# TECHNICAL RESEARCH REPORT

Optimal Wavelet Basis Selection for Signal Representation

*by Y. Zhuang, J.S. Baras*

**CSHCN T.R. 94-7**
**(ISR T.R. 94-3)**

CENTER FOR SATELLITE AND HYBRID COMMUNICATION NETWORKS

**Optimal wavelet basis selection for signal representation**

Yan Zhuang [1] and John S. Baras [2]
Institute for Systems Research and
Department of Electrical Engineering
The University of Maryland, College Park, MD 20742

**ABSTRACT**

We study the problem of choosing the optimal wavelet basis with compact support for signal representation and provide a general algorithm for computing the optimal wavelet basis. We first briefly review the multiresolution property of wavelet decomposition and the conditions for generating a basis of compactly supported discrete wavelets in terms of properties of quadrature mirror filter (QMF) banks. We then parametrize the mother wavelet and scaling function through a set of real coefficients. We further introduce the concept of information measure as a distance measure between the signal and its projection onto the subspace spanned by the wavelet basis in which the signal is to be reconstructed. The optimal basis for a given signal is obtained through minimizing this information measure. We have obtained explicitly the sensitivity of dilations and shifts of the mother wavelet with respect to the coefficient set. A systematic approach is developed here to derive the information gradient with respect to the parameter set for a given square integrable signal and the optimal wavelet basis. A gradient based optimization algorithm is developed in this paper for computing the optimal wavelet basis.

**Keywords:** wavelets, data compression, signal processing.

# 1 INTRODUCTION

The last few years have witnessed extensive research interest and activities in wavelet theory and its applications in signal processing, image processing and many other fields [1, 2]. The most attractive features of wavelet theory are the multiresolution property and time and frequency localization ability. The wavelet transform decomposes a signal to its components at different resolutions. Its application actually simplifies the description of signals and provides analysis at different levels of detail. There are some successful applications of these properties in the fields of signal processing, speech processing and especially in image processing [3, 4, 5]. Wavelet transform differs from short-time Fourier transform (STFT) in the sense of producing a varying time-frequency window for signal representation. It admits nonuniform bandwidths, so that the bandwidth is higher at higher frequencies, which makes it possible to implement the wavelet transform through different levels of decimation in filter banks.

We know that wavelet functions can be used for function approximation and finite energy signal representations which are useful in signal processing and system identification. The wavelet basis is generated by dilating and shifting a single mother wavelet function $\psi(t)$. The wavelet function is not unique and its design can be related to that of a power symmetric FIR low pass filter. Obviously, different wavelets $\psi(t)$ shall yield different wavelet bases. However, appropriate selection of the wavelet for signal representation can result to maximal benefits of this new technique. Different wavelet functions may be suitable for different signals or functions to be represented or to be approximated. It is reasonable to think that if a wavelet contains enough information about a signal to be represented, the wavelet system is going to be simplified in terms of the level of required resolution, which reduces the computational complexity of the problem for system implementation. This paper addresses the issue of finding the optimal mother wavelet function to span the appropriate feature space for signal representation.

The key to choosing the optimal wavelet basis lies in the appropriate parameterization and the adequate performance measure in addition to the accurate interpretation of physical phenomena. A method was proposed for choosing a wavelet for signal representation based on minimizing an upper bound of the $L^2$ norm of error [6, 7] in approximating the signal up to the desired scale. Coifman et al. derived an entropy based algorithm for selecting the best basis from a library of wavelet packets [8]. However, a direct method to systematically generate the optimal orthonormal discrete wavelet basis with compact support has not been developed as yet. We proposed

[1]Email: yzhuang@src.umd.edu. Tel.: (301) 405-6578.
[2]Martin Marietta Chair in Systems Engineering. Email: baras@src.umd.edu. Tel.: (301) 405-6606.

an information measure based approach for constructing the optimal discrete wavelet basis with compact support in our earlier work on adaptive wavelet neural networks [9]. We shall provide here a direct approach to calculate the optimal discrete wavelet basis.

This paper will study the problem of selecting the optimal wavelet basis with compact support of an appropriate size. We first review briefly the multiresolution property of wavelet functions and the conditions for generating a basis of compactly supported discrete wavelets in terms of properties of quadrature mirror filter (QMF) banks [10]. We then introduce the concepts of information measure as a distance measure and the optimal discrete orthonormal wavelet basis under the information measure. A systematic approach is being developed here to derive the information gradient and the optimal wavelet basis. This approach can be implemented in real time systems due to our parameterization.

