Appears in: IEEE Trans. on CAS II - April 1994

OPTIMAL WAVELET REPRESENTATION OF SIGNALS AND THE WAVELET SAMPLING THEOREM

R. A. Gopinath, J. E. Odegard, and C. S. Burrus
Department of Electrical and Computer Engineering,
Rice University, Houston, TX-77251

CML TR92-05
15th April '92
Revised Aug 12 '93
Revised Dec 20 '93

OPTIMAL WAVELET REPRESENTATION OF SIGNALS AND THE WAVELET SAMPLING THEOREM†

R.A. Gopinath, J.E. Odegard, and C.S. Burrus
Department of Electrical and Computer Engineering,
Rice University, Houston, TX-77251

CML TR92-05
15th April '92
Revised Aug 12 '93
Revised Dec 20 '93

Abstract

The wavelet representation using orthonormal wavelet bases has received widespread attention. Recently M-band orthonormal wavelet bases have been constructed and compactly supported M-band wavelets have been parameterized [15, 12, 32, 17]. This paper gives the theory and algorithms for obtaining the optimal wavelet multiresolution analysis for the representation of a given signal at a predetermined scale in a variety of error norms [23]. Moreover, for classes of signals, this paper gives the theory and algorithms for designing the *robust* wavelet multiresolution analysis that minimizes the worst case approximation error among all signals in the class. All results are derived for the general M-band multiresolution analysis. An efficient numerical scheme is also described for the design of the optimal wavelet multiresolution analysis when the least-squared error criterion is used. Wavelet theory introduces the concept of scale which is analogous to the concept of frequency in Fourier analysis. This paper introduces *essentially scalelimited signals* and shows that bandlimited signals are essentially scalelimited, and gives the *wavelet sampling theorem*, which states that the scaling function expansion coefficients of a function with respect to a M-band wavelet basis, at a certain scale (and above) completely specify a bandlimited signal (i.e., behave like Nyquist (or higher) rate samples).

Contact Address:
Ramesh A. Gopinath
Speech Recognition Group
IBM T.J. Watson Research
Hawthorne, NY
Phone (914) 784-6548
email: rameshg@watson.ibm.com

Contact Address:
Jan Erik Odegard
Department of ECE,
Rice University,
Houston, TX-77251
Phone (713) 527-8750 x3508
email: odegard@rice.edu

†This work was supported by AFOSR under grant 90-0334 funded by DARPA and Bell-Northern Research

1 Introduction

Orthonormal (ON) wavelet bases give a novel way to decompose a signal into orthogonal frequency channels that have the same bandwidth on a logarithmic scale [6, 7, 15]. The wavelet basis functions are generated by the translates and dilates (by powers of *two*) of a single function, the *wavelet* $\psi_1(t)$. Such wavelet bases will be referred to as 2-band wavelets. All 2-band compactly supported ON wavelets can be parameterized by a finite number of parameters [6, 7]. For a given application, therefore, one can design the optimal 2-band wavelet with respect to any objective criterion by a finite dimensional optimization. Moreover the parameterization also gives an efficient algorithm for the computation of the (discrete) wavelet transform (expansion coefficients in the wavelet basis). Wavelet bases also (unlike the Fourier basis) constitute unconditional bases for all the L^p spaces, a fact of great importance in harmonic analysis [6]. Recently, M-band orthonormal wavelet bases, where the basis functions are generated by translates and dilates (by powers of M) have been constructed. An important difference between the cases $M = 2$ and $M > 2$ is that in the latter the ON wavelet basis is generated by more than one wavelet (precisely $(M-1)$ of them). All M-band wavelet bases discussed in this paper will be associated with a special function called the *scaling function* $\psi_0(t)$.

The potential applications of wavelets have received widespread attention in signal processing and applied mathematics [3, 31, 22, 2, 5, 9, 14, 9, 1]. Typically the first step in wavelet analysis is the approximation of a signal in a wavelet basis at some prescribed scale. This paper addresses the following two problems:

1. Given a signal $f(t)$, a dilation factor M, and a prescribed scale J, what is the optimal wavelet representation (among all compactly supported wavelets of a fixed support) that represents $f(t)$ at resolution J. The optimality is measured with respect to minimization of frequency domain L^p norm of the approximation error. The approximation at resolution J depends only on the scaling function $\psi_0(t)$.

2. Given a class of signals, what is the choice of wavelets that minimizes the worst case approximation error among all the signals in the class? M, J and the support size of the wavelets are fixed as in the previous problem. The class of signals considered are the frequency domain L^p class.

