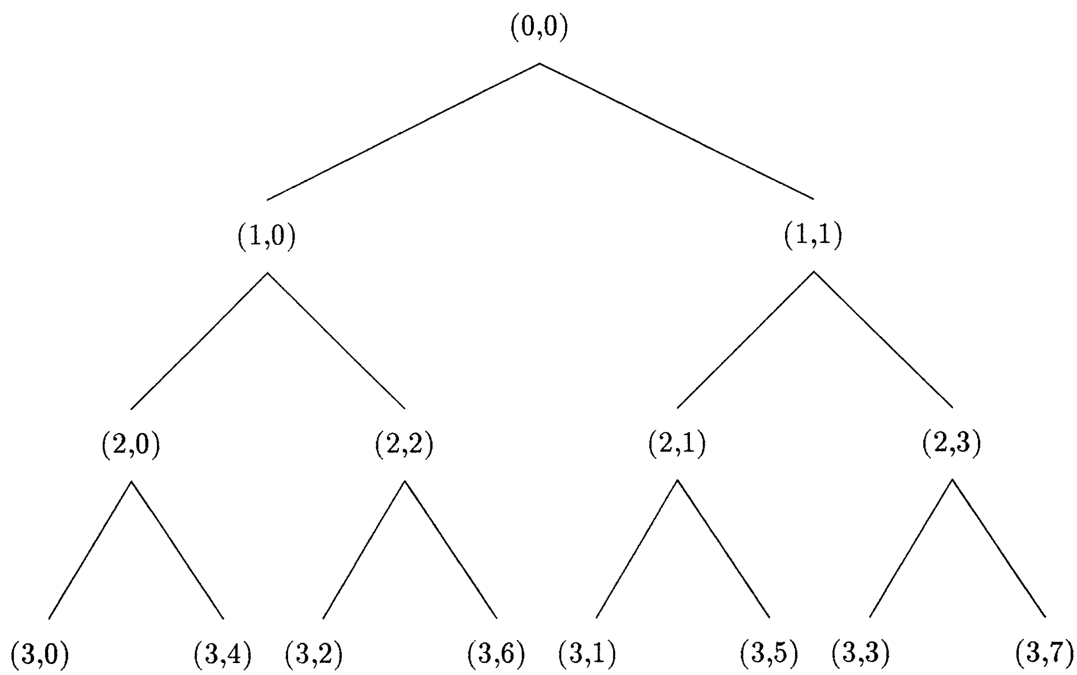

Figure 2: Time-localization tree for a 3 resolution level wavelet analysis. (Nodes are indexed by parameters (j, p).)

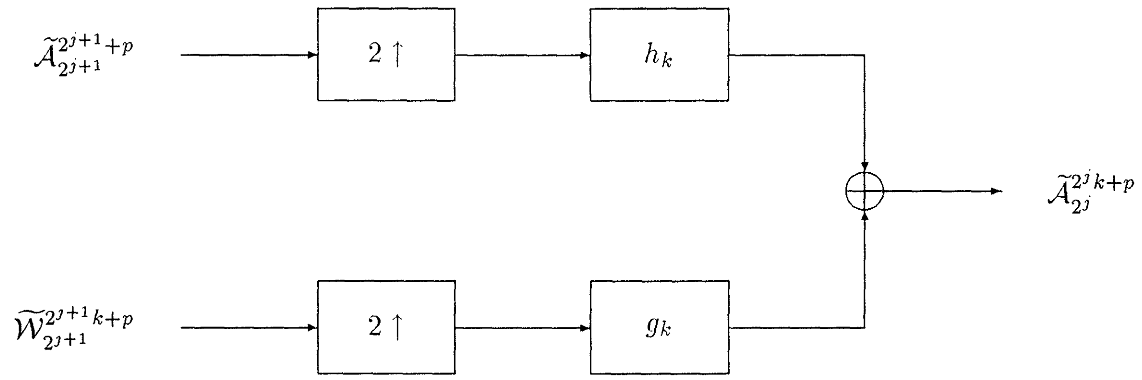

Figure 3: Synthesis filter bank for signal reconstruction from its wavelet coefficients, when $0 \leq p < 2^j$.

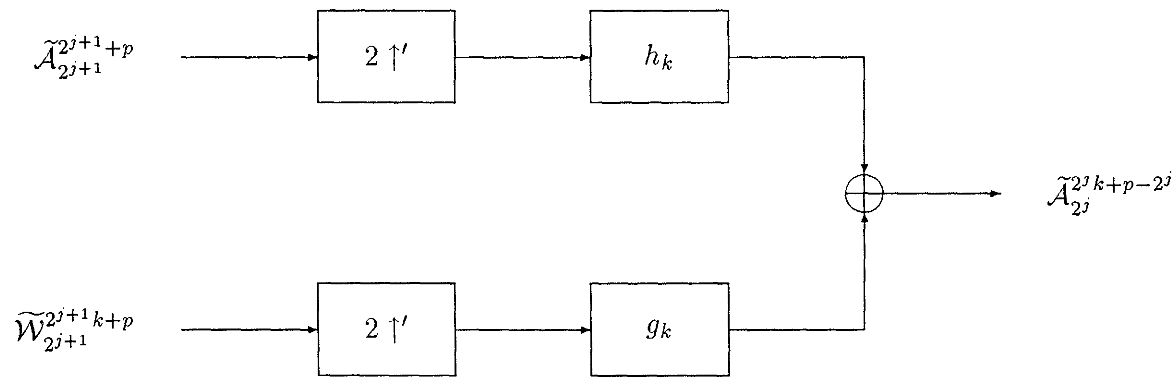

Figure 4: Synthesis filter bank for signal reconstruction from its wavelet coefficients, when $2^j \leq p < 2^{j+1}$.

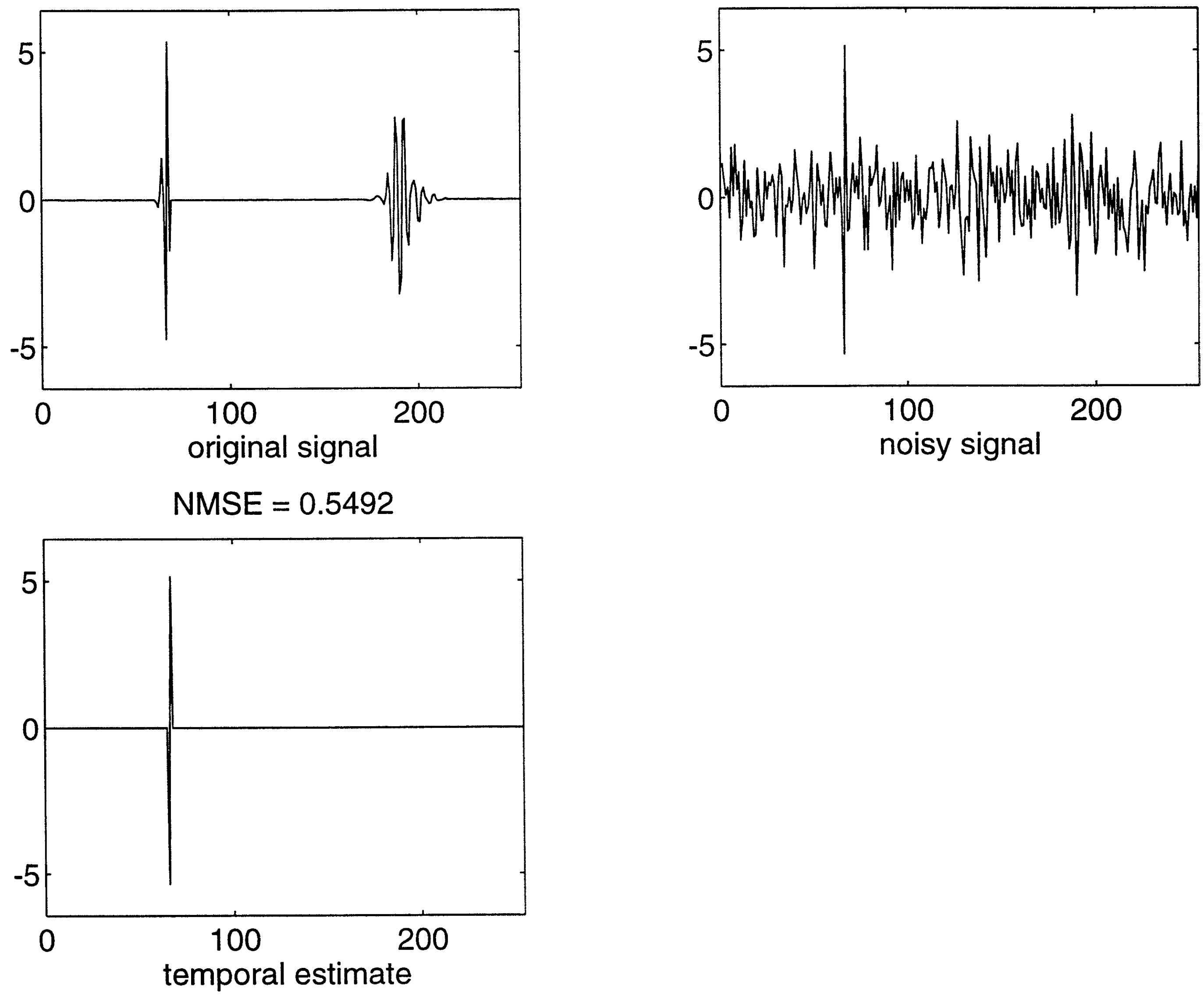

Figure 5: First example of noisy transients (SNR = -3.571 dB).

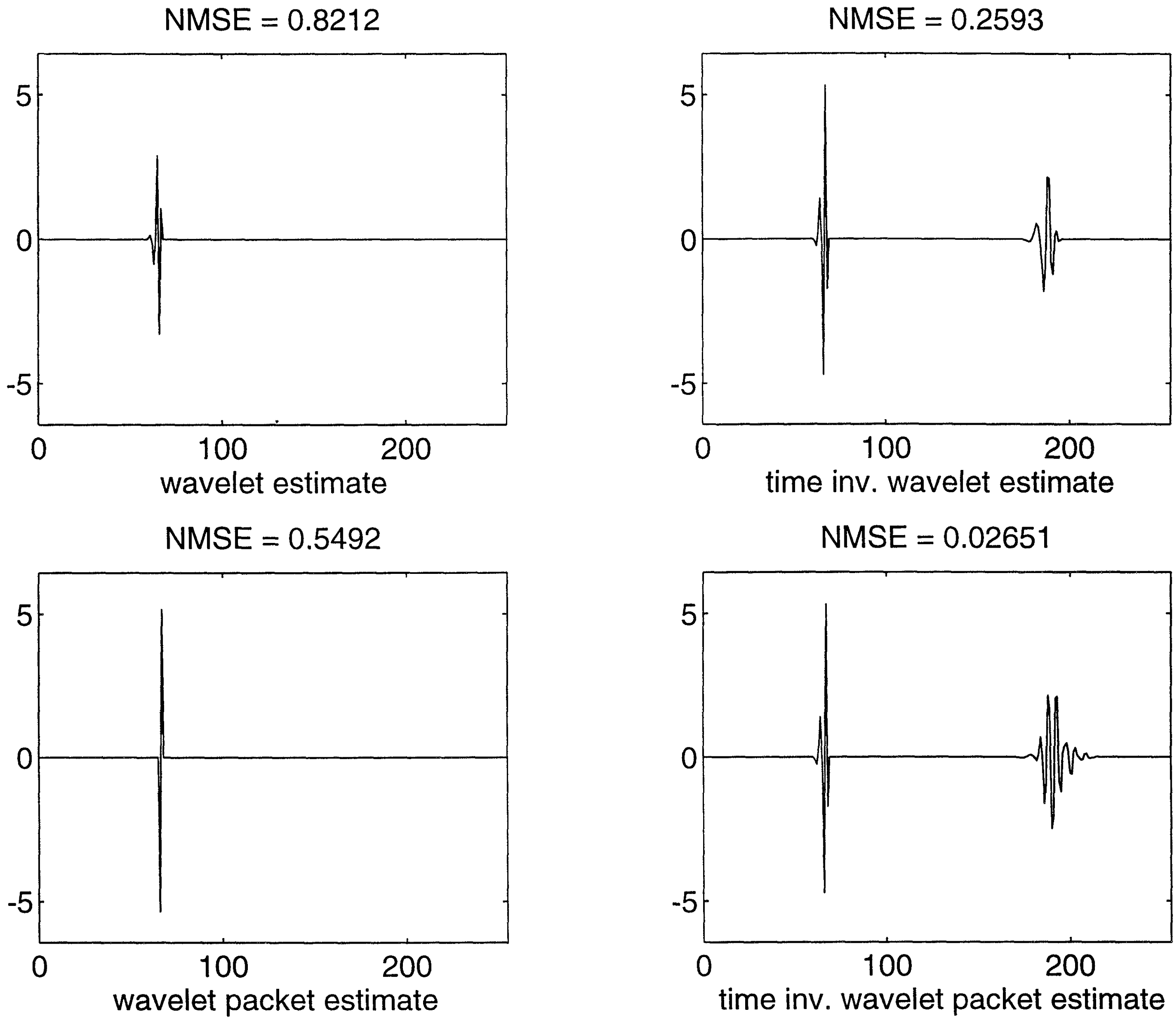

Figure 6: Estimation of the signal in Fig. 5.

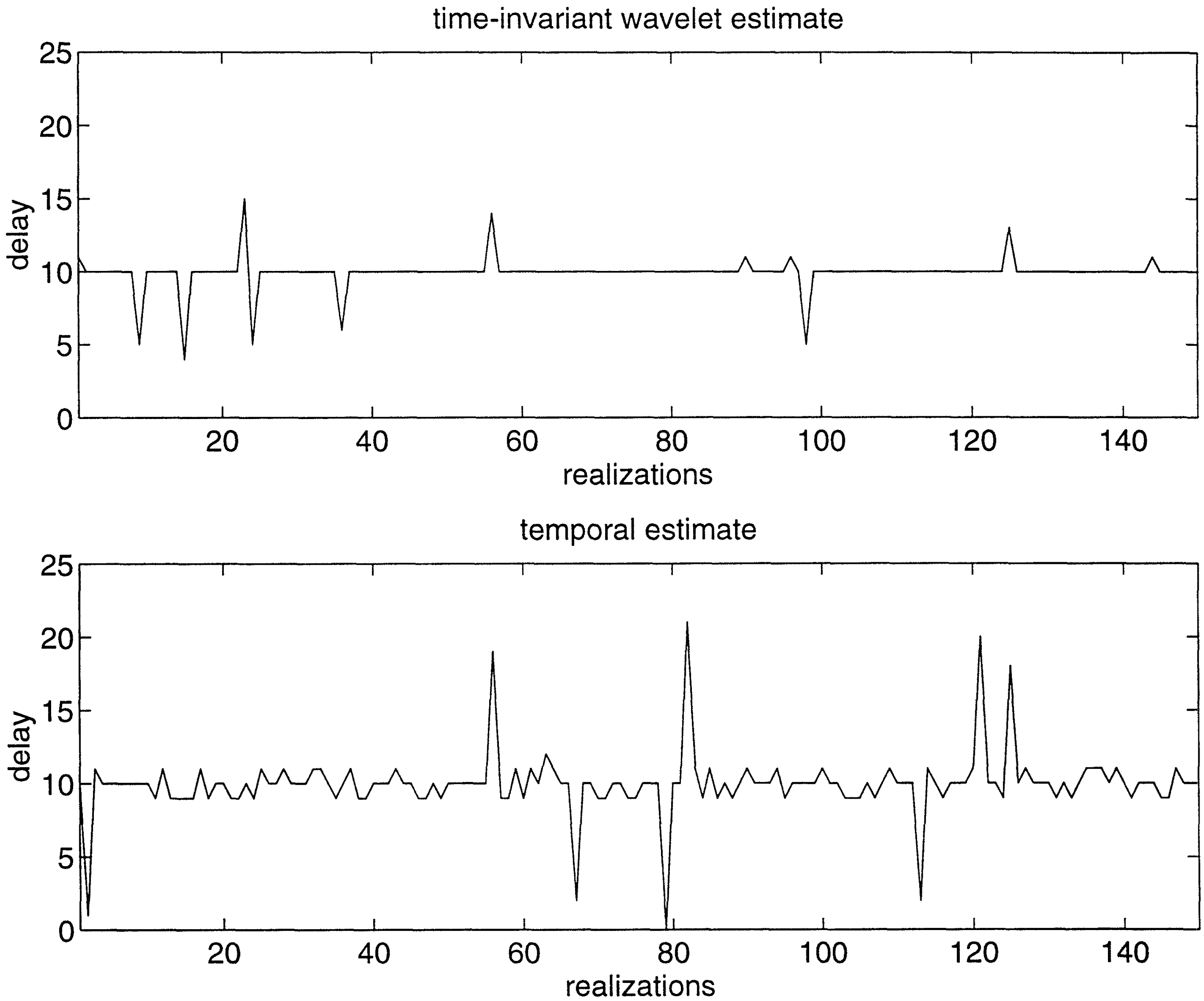

Figure 7: Delay estimates using a time invariant wavelet representation and a classical correlation method.