# 2 WAVELET TRANSFORM AND QMF BANKS

## 2.1 Wavelet Transform

All the basis functions are dilations and shifts of a single function called the mother wavelet. A general form is

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}), \tag{1}$$

where $a \in R^+$, $b \in R$. The parameters $a$ and $b$ provide scaling and shift of the original function $\psi(t)$. The wavelet transform is defined as

$$X_w(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}\psi^*(\frac{t-b}{a})x(t)dt. \tag{2}$$

The discretized version of the wavelet basis functions is

$$\psi_{m,n}(t) = a_0^{-m/2}\psi(a_0^{-m}t - nb_0),\ m,n \in Z,\ a_0 > 0, b_0 \neq 0, \tag{3}$$

which corresponds to $a = a_0^m$ and $b = na_0^m b_0$, where the size of the shift depends on the scaling factor. We are interested in the dyadic case, that is, $a_0 = 2$ and $b_0 = 1$. It was shown [11] that it is possible to construct a mother wavelet function $\psi(x) \in L^2(R)$ such that for $j,l \in Z$, $\{\psi_{j,l}(x)\}_{j,l\in Z}$ with

$$\psi_{j,l}(x) = 2^{-j/2}\psi(2^{-j}x - l) \tag{4}$$

is an orthonormal basis of $L^2(R)$. Any signal in $L^2(R)$ can be decomposed to its components in different scales in subspaces of $L^2(R)$ of corresponding resolutions and the reverse is true when the regularity condition for the base wavelet $\psi(x)$ is introduced [2, 11]. The base wavelet function $\psi(x)$ plays a central role in this formulation.

## 2.2 Multiresolution Approximation

A multiresolution approximation [11] of $L^2(R)$ is a sequence $\{V_j\}_{j\in Z}$ of closed subspaces of $L^2(R)$ such that the following hold (with $Z$ denoting the set of all integers),

(I).

$$V_j \subset V_{j-1},\ \forall j \in Z \tag{5}$$

$$\bigcup_{j=-\infty}^{+\infty} V_j \text{ is dense in } L^2(R) \text{ and } \bigcap_{j=-\infty}^{+\infty} V_j = \{0\} \tag{6}$$

(II).

$$f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j-1},\ \forall j \in Z \tag{7}$$

(III).

$$f(x) \in V_j \Longrightarrow f(x - 2^j k) \in V_j,\ k \in Z \tag{8}$$

and there is a scaling function $\phi(x) \in L^2(R)$, such that, for all $j \in Z$,

$$\phi_{j,l} = 2^{-j/2}\phi(2^j x - l))_{l \in Z} \tag{9}$$

is an orthonormal basis of $V_j$ with $V_j \subset V_{j-1}$.

With this setting, $W_j$, the complement of $V_j \subset V_{j-1}$, can be expressed as

$$V_j \oplus W_j = V_{j-1}, \tag{10}$$

with

$$V_J = \oplus_{j=J+1}^{\infty} W_j. \tag{11}$$

For all $j$, there is a wavelet function $\psi(x)$, such that,

$$\psi_{j,l}(x) = 2^{-j/2}\psi(2^{-j}x - l))_{l \in Z} \tag{12}$$

is an orthonormal basis of $W_j$. The additional information in an approximation at resolution $2^{-j}$ compared with the resolution $2^{-j+1}$ is contained in the subspace $W_j$, the orthogonal complement of $V_j \in V_{j-1}$. If we define $P_{V_j}$ to be a projection operator in $L^2(R)$ and $I$ to be the identity operator, then

$$P_{V_j} \to I, \text{ as } j \to -\infty. \tag{13}$$

Any square integrable function $f(x) \in L(R^2)$ can be represented as

$$f(x) = \sum_{j,l} w_{j,l}\psi_{j,l}(x), \tag{14}$$

the coefficients $w_{j,l}$ carry the information of $f(x)$ near frequency $2^{-j}$ and near $x = 2^j l$.

## 2.3 Orthonormal Wavelet Basis and QMF

A particular useful setup for our problem is a basis of discrete orthonormal wavelets with compact support. It is useful for real time implementation on digital computers. The compactness of support provides a means of isolation and detection of signals at a certain region, which has proven useful in signal processing problems. Our interest is in parameterizing the discrete wavelet basis functions with a finite number of parameters to generate the optimal wavelet basis for signal representation.

¿From the multiresolution property of wavelets due to Mallat [11] , for $\phi(t) \in V_j$, we have $\phi(2t) \in V_{j-1}$ and $\phi(2t-n)$ is a basis for the space $V_{j-1}$. Hence, we have the expression for the scaling function $\phi(t)$ with $t$ denoting time as [2]

$$\phi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} c_k\phi(2t-k). \tag{15}$$

The corresponding discrete wavelet is given by

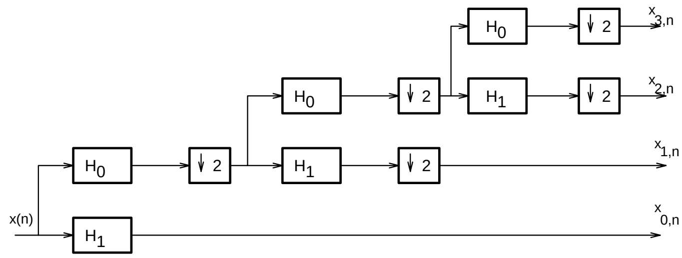
$$\psi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} d_k\phi(2t-k), \tag{16}$$

where the coefficient $\sqrt{2}$ is for normalization purposes. These are the two fundamental equations for the scaling function $\phi(t)$ and wavelet function $\psi(t)$ which is determined by the scaling function $\phi(t)$. The scaling function is to be parameterized by a finite set of parameters as we proceed. Let us denote $h_0(k) = c_k$ and $h_1(k) = d_k$ and take their Fourier transforms

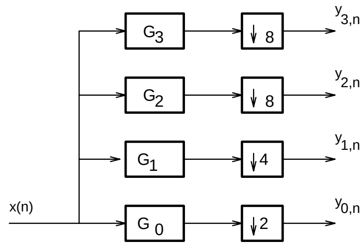
$$H_0(e^{j\omega}) = \sum_k h_0(k)e^{-j\omega k}, \tag{17}$$

and

$$H_1(e^{j\omega}) = \sum_k h_1(k)e^{-j\omega k}. \tag{18}$$



(a)



(b)

Figure 1: (a) Wavelet analysis QMF bank and (b) its equivalent four channel system

The coefficients $\{c_k\}$ and $\{d_k\}$ can be identified with the impulse response of a low pass filter and a high pass filter respectively. The frequency domain versions of the fundamental equations are available by taking the Fourier transform of Equation (15) and Equation (16) with $\Phi(\omega)$ and $\Psi(\omega)$ being their Fourier transforms respectively.

$$\Phi(\omega) = \frac{1}{\sqrt{2}} H_0(e^{j\omega/2})\Phi(\omega/2) \tag{19}$$

and

$$\Psi(\omega) = \frac{1}{\sqrt{2}} H_1(e^{j\omega/2})\Phi(\omega/2) \tag{20}$$

These two equations can be used recursively to generate the scaling and wavelet functions.

We need to consider the case when $H_0(z)$ is a causal FIR filter, i.e., there are only finite many nonzero $c_k$ for the filter. Without loss of generality, we assume that $c_k \neq 0$ when $k \in [0, K]$ where $K$ is a positive odd integer. The scaling function $\phi(t)$ can be nonzero only on $[0, K]$ due to the finite duration of the sequence $\{c_k\}$. The base wavelet function obtained through $\phi(t)$ is also compactly supported. With the FIR assumption, the fundamental equation for the scaling function becomes

$$\phi(t) = \sqrt{2} \sum_{k=0}^{K} c_k \phi(2t - k). \tag{21}$$

The corresponding discrete mother wavelet is given by

$$\psi(t) = \sqrt{2} \sum_{k=0}^{K} d_k \phi(2t - k), \tag{22}$$

We need to find the conditions for the generated wavelet function to produce an orthonormal basis for a subspace of $L^2(R)$ for function approximation and signal representation. Interesting enough, the dyadic orthonormal wavelet functions can be related to binary tree structured QMF banks constructed from the two basic FIR filters which determine the scaling function and the wavelet function.

Figure 1 (a) shows a three level dyadic tree structured QMF bank for wavelet transformation. The input sequence $x(n)$ is decomposed to different resolutions by passing the signal through the QMF bank. The output $x_{i,n}$

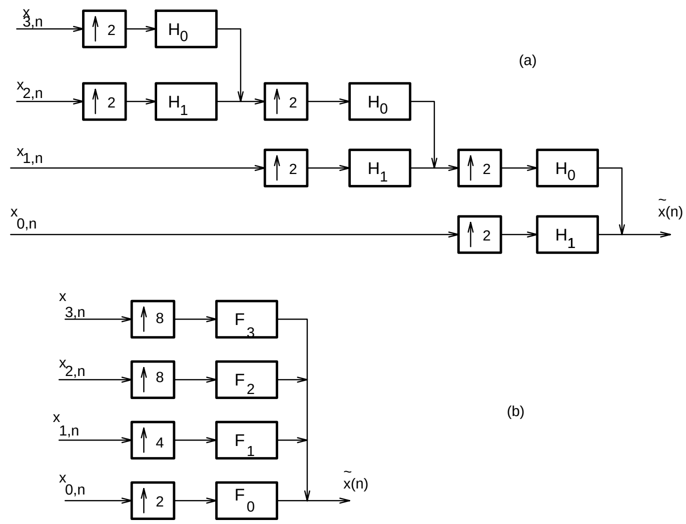



Figure 2: (a) Wavelet synthesis QMF bank and (b) its equivalent four channel system

are the related wavelet coefficients. Figure 1 (b) provides the equivalent four channel filter bank which is derived from (a) through block transforms. Figure 2 demonstrates the corresponding synthesis QMF bank whose input is the sequence of the wavelet coefficients while the output is the reconstructed signal $\tilde{x}(n)$ which is the wavelet representation of the original signal.

**Theorem 2.1** *Let $H_0(z)$ and $H_1(z)$ be causal FIR filters, then the scaling function $\phi(t)$ and the wavelet function $\psi(t)$ generated by the QMF bank of Figure 1 and Figure 2 are causal with finite duration $Kb_0$. Further, if $H_0(z)$ and $H_1(z)$ satisfy the paraunitary condition* [10] *, $|H_0(1)| = \sqrt{2}$ and $H_0(e^{j\omega}) \neq 0$ while $|\omega| < \pi/2$, the wavelet functions $\psi_{j,l}(t)$ are orthonormal.*

The condition imposed for orthonormality of wavelets can be relaxed when the number of levels of the QMF tree is finite, in this case, both the scaling function and the base wavelet function are obtained through finite recursion by using Equation (19) and Equation (20) respectively, i.e., the paraunitary condition alone is enough to guarantee the orthonormality of the wavelet functions. A proof of this fact in the frequency domain is provided in [10] . As a matter of fact, this is the usual situation in practical application and implementation.

**Lemma 2.1** *Compactly supported scaling function and wavelets generated through the finite recursion are orthonormal if the matrix*

$$\mathcal{H}(\omega) = \begin{bmatrix} H_0(e^{j\omega}) & H_1(e^{j\omega}) \\ H_0(e^{j(\omega+\pi)}) & H_1(e^{j(\omega+\pi)}) \end{bmatrix} \tag{23}$$

*is paraunitary for all $\omega$ for the two-channel quadrature mirror filter (QMF) bank.*

This is the constraint that the parameters $c_k$ should satisfy. In particular, the cross-filter orthonormality implied by the paraunitary property, is satisfied by the choice of

$$H_1(z) = -z^{-K} H_0(-z^{-1}), \text{ K odd} \tag{24}$$

or in the time domain,

$$h_1(k) = (-1)^k h_0(K-k). \tag{25}$$

As we can see from the above, both the scaling function and the wavelet function depend on the selection of $\{c_k\}$ for $k \in [0, K]$. As a consequence, the dilations and shifts of the base wavelet depend on the selection of this set of parameters subject to the paraunitary condition imposed on the filters of the QMF bank.

# 3 OPTIMAL DISCRETE WAVELET BASIS WITH COMPACT SUPPORT

## 3.1 Parameterization of Wavelet Functions and Information Measures

We first introduce a distance measure for optimization purpose. Inspired by the work in [8] , we define an additive information measure of entropy type and the optimal basis as the following.

**Definition 3.1** *A non negative map $\mathcal{M}$ from a sequence $\{f_i\}$ to $R$ is called an additive information measure if $\mathcal{M}(0) = 0$ and $\mathcal{M}(\sum_i f_i) = \sum_i \mathcal{M}(f_i)$.*

**Definition 3.2** *Let $x \in R^N$ be a fixed vector and $\mathcal{B}$ denote the collection of all orthonormal bases of dimension $N$, a basis $B \in \mathcal{B}$ is said to be optimal if $\mathcal{M}(Bx)$ is minimal for all bases in $\mathcal{B}$ with respect to the vector $x$.*

We shall define a distance measure between a signal and its decompositions to subspaces of $L^2(R)$ motivated by Shannon entropy (Shannon's formula) [12]

$$H(X) = H(P) = -\sum_{x \in X} P(x) \log P(x), \tag{26}$$

which is interpreted as a measure of the information content of a random variable $X$ with distribution $P_x = P$ in information theory.

**Definition 3.3** *Let $H$ be a Hilbert space which is an orthogonal direct sum*

$$H = \oplus \sum H_i, \tag{27}$$

*a map $\mathcal{E}$ is called decomposition entropy if*

$$\mathcal{E}(v, \Psi) = -\sum \frac{\|v_i\|^2}{\|v\|^2} \log \frac{\|v_i\|^2}{\|v\|^2} \tag{28}$$