Problem 1 has been addressed by Tewfik, Sinha and Jorgensen [29] for the special case $M = 2$ (i.e., for Daubechies' orthonormal wavelet bases). Since the expression for the exact approximation error is unwieldy and complicated, the approach in [29] is to obtain upper and lower bounds on the approximation error, and devise a numerical scheme to minimize the upper bound. This gives a *sub-optimal* solution to the approximation problem that is relatively efficient to implement. Our approach to Problem 1 is based on the following crucial assumption: the signals being analyzed are bandlimited. This constraint, we show, can be used to considerably simplify the expression for the approximation error. Using this expression, and the parameterization of M-band wavelet tight frames (WTFs) [15], we devise an efficient numerical scheme to solve Problem 1 and illustrate it with examples, paying particular emphasis to the case of L^2 error norm. As for Problem 2, we show that the approximation error can be considered as an operator acting on any L^p class of signals. Then solving Problem 2 is equivalent to minimizing the induced norm of this operator.

In many signal processing applications the L^2 or energy norm is the most useful measure of approximation error. However, the frequency domain L^∞ norm and more generally L^p norms are sometimes useful in signal processing approximation problems (for example the Chebychev error in filter design). Since the solutions to both problems considered in this paper for the general L^p case is not more complicated than that of the L^2 case, we present the general results (with examples).

In wavelet analysis *resolution* or *scale* plays a role analogous to *frequency* in Fourier analysis. A natural question is whether a notion similar to bandlimitedness exists in wavelet analysis. This question has been investigated in [23], where the authors introduce the notion of *essentially scalelimited* signals. Our solution to Problem 2 shows that all bandlimited signals are essentially scalelimited (i.e., the class of essentially scalelimited signals is rich).

Practically wavelet expansion coefficients are computed as follows: First scaling function expansion coefficients at some scale J are computed. The wavelet expansion coefficients (for all scales $j < J$) are computed from the scaling expansion coefficients using a filter bank. In most applications the samples of the signals are themselves taken to be the scaling expansion coefficients. There are several ways to see why this approximation works well - for large scales the scaling function approximates the delta function, the samples give a third order approximation of the expansion coefficients etc. All this suggests interpreting the expansion coefficients as generalized samples of a signal. Indeed, for bandlimited signals (using the intuition gained from the solution of Problem 1) we show that the scaling function expansion coefficients completely specify the signal (i.e., have the same information as the Nyquist rate samples), a result which we refer to as the *wavelet sampling theorem.* The main purpose of this theorem is to give an interpretation for the scaling expansion coefficients as generalized samples.

1.1 Outline of the paper

Section 2 gives a review of M-band compactly supported wavelet tight frames and their parameterization. Section 3 derives the main expression for the approximation error, and describes the design of optimal and robust wavelets. Section 4 outlines an efficient numerical algorithm for the design of optimal wavelets with the L^2 error norm and gives examples of optimal designs.

1.2 Notational Preliminaries

We denote the set of real numbers and integers by $\mathbb{R}$ and $\mathbb{Z}$ respectively. $L^p(\mathbb{R})$ denotes the vector space of measurable one-dimensional functions, $f(t)$, satisfying

$$\int_{\mathbb{R}} dt\, |f(t)|^p < \infty.$$

In particular for $p = 2$ and $f(t), g(t) \in L^2(\mathbb{R})$, the $L^2(\mathbb{R})$ inner product is denoted by

$$\langle f(t), g(t) \rangle = \int_{\mathbb{R}} dt f(t) g^*(t).$$

The $L^p(\mathbb{R})$ norm of a function, $f(t) \in L^p(\mathbb{R})$, is given by

$$\|f\|_p^p = \int_{\mathbb{R}} dt\, |f(t)|^p$$

and for $p = 2$ we denote the $L^2(\mathbb{R})$ norm by $\|\cdot\| = \|\cdot\|_2$ (i.e., the subscript may be omitted) which in terms of the $L^2(\mathbb{R})$ inner product becomes: $\|f\|_2^2 = \langle f(t), f(t) \rangle$. $L^p_{loc}(\mathbb{R})$ will denote the space of functions that are locally $L^p(\mathbb{R})$ (i.e., when viewed through any compactly supported or finite duration window). The Fourier transform of a function $f(t) \in L^2(\mathbb{R})$ is denoted by $\hat{f}(\omega)$ and defined by

$$\hat{f}(\omega) = \int dt f(t) e^{-\imath \omega t},$$

where $\imath = \sqrt{-1}$. The convolution of two functions $f(t), g(t) \in L^2(\mathbb{R})$ is denoted

$$f * g(t) = (f(\tau) * g(\tau))(t) = \int_{\mathbb{R}} d\tau f(\tau) g(t - \tau).$$

For a sequence $h(n)$, $H(z)$ will denote its $\mathcal{Z}$-transform. $[\downarrow M]$ will denote the M-fold downsampling defined by

$$y(n) = [\downarrow M]\, x(n) = x(Mn)$$

and $[\uparrow M]$ denotes the M-fold upsampling operator defined by

$$y(n) = [\uparrow M]\, x(n) = \begin{cