Entitled:

"Existence and Construction of Optimal Wavelet
Basis for Signal Representation"

Authors:

(with Y Zhuang)

Journal

*IEEE Transactions on Signal Processing*

# Existence and Construction of Optimal Wavelet Basis for Signal Representation *

Yan Zhuang[†] and John S. Baras [‡]

Institute for Systems Research and Department of Electrical Engineering

The University of Maryland, College Park, MD 20742

## Abstract

We study the problem of choosing an 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 decompositions and the conditions for generating an orthonormal basis of compactly supported discrete wavelets in terms of properties of quadrature mirror filter (QMF) banks. We then parameterize the mother wavelet and scaling function of wavelet systems through a set of real coefficients. We further introduce the concept of decomposition entropy as an information measure to describe the distance between a given signal and its projection onto the subspace spanned by the wavelet basis in which the signal is to be reconstructed. The optimal basis for the 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 in this paper to derive the information gradient with respect to the parameter set from a given square integrable signal and an initial discrete basis of wavelets. The existence of the optimal basis for the wavelets has been proven in this paper. A gradient based optimization algorithm is developed for computing the optimal wavelet basis.

# 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, 4]. The most attractive features of the wavelet transform are the multiresolution property and time and frequency localization ability. The wavelet transform decomposes a signal into its components at different resolutions. Its application actually simplifies the description of signals and provides analysis at different levels of detail. There are many applications of these properties in the areas of signal processing, speech processing and especially in image processing [13, 11, 9]. The wavelet transform differs from the 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. Appropriate selection of the wavelet for signal representation can result in maximal benefits of this new technique. One of the fundamental questions in wavelet applications, how to construct the optimal wavelet basis for a specific application, remains open. It is reasonable to anticipate 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. It is widely accepted that selection of wavelets can impact the performance of algorithms and systems in many applications. This paper addresses the issue of finding an 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 [14, 8] 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 [2]. However, a direct method to systematically generate the optimal orthonormal discrete wavelet basis with compact support has not been developed as of yet. Our work attempts to fill this gap by introducing a systematic approach for constructing the signal based optimal wavelet basis. We proposed an information measure based approach for constructing the optimal discrete wavelet basis with compact support in our previous work on adaptive wavelet neural networks [16]. We shall provide here a direct approach to calculate the optimal discrete wavelet basis. This approach can be implemented in real time systems due to our parameterization.

The rest of this paper is organized as follows. In section II, we 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 [15]. In section III, we 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 developed to derive the information gradient and the optimal wavelet basis. In section IV, we provide a proof of existence of the optimal wavelet basis. In section V, we develop an algorithm for constructing the optimal basis for wavelets with compact support. In section VI, we summarize our conclusions.

## 2    Wavelet Transform and QMF Banks

This section reviews the properties of wavelet transforms and QMF banks as well as the relationship between them for generating orthonormal discrete wavelets. It makes use of the multiresolution property of wavelets as studied by Mallat [10], the link between wavelets and filter banks from Daubechies [3], and the structures of multirate systems and quadrature mirror filters (QMF) of Vaidyanathan [15], and Herley [5].

3

## 2.1 Wavelet Transform

The basis functions of a wavelet system are dilations and shifts of a single function called the mother wavelet. A general form of the basis functions 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 [10] 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 into its components in different scales in subspaces of $L^2(R)$ of corresponding resolutions, the reverse is true when the regularity condition for the mother wavelet $\psi(x)$ is introduced [4, 10]. The base wavelet function $\psi(x)$ plays a central role in this formulation.

## 2.2 Multiresolution Approximation

A multiresolution approximation due to Mallat [10] 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).

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

4

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

(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 particularly 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 [10], 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 [4]

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

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

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

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

6

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 finitely 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. Interestingly 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 into different resolutions by passing the signal through the QMF bank. The output $x_{i,n}$ are the related wavelet coefficients. Figure 1 (b) provides the
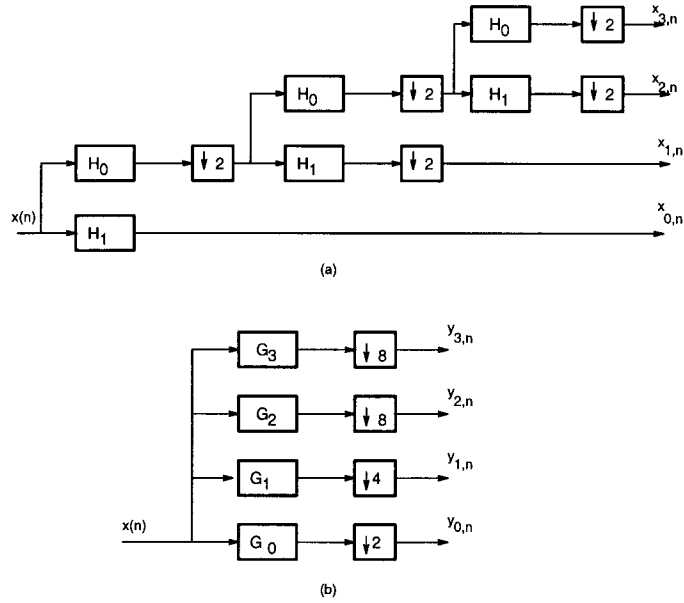
Figure 1: (a) Wavelet analysis QMF bank and (b) its equivalent four channel system
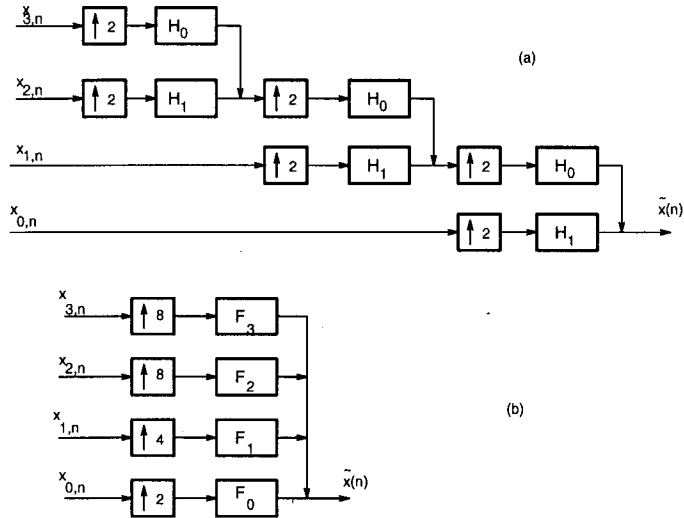


Figure 2: (a) Wavelet synthesis QMF bank and (b) its equivalent four channel system
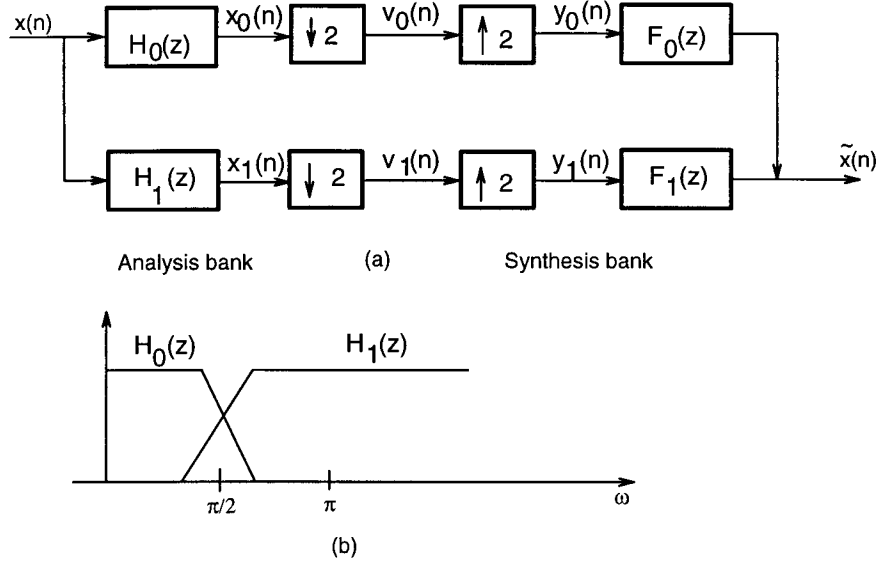
Figure 3: Quadrature mirror filter bank (a) and its response (b)

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.

**Definition 2.1 (Vaidyanathan [15] p196)** *A filter bank is said to be a perfect reconstruction system if its transfer function is a pure delay, free from aliasing, amplitude distortion and phase distortion, i.e., if $\tilde{x}(n) = cx(n - n_0)$.*

**Definition 2.2 (Vaidyanathan [15] p288)** *A rational transfer matrix $H(z)$ is said to be paraunitary if*

$$\tilde{H}(z)H(z) = dI, \quad \text{for all } z \tag{23}$$

*where $d$ is a constant, $I$ is the identity matrix.*

The QMF bank in Figure 3 is a general setting. The reconstructed signal is given by

$$\tilde{X}(z) = \frac{1}{2}[H_0(z)F_0(z) + H_1(z)F_1(z)]X(z)$$

$$+ \frac{1}{2}[H_0(-z)F_0(z) + H_1(-z)F_1(z)]X(-z). \tag{24}$$

9

The $X(-z)$ term represents the aliasing component of the system due to decimation of the input signal $x(n)$. The matrix notation of the above is

$$\tilde{X}(z) = \frac{1}{2} \begin{bmatrix} X(z) & X(-z) \end{bmatrix} \begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z) \end{bmatrix} \begin{bmatrix} F_0(z) \\ F_1(z) \end{bmatrix} \tag{25}$$

where $H(z)$ is the alias-component (AC) matrix of the system which is given by

$$\mathcal{H}(z) = \begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z) \end{bmatrix} . \tag{26}$$

The conditions for this filter bank to be free of aliasing are the following [15],

$$F_0(z) = H_1(-z), \; F_1(z) = -H_0(-z), \tag{27}$$

which cause the coefficient of the $X(-z)$ to be zero and thus cancel the aliasing component. In fact, only the filter $H_0(z)$ needs to be designed for the system and the rest filters are given by the above relation. If we further assume that $H_0(z)$ be power symmetric, i.e.,

$$\tilde{H}_0(z)H_0(z) + \tilde{H}_0(-z)H_0(-z) = 1, \tag{28}$$

and assign

$$H_1(z) = -z^{-N}\tilde{H}_0(-z) \tag{29}$$

where $N$ is an odd integer, then, the transfer function of the QMF bank reduces to a pure delay of $N$ unit time. Thus, $\tilde{x}(n) = \frac{1}{2}x(n - N)$ and the QMF bank forms a perfect reconstruction system. The following theorem is a revised version of a theorem due to Vaidyanathan [15, p530].

**Theorem 2.1 (Vaidyanathan [15] p530)** *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 parauni-tary condition, $|H_0(1)| = \sqrt{2}$ and $H_0(e^{j\omega}) \neq 0$ while $|\omega| < \pi/2$, the wavelet functions $\psi_{j,l}(t)$ are orthonormal.*

10

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 mother 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 can be found in [15]. As a matter of fact, this is the usual situation in practical application and implementation.

We summarize the conditions for compactly supported scaling functions and wavelets to be orthonormal based on the theorem above and the properties of AC matrices of the two channel QMF banks as a lemma as follows.

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

$$\mathcal{H}(\omega) = \left[ \begin{array}{cc} H_0(e^{j\omega}) & H_1(e^{j\omega}) \\ H_0(e^{j(\omega+\pi)}) & H_1(e^{j(\omega+\pi)}) \end{array} \right] \tag{30}$$

*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

$$H_1(z) = -z^{-K} H_0(-z^{-1}), \text{ K odd} \tag{31}$$

or in the time domain,

$$h_1(k) = (-1)^k h_0(K - k). \tag{32}$$

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 mother wavelet depend on the selection of this set of parameters subject to the paraunitary condition imposed on the filters of the QMF bank.

11

# 3 Optimal Discrete Wavelet Basis with Compact Support

In this section, we parameterize the discrete wavelets, introduce an additive information measure and derive the information gradient for computing the optimal basis for wavelets.

## 3.1 Parameterization of Wavelet Functions and Information Measures

We first introduce a distance measure for optimization purpose. Inspired by the work in [2], we define an additive information measure of entropy type and the optimal basis as the following.

**Definition 3.1 (Coifman et al [2])** *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) [6]

$$H(X) = H(P) = -\sum_{x \in X} P(x) \log P(x) \tag{33}$$

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 of its subspaces $H_i$*

$$H = \oplus \sum H_i, \tag{34}$$

*and let $B$ be a basis spanning $H$; a map $\mathcal{E}$ is called decomposition entropy if*

$$\mathcal{E}(v, B) = -\sum \frac{\|v_i\|^2}{\|v\|^2} \log \frac{\|v_i\|^2}{\|v\|^2} \tag{35}$$

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

$$v = \oplus \sum v_i, v_i \in H_i, \tag{36}$$

*and we set*

$$p \log p = 0, \quad when \ p = 0. \tag{37}$$

Entropy is a good measure for signal concentration in signal processing and information theory. The value of $\exp \mathcal{E}(v, B)$ 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. For a source of a finite number of independent signals, such as a digital image considered as a source of independent pixels, its entropy is maximum for uniform distribution [7, p42]. If the entropy value is less than the maximum, then, this implies that a higher concentration of the signal energy over certain bands exists. 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 nonuniform energy distribution of the signal or system in consideration over its energy spectrum. The optimization of the wavelet basis finds 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 a 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.

**Definition 3.4** *Let $\Psi$ be a basis given as above; a basis 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 $f(t) \in L^2$.*
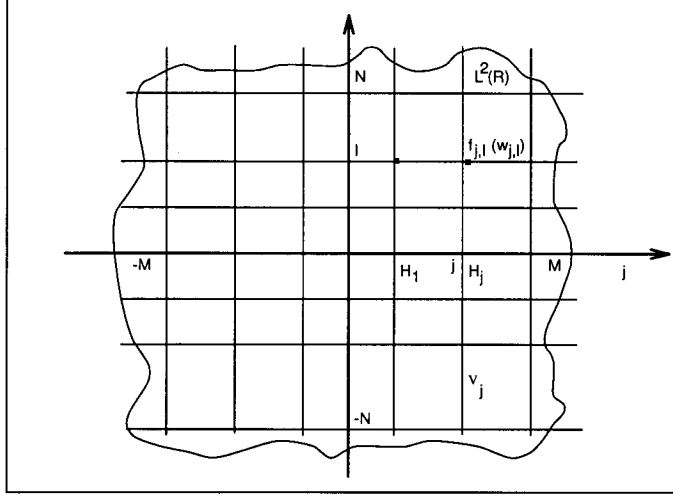
Figure 4: Mesh structure of the projection space

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

$$V_J = \oplus_{j=J+1}^{\infty} H_j \tag{38}$$

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

$$f(x) = \sum_{j=-M}^{M} \sum_{l=-N}^{N} w_{j,l} \psi_{j,l}(x). \tag{39}$$

The subspaces used to approximate the function $f(x)$ has a mesh of size $(2M + 1) \times (2N + 1)$ as in Figure 4.

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 into the orthogonal subspaces as

$$f(t) = \sum_{j=-M}^{M} \sum_{l=-N}^{N} f_{j,l} \psi_{j,l}(t). \tag{40}$$

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 $\Psi_j f(t)$.

14

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

$$\lambda(\Psi, v) = -\sum_j \|v_j\|^2 \log \|v_j\|^2 \tag{41}$$

which relates to the decomposition entropy through

$$\mathcal{E}(v, \Psi) = \|v\|^{-2} \lambda(\Phi, v) + \log \|v\|^2 (2M + 1). \tag{42}$$

As shown in the expression above, 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 (41) is an additive measure. Since the two functionals above share the same set of minimal points, we minimize the functional $\lambda(\Psi, f)$ 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

$$\|f_j(t)\| = \left\| P_{H_j}[f(t)] \right\| \tag{43}$$

where

$$\|f_j\|^2 = \sum_{l=-N}^{N} f_{j,l}^2. \tag{44}$$

Similarly, the norm of the decomposed signal is given by

$$\|f(t)\|^2 = \sum_{j=-M}^{M} \|f_j\|^2. \tag{45}$$

## 3.2 Sensitivity Gradient of Wavelet Components

We need to further find $\frac{\partial f_{j,l}}{\partial c_k}$ 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 measure and the properties of QMF discussed earlier, we derive an explicit expression for analyzing the sensitivity of a dilated and shifted wavelet function versus the parameter set as follows.

**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*

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

$$\left. + (-1)^n \sqrt{2} c_{K-n} \phi(2^{-j+2}t - 4l - 2n - k) \right]. \tag{46}$$

*Proof:*

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

$$\frac{\partial \psi_{j,l}}{\partial c_k} = \sqrt{2^{-j+1}} \frac{\partial}{\partial c_k} \sum_n h_1(n) \phi(2^{-j+1}t - 2l - n) \tag{47}$$

or

$$\frac{\partial \psi_{j,l}}{\partial c_k} = \sqrt{2^{-j+1}} \sum_n \left[ \frac{\partial h_1(n)}{\partial c_k} \phi(2^{-j+1}t - 2l - n) + h_1(n) \frac{\partial}{\partial c_k} \phi(2^{-j+1}t - 2l - n) \right]. \tag{48}$$

From Equation (21), we have

$$\frac{\partial \phi(t)}{\partial c_k} = \sqrt{2} \phi(2t - k). \tag{49}$$

hence,

$$\frac{\partial}{\partial c_k} \phi(2^{-j+1}t - 2l - n) = \sqrt{2} \phi(2^{-j+2}t - 4l - 2n - k). \tag{50}$$

We next need to find $\frac{\partial h_1(n)}{\partial c_k}$. From the time domain relation (32) of the QMF, we have ,

$$h_1(n) = (-1)^n h_0(K - n) \tag{51}$$

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

$$h_1(n) = \frac{\partial}{\partial c_k}(-1)^n c_{K-n} \tag{52}$$

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

$$\frac{\partial h_1(n)}{\partial c_k} = (-1)^{K-k}. \tag{53}$$

The lemma is proven through (50) and (53).

$\square$

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 $\{c_k\}$. The relation to be described shall provide a clue for developing an algorithm to find the optimal mother wavelet function for signal representation.

**Theorem 3.1** *Let $\lambda(\cdot, \cdot)$ be the additive information measure as defined in (41) and $f(t) \in L^2(R)$ be a fixed signal. Let $[0, K]$ be the compact support for $\{c_k\}$ and let $\Psi$ be the corresponding wavelet basis from dilations and shifts of the mother wavelet $\psi(t)$. Then the gradient of the information measure with respect to the parameter set $\{c_k\}$ for the given signal is given by*

$$\begin{aligned}
\frac{\partial \lambda(\Psi, f(t))}{\partial c_k} &= -\sqrt{2^{-j+2}} \sum_j \sum_l \log 2 \, \|f_j\|^2 \\
&\quad \cdot f_{j,l} \sum_n \left[ (-1)^{K-k} \left\langle f(t), \phi(2^{-j+1}t - 2l - n) \right\rangle \right. \\
&\quad \left. + (-1)^n c_{K-n} \left\langle f(t), \phi(2^{-j+2}t - 4l - 2n - k) \right\rangle \right].
\end{aligned} \tag{54}$$

*Proof:*

By the chain rule, we have the information gradient

$$\frac{\partial \lambda(\Psi, f(t))}{\partial c_k} = \sum_j \frac{\partial \lambda(\Psi, f(t))}{\partial \|f_j\|^2} \frac{\partial \|f_j\|^2}{\partial c_k}. \tag{55}$$

17

The definition of information measure $\lambda(f(t))$ in (41) yields

$$
\begin{aligned}
\frac{\partial \lambda(\Psi, f(t))}{\partial \|f_j\|^2} &= -\log \|f_j\|^2 - 1 \\
&= -\log 2 \|f_j\|^2 \qquad (56)
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial \|f_j\|^2}{\partial c_k} &= \frac{\partial}{\partial c_k} \sum_l f_{j,l}^2 \\
&= 2 \sum_l f_{j,l} \frac{\partial f_{j,l}}{\partial c_k}. \qquad (57)
\end{aligned}
$$

We have so far

$$
\frac{\partial \lambda(\Psi, f(t))}{\partial c_k} = -2 \sum_j \sum_l \log 2 \|f_j\|^2 f_{j,l} \frac{\partial f_{j,l}}{\partial c_k}. \qquad (58)
$$

Since

$$
\frac{\partial f_{j,l}}{\partial c_k} = \left\langle f(t), \frac{\partial \psi_{j,l}}{\partial c_k} \right\rangle, \qquad (59)
$$

the result from the previous lemma concludes the proof.

$\square$

This theorem demonstrates an explicit relation among the gradient of the additive information measure, parameter set $\{c_k\}$ and the measured signal $f(t)$. It points to the direction for updating the parameter set to reduce the information measure and thus will facilitate the search for the optimal wavelet basis. 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 mother 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 5 describes this process.
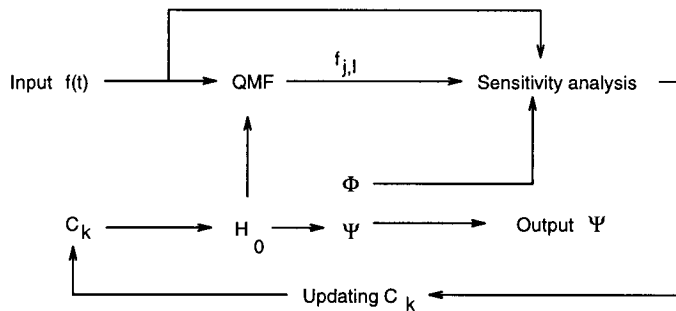
Figure 5: Flow chart for computing $\psi(t)$

# 4   Existence of the Optimal Wavelet Basis

An existence theorem is proven in this section through the decomposition of unitary matrices and lattice structures of QMF.

## 4.1   Decomposition of Unitary Matrices

We shall show that the paraunitary matrix discussed earlier can be decomposed to its minimum forms and thus be parameterized for the optimization purpose. We first introduce some basic definitions and notations to facilitate the description of our problems. Let $F$ be a subfield of $C$ closed under complex conjugation, let $U$ be the multiplication group of two by two unitary matrices over the ring of Laurent polynomials $F(z)$ with $|z| = 1$. Define a natural conjugate operation denoted by $\tilde{\phantom{x}}$ which takes the complex conjugate of $F$ and replaces $z$ by its reciprocal. Denote $\hat{U}$ to be the subgroup of matrices in $U$ which have determinant 1 and equal identity matrix when evaluated at $z = 1$. Use $UI$ to represent the two by two multiplication group of unitary matrices over the subfield $F$. It is obvious that the aliasing component (AC) matrix in (26) belongs to $U$. From the previous definition of a paraunitary matrix, it can easily be converted as a unitary matrix here with a scaling factor $d$. Without further explanation, we shall identify those two in the rest of this section.

**Theorem 4.1 (Pollen [12])** *For every element $M \in U$, there exist unique*

$$A \in UI, \; B \in \begin{bmatrix} 1 & 0 \\ 0 & z^k \end{bmatrix}, C \in \hat{U}$$

19

*with $k \in Z$ such that $M = ABC$.*

The uniqueness holds upon the choice of the parameter $k$ of matrix $B$. To proceed further, we define the order of a matrix in $U$ to be the highest order of the Laurent polynomial appearing in the matrix. This theorem shows [12] that we can factor the matrix in $U$ into the product of a zero order polynomial matrix, a monomial matrix and a matrix which has unit determinant and is equal to the identity matrix upon evaluation at $z = 1$. The process can be repeated when the order of the factor $C$ is greater than one. Hence, we may have

$$
\begin{aligned}
M &= ABC \\
&= ABA_1B_1C_1 \\
&= ABA_1B_1A_2B_2 \cdots A_mB_mC_m
\end{aligned}
\tag{60}
$$

where $m \leq deg(M)$. The matrices in $U$ have a nice property of being able to be expressed as a unique product of a minimal number of simple factors in $\hat{U}$.

**Definition 4.1** *A matrix $X \in \hat{U}$ is called a simple factor if its degree is not more than one.*

**Lemma 4.1** *A $2 \times 2$ matrix $X$ is a simple factor if*

$$
X = \begin{bmatrix} a + (1-a)z & b - bz \\ -(b - bz) & a + (1-a)/z \end{bmatrix}
$$

*with $|a - \frac{1}{2}|^2 + |b|^2 = (\frac{1}{2})^2$.*

The set of simple factors is a subset of the collection of all the degree 0 or 1 elements of $UI$.

*Proof:*

Obviously, $X$ is an identity matrix when $z = 1$. The circle condition regarding $a$ and $b$ in the theorem proves that $det(X) = 1$.

$\square$

The topology of the factor is shown here to be a circle when the field is real, or a 2-sphere for complex field. The set of all simple factors is a subset of the set composed of all the matrices of degree zero and one.

The unique factorization theorem for second order matrices in $\hat{U}$ due to Pollen is introduced here without proof for the parameterization of the unitary matrices.

**Theorem 4.2 (Pollen [12])** *Let $M \in \hat{U}$ with deg(M)$> 0$, then there exist simple factors $A_i$ and $B_i$ such that*

$$M = \tilde{A}_1 B_1 \tilde{A}_2 B_2 \cdots \tilde{A}_m B_m.$$

*This expression is the factorization of $M$ as a product of $2m$ simple factor elements and factor inverse. Furthermore, the factorization is unique. A slightly weaker form of the decomposition is that $M$ is factorized in to a product of $m$ elements of degree one, i.e., $M = C_1 C_2 \cdots C_m$. However, the elements may not be simple factors.*

This theorem implies that the original parameters of the matrix can be parameterized by the factors in a simple form and the factors can be derived from the coefficient matrices of the original unitary matrix $M$. This property is especially useful since we consider the paraunitary property of the filter transfer function $H_0$ and $H_1$ on the unit circle in the $z$ plane which falls into the realm of this unique factorization theorem.

## 4.2   Givens Rotation and QMF Lattice Structures

This section intends to transform the constrained optimization of the optimal wavelet function into an unconstrained problem through parameterization. The decomposition theorem shows that it is possible to formulate the problem with $z$ on the unit circle, which is convenient to be treated by trigonometry functions. We need to find an appropriate mapping between the parameters of the QMF and the corresponding set of unconstrained parameters on the unit circle. We then can apply the numerical procedure discussed earlier to find the right QMF bank and thus to find the optimal wavelets. One of the convenient ways is to use Givens rotation as follows.

From the decomposition theorem 4.1, for any matrix in $U$ with degree greater than or equal to one, we can apply the theorem again until all the factors have degree zero or one as defined above.
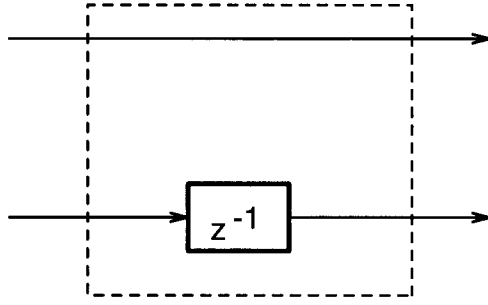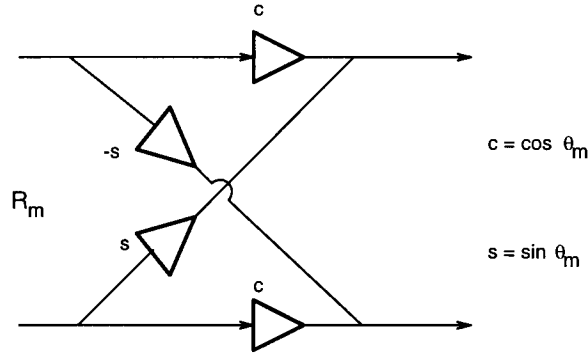
Figure 6: Block structure of the unit factor



Figure 7: A structure block of Givens rotation

The resulting factors are of the form of a $2 \times 2$ paraunitary transfer matrix

$$\Lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix}. \tag{61}$$

Figure 6 shows the block structure of this basic factor in $U$. We notice that the product of two unitary matrices is still unitary. We further introduce the transfer matrix

$$R_m = \begin{bmatrix} cos\theta_m & sin\theta_m \\ -sin\theta_m & cos\theta_m \end{bmatrix}, \tag{62}$$

where $\theta_m$ is real. This is the Givens rotation which is used to parameterize the AC matrix on the unit circle. The block structure of the Givens rotation is shown in Figure 7. It can be easily shown that the determinant of the AC matrix (26) is $\beta z^{-K}$ where $\beta$ is a constant coefficient for the low pass filter $H_0(z)$ of degree $K$. At most, the AC matrix has $K$ such factors in its decomposition form.

Let $H_K(z)$ denote the $2 \times 2$ aliasing component matrix of degree $K$; the factorization through

Givens rotations is given by

$$H_K(z) = \alpha R_0 \Lambda_0 R_1 \Lambda_1 \cdots R_{K-1} \Lambda_{K-1} R_K \Lambda_K \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix}, \tag{63}$$

where $\alpha$ is a constant. The structure of the factorization is a cascade system in which the basic blocks $R_i$ and $\Lambda(z)$ appear alternatively.

The parameters of the QMF bank, i.e., of the low pass filter $H_0(z)$, are functions of those angles $\theta_m$ with $m$ runs from zero to $K$. The number of parameters of the QMF and that of the factorization through applying Givens rotation are the same. Hence, the parameter set $\{c_k\}$ can be expressed as a continuous nonlinear function of those angles which take their values in the closed interval $[0, 2\pi]$. We represent the relation as

$$C = P(\theta) \tag{64}$$

where $P$ denotes the nonlinear function of the parameterization, $C$ and $\theta$ represent the corresponding $K + 1$ dimensional vectors of parameters. We can minimize (42) through a gradient based method with

$$\frac{\partial \lambda(\Psi, f(t))}{\partial \theta_k} = \sum_{m=0}^{K} \frac{\partial \lambda(\Psi, f(t))}{\partial c_m} \frac{\partial c_m}{\partial \theta_k}. \tag{65}$$

The optimization of the cost functional (42) turns into an unconstrained optimization on the compact set $[0, 2\pi]$.

With the parameterization of the low pass filter $H_0(z)$ by the rotation vector $\theta$ over set $[0, 2\pi]^{K+1}$, the existence of the global minimum is guaranteed although the gradient based algorithm may stop at a local minimum. The parameter set $C$ is computed from the final set of $\theta$'s to further improve the optimal mother wavelet $\psi(t)$ and the optimal compactly supported orthonormal wavelet basis. The next theorem is a main result of this paper.

**Theorem 4.3 (Existence of an optimal wavelet basis)** *Let $\lambda(\cdot, \cdot)$ be the additive information measure as defined in (41) and let $f(t) \in L^2(R)$ be an arbitrary finite energy function. Then there*

23

*exists an optimal mother wavelet function $\psi^*(t)$ such that, $\lambda(\Psi^*, f(t)) = \lambda_{min}$ where $\Psi^*$ is the wavelet basis generated by the dilations and shifts of $\psi^*(t)$.*

*Proof:*

By the definitions, $\lambda$ is continuous with respect to its arguments, $\Psi$, $f(t)$ and the parameter set $C$, the nonlinear function in (64) is a continuous bijection with $\theta \in [0, 2\pi]^{K+1}$, a compact set, as the consequence, the information measure parameterized by the angular vector $\theta$ is a continuous functional defined on a compact manifold $[0, 2\pi]^{K+1}$, hence, there exists a vector $\theta^* \in [0, 2\pi]^{K+1}$, at which, $\lambda$ attains its minimum. The optimal wavelet is thus determined through $C$, the coefficients of the low pass filter in the QMF bank.

$\square$

The existence theorem guarantees that the global optimal solution does exist. However, as mentioned earlier, the computation may end up at a local minimal depending on the starting point and thus yields a suboptimal solution which in many cases may still provide satisfactory performance.

## 5 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 mother wavelet function $\psi$ can be derived afterwards. Equipped with the theorems above, 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 5.1** *Computation of the optimal wavelet basis*

*Step 1: Set $i := 1$,*
$\qquad \lambda_0 := 0$,
$\qquad$ *mesh parameters $M, N$;*
$\qquad$ *Initialize vector $C_0$;*

24

*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 mother 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 an appropriate order. This is realized by a modified algorithm as follows.

**Algorithm 5.2** *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.

25

# 6 Conclusions

This paper has provided a direct approach to construct an optimal orthonormal wavelet basis with compact support for signal representation and compression. 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 mother 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 mother wavelet allows an explicit expression of information gradient with respect to the optimization parameters and thus paves the way to efficient basis selection algorithms.

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, system identification and progressive system modeling. It also finds applications in multiresolution feature extraction and classification of signals. Our methodology provides means of improving current wavelet applications by providing a practical approach for generating the optimal wavelet basis. In the context of pattern recognition, it is also a way to construct a feature space and to partition the signal space according to its representatives. The results of this research shall appear in a forthcoming paper. The parameterization of the cost functionals described above is very helpful; other forms of measures or cost functions may be introduced depending on the contexts of actual physical problems.

# References

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

[2] 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.

[3] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

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

[5] C. Herley. *Wavelets and Filter Banks*. PhD thesis, Columbia University, 1993.

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

[7] A.K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., Englewood Cliffs, NJ 07632, 1989.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] D. Pollen. $SU_I(2, F[z, 1/z])$ for $F$ a subfield of C. *Journal of the American Math. Society*, 3(3):611–624, July 1990.

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

[14] 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.

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

[16] 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.