*for $v \in H$, $\|v\| \neq 0$, such that*

$$v = \oplus \sum v_i, v_i \in H_i, \tag{29}$$

*and we set*

$$p \log p = 0, \text{ when } p = 0. \tag{30}$$

Entropy is a good measure for signal concentration in signal processing and information theory. The value of $\exp \mathcal{E}(v)$ is proportional to the number of coefficients and the length of code words necessary to represent the signal to a fixed mean error and to error-less coding respectively. The number $\frac{\|v_i\|^2}{\|v\|^2}$ is the equivalent probability measure in the decomposition entropy which is the stochastic approximation of Shannon entropy since the density function of the signal is unknown. Entropy obtains its maximum when energy of the signal is uniformly distributed in its frequency domain. On the contrary, lower entropy value means higher concentration of the signal energy over certain frequency bands. In our formulation, energy concentration is identified with a model of lower order or networks with less complexity. The implication of using entropy as a performance measure takes advantage of the fact of the nonuniform energy distribution of the signal or system in consideration over its energy spectrum. The optimization of the wavelet basis is finding the suitable wavelet for a certain class of signals which have energy concentration at certain frequency bands. In other words, we are seeking a representative of certain class of signals to generate suitable subspaces in which the decomposition entropy is minimized or equivalently that the energy of the signal is concentrated.

Let $\psi(t)$ be the mother wavelet function and let $\Psi(t)$ represent the orthonormal discrete wavelet basis of $L^2$ generated by dilation and shifting of $\psi(t)$, similarly, we define $\Psi_j$ to be the basis of $H_j$. We write $\Psi(t) = \{\psi_{j,l}(t)\}$ and $\Psi_j(t) = \{\psi_{j,l}(t)\}_{l \in Z}$ respectively. We treat both $\Psi(t)$ and $\Psi_j$ as operators and thus define the following.

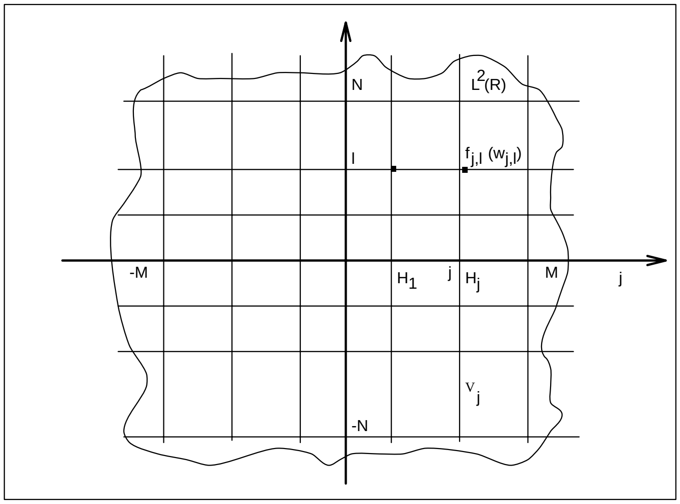



Figure 3: Mesh structure of the projection space

**Definition 3.4** *Let $\Psi$ be a basis given above, a base operation is defined to be a map from $L^2(R)$ to a set of real numbers, i.e., $\Psi(t)f(t) = \{f_{j,l}\}_{j,l \in Z}$ where $f_{j,l} = \langle f(t), \psi_{j,l}(t)\rangle$ for all $f(t) \in L^2$.*

Consider $V_J$, the subspace of $L^2(R)$, with

$$V_J = \oplus_{j=J+1}^{\infty} H_j, \tag{31}$$

and Equation (14), let $M$ and $N$ be appropriate positive integers, we truncate the approximation in Equation (14) to a scale up to $M$, we have

$$f(x) = \sum_{j=-M}^{M} \sum_{l=-N}^{N} w_{j,l}\psi_{j,l}(x). \tag{32}$$

The subspaces used to approximate the function $f(x)$ has a mesh of size $(2M+1) \times (2N+1)$ as in Figure 3.

Given a function or signal $f(t) \in L^2(R)$ and a base wavelet function $\psi(t)$ with a finite mesh of size $(2M+1) \times (2N+1)$, we decompose the signal to the orthogonal subspaces as

$$f(t) = \sum_{j=-M}^{M} \sum_{l=-N}^{N} f_{j,l}\psi_{j,l}(t). \tag{33}$$

We are going to find the optimal wavelet base function $\psi(t)$ for a given signal $f(t)$ such that the additive information measure $\mathcal{M}$ is minimized. The result of the base operation $\Psi f(t)$ appears as the weights on the nodes of the mesh. The weights on the vertical line with coordinate $j$ form the number set produced by $\Psi_j f(t)$.

Although the decomposition entropy is a good measure for the "distance", it is not an additive type of map because the norm $\|v\|$ is used to scale the vector. We thus further introduce a cost functional

$$\lambda(\Psi, v) = -\sum_j \|v_j\|^2 \log \|v_j\|^2, \tag{34}$$

which relates to the decomposition entropy through

$$\mathcal{E}(v, \Psi) = \|v\|^{-2} \lambda(\Phi, v) + \log \|v\|^2 (2M+1). \tag{35}$$

As shown in the above expression, the cost functional $\lambda$ takes the wavelet basis $\Psi$ and the signal vector $v$ as its arguments. For any fixed signal, it is a functional of the basis and hence that of the wavelet function $\psi(t)$. The function in (34) is an additive measure. Since the above two functionals share the same set of minimal points, we minimize the functional $\lambda(\Phi, f)$ to find the optimal wavelet basis through multiresolution decomposition of a given signal of finite energy.

The weight of the decomposition of the signal $f(t)$ on a subspace $H_j$ is measured by a subnorm $\|f_j\|$ defined as

$$\|f_j(t)\| = \left\|P_{H_j}[f(t)]\right\|, \tag{36}$$

where

$$\|f_j\|^2 = \sum_{l=-N}^{N} f_{j,l}^2. \tag{37}$$

Similarly, the norm of the decomposed signal is given by

$$\|f(t)\|^2 = \sum_{j=-M}^{M} \|f_j\|^2. \tag{38}$$

## 3.2 Sensitivity Gradient of Wavelet Components

We need to further compute $\frac{\partial f_{j,l}}{\partial c_k}$ which is a measure of the sensitivity of the components of the signal decomposition with respect to a wavelet basis versus the change of the defining parameter set of the mother wavelet. One can compute this quantity through numerical methods from the relations and definitions. Based on the definition of information gradient and the properties of QMF discussed earlier, we derive an explicit expression below.

**Lemma 3.1** *The sensitivity gradient* $\frac{\partial \psi_{j,l}}{\partial c_k}$ *of the component* $\psi_{j,l}$ *of the wavelet basis* $\Psi$ *with respect to the parameter* $c_k$ *is given by*

$$\begin{aligned}\frac{\partial \psi_{j,l}}{\partial c_k} &= \sqrt{2^{-j+1}} \sum_n \left[(-1)^{K-k}\phi(2^{-j+1}t - 2l - n)\right.\\ &\quad \left. + (-1)^n \sqrt{2} c_{K-n} \phi(2^{-j+2}t - 4l - 2n - k)\right].\end{aligned} \tag{39}$$

*Proof:*
From the fundamental equation of wavelets (22) and the wavelet basis function,

$$\frac{\partial \psi_{j,l}}{\partial c_k} = \sqrt{2^{-j+1}} \frac{\partial}{\partial c_k} \sum_n h_1(n)\phi(2^{-j+1}t - 2l - n), \tag{40}$$

or

$$\frac{\partial \psi_{j,l}}{\partial c_k} = \sqrt{2^{-j+1}} \sum_n \left[\frac{\partial h_1(n)}{\partial c_k}\phi(2^{-j+1}t - 2l - n) + h_1(n)\frac{\partial}{\partial c_k}\phi(2^{-j+1}t - 2l - n)\right]. \tag{41}$$

From Equation (21), we have

$$\frac{\partial \phi(t)}{\partial c_k} = \sqrt{2}\phi(2t - k). \tag{42}$$

hence,

$$\frac{\partial}{\partial c_k}\phi(2^{-j+1}t - 2l - n) = \sqrt{2}\phi(2^{-j+2}t - 4l - 2n - k). \tag{43}$$

We next need to find $\frac{\partial h_1(n)}{\partial c_k}$. From the time domain relation (25) of the QMF, we have ,

$$h_1(n) = (-1)^n h_0(K - n) \tag{44}$$

with $h_0$ being compactly supported on $[0, K]$. Thus,

$$h_1(n) = \frac{\partial}{\partial c_k}(-1)^n c_{K-n}, \tag{45}$$

and there is only one nonzero term when $K - n = k$. This yields,

$$\frac{\partial h_1(n)}{\partial c_k} = (-1)^{K-k}. \tag{46}$$

The lemma is proven through (43) and (46).

□

This lemma establishes a direct link between the rate of change of the components in the basis $\Psi$ and the variations of the parameters in the fundamental equations of wavelets, which leads to the next theorem. We introduce the following theorem to show the relationship between the information measure and the parameter set $\{c_k\}$ and the relation here shall provide a clue for developing an algorithm to find the optimal base wavelet function for signal representation.

**Theorem 3.1** *Let $\lambda(\cdot,\cdot)$ be the additive information measure and $[0, K]$ be the compact support for $\{c_k\}$ and $\Psi$ be the corresponding wavelet basis from dilations and shifts of the wavelet $\psi(t)$. Let $f(t)$ be a fixed signal in $L^2(R)$. Then the gradient of the information measure with respect to the parameter set $\{c_k\}$ for the given signal is described by*

$$\begin{aligned} \frac{\partial \lambda(\Psi, f(t))}{\partial c_k} &= -\sqrt{2^{-j+2}} \sum_j \sum_l \log 2 \left\| f_j \right\|^2 \\ &\quad \cdot f_{j,l} \sum_n \left[ (-1)^{K-k} \left\langle f(t), \phi(2^{-j+1}t - 2l - n) \right\rangle \right. \\ &\quad \left. + (-1)^n c_{K-n} \left\langle f(t), \phi(2^{-j+2}t - 4l - 2n - k) \right\rangle \right]. \end{aligned} \tag{47}$$

*Proof:*
By the chain rule, we have the information gradient

$$\frac{\partial \lambda(\Psi, f(t))}{\partial c_k} = \sum_j \frac{\partial \lambda(\Psi, f(t))}{\partial \|f_j\|^2} \frac{\partial \|f_j\|^2}{\partial c_k}. \tag{48}$$

The definition of information measure $\lambda(f(t))$ in (34) yields,

$$\begin{aligned} \frac{\partial \lambda(\Psi, f(t))}{\partial \|f_j\|^2} &= -\log \|f_j\|^2 - 1 \\ &= -\log 2 \|f_j\|^2, \end{aligned} \tag{49}$$

with 2 being the base of the *log* function. We use the chain rule again,

$$\begin{aligned} \frac{\partial \|f_j\|^2}{\partial c_k} &= \frac{\partial}{\partial c_k} \sum_l f_{j,l}^2 \\ &= 2 \sum_l f_{j,l} \frac{\partial f_{j,l}}{\partial c_k}. \end{aligned} \tag{50}$$

We have so far

$$\frac{\partial \lambda(\Psi, f(t))}{\partial c_k} = -2 \sum_j \sum_l \log 2 \|f_j\|^2 f_{j,l} \frac{\partial f_{j,l}}{\partial c_k}. \tag{51}$$

Since

$$\frac{\partial f_{j,l}}{\partial c_k} = \left\langle f(t), \frac{\partial \psi_{j,l}}{\partial c_k} \right\rangle, \tag{52}$$

the result from the previous lemma concludes the proof.

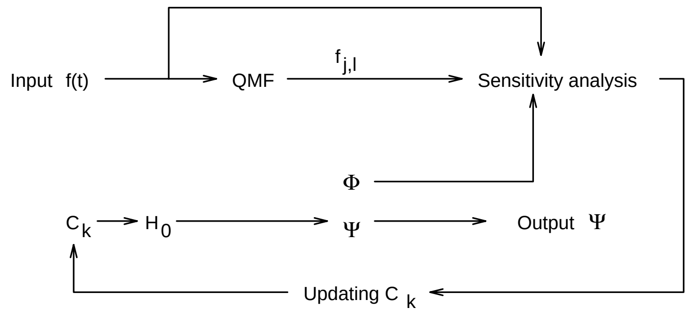



Figure 4: Flow chart for computing $\psi(t)$

□

This theorem demonstrates an explicit relation among the gradient of the additive information measure, the parameter set $\{c_k\}$ and the measured signal $f(t)$. It will facilitate the search for the optimal wavelet basis due to our parameterization and the information measure. This algorithm starts by assigning an initial set of parameters which form the low pass filter of the QMF bank which is followed by the generation of both the scaling function $\phi(t)$ and the wavelet $\psi(t)$ through a recursive process. The wavelet decomposition is implemented through passing the input signal through the QMF bank composed of $H_0(z)$ and $H_1(z)$. The flow chart in Figure 4 describes this process.

# 4 ALGORITHMS

We have identified the problem of finding the optimal wavelet basis $\Psi$ with that of finding a parameter set $\{c_k\}$ such that the additive information measure $\lambda$ is minimized. Once the set $\{c_k\}$ is determined, both the scaling function $\phi$ and the base wavelet function $\psi$ can be derived afterwards. Equipped with the above theorem, the information gradient is available, and different optimization schemes can be applied to solve this problem. We have developed a basis selection algorithm based on the steepest descent method as follows. To simplify notation, we denote the parameter set $\{c_0 c_1 \cdots c_{K-l}\}$ by a vector $C$.

**Algorithm 4.1** *Computation of the optimal wavelet basis*

*Step 1: Set* $i := 1$,
$\lambda_0 := 0$,
*mesh parameters* $M, N$;
*Initialize vector* $C_0$;
*Input* $f(t)$.
*Step 2: If* $C_i$ *dose not satisfy the constraint,*
*then, modify* $C_i$ *and repeat Step 2.*
*Step 3:* $C_i := C_{i-1} + p_{i-1}\frac{\partial\lambda}{\partial C_{i-1}}$.
*Step 4: Compute* $\phi$ *and* $\psi$.
*Step 5: Compute* $\lambda$.
*Step 6: If* $|\lambda_i - \lambda_{i-1}| > \epsilon$,
$i := i + 1$, *go to Step 2.*
*Step 7: Output the optimal basis* $\Psi$ *and stop.*

The mesh size is governed by the choice of parameters $M$ and $N$. Obviously, when $M$ and $N$ turn to infinity, the supporting subspace spanned by the dilations and shifts of the base wavelet turns to the space $L^2(R)$. The size of the mesh is identified with the complexity of the resulting wavelet system. The constraint on the parameter $c_k$ is dominated by the unitary property of the QMF bank which can be transformed into a set of algebraic equations. The parameters M and N can be predetermined by the time and frequency localization property of the signal in consideration. We can also perform an adaptation scheme to generate the system with the appropriate order. This is realized by a modified algorithm as follows.

**Algorithm 4.2** *Computation of the optimal wavelet basis with variable mesh size.*

*Step 1: Set* $i := 1$,
$\lambda_0 := 0$,
*mesh parameters* $M, N$;
*Initialize vector* $C_0$;
*Input* $f(t)$.
*Step 2: If* $C_i$ *does not satisfy the constraint,*
*then, modify* $C_i$ *and repeat Step 2.*
*Step 3:* $C_i := C_{i-1} + p_{i-1} \frac{\partial \lambda}{\partial C_{i-1}}$.
*Step 4: Compute* $\phi$ *and* $\psi$.
*Step 5: Compute* $\lambda$.
*Step 6: If* $|\lambda_i - \lambda_{i-1}| > \epsilon$,
$i := i + 1$,
$M := M + 1$,
$N := N + 1$, *go to Step 2.*
*Step 7: Output the optimal basis* $\Psi$ *and stop.*

This algorithm starts from an initial mesh size determined by $M$ and $N$ in step 1. While updating the parameter set $\{C_i\}$, the algorithm adjusts the size of the mesh until the error tolerance is met to finish the iterative process. The sequence of order updating and parameter updating can be organized adequately for reducing computation complexity.

# 5 CONCLUSIONS

This paper has provided a direct approach to construct an optimal orthonormal wavelet basis with compact support for signal representation. The cost functional, an additive information measure, is introduced based on the decomposition entropy of the given signal with respect to an initial wavelet basis. This entropy measures the nonuniform energy concentration of the given signal of finite energy in the sense of being square integrable. The sensitivity of each dilation and shift of the base wavelet function $\psi(t)$ with respect to the governing coefficients has been found, which establishes the gradient of the information measure versus the parameter set. The parameterization of both the information measure and the base wavelet allows an explicit expression of information gradient with respect to the optimization parameters and thus paves the way to an efficient basis selection algorithm.

The constraint on the optimal basis selection can be removed through appropriate parameterization and structure analysis of the corresponding QMF bank to yield an unconstrained optimization problem. The results of this research shall appear in a forthcoming paper. This methodology of the optimal basis selection in a general setting is useful not only for signal approximation and reconstruction in $L^2(R)$ but also for data compression and system identification. In the context of pattern recognition, it is also a way to construct the feature space and for partitioning the signal space according to its representatives. The parametrization of the cost functionals described here is very helpful; other forms of measures or cost functions may be introduced depending on the contexts of the actual physical problems.

# 6 ACKNOWLEDGMENTS

This research was partially supported by NSF grant NSFD CDR 8803012 through the Engineering Research Center's Program.

# References

[1] C.K. Chui. *Wavelets: A tutorial in Theory and Applications*. Acamemic Press, Inc., 250 Sixty Ave, San Diego, CA 92101, 1992.

[2] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, 3600 University City Science Center, Philadelphia PA, 1992.

[3] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, pages 14–38, Oct. 1991.

[4] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[5] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. on Pattern Analysis and Machine intelligence*, 14(7):710–732, July 1992.

[6] A.H. Tewfik, D. Sinha, and P. Jorgensen. On the optimal choice of a wavelet for signal representation. *IEEE Trans. on Information Theory*, 38(2):747–765, Mar. 1992.

[7] P. Jorgensen. Choosing discrete orthogonal wavelets for signal analysis and approximation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages III–308–311, Minnesota, Minneapolis, April 27-30 1993.

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

[9] Y. Zhuang and J.S. Baras. Identification of infinite dimensional systems via adaptive wavelet neural networks. Tech. Report 93-64, Institute for Systems Reseaarch, The University of Maryland, College Park, MD 20742, Aug. 1993.

[10] P.P. Vaidyanathan. *Multirate Systems and Filterbanks*. Prentice hall Signal Processing Series. P T R Prentice-Hall, Inc., Englewood Cliffs, NJ 07632, 1993.

[11] S.G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Transactions of The American Mathematical Society*, 315(1):69–87, Sept. 1989.

[12] I. Csiszár and J. Körner. *Information Theory*. Akadémiai Kiadó, Budapest, Hungary, 1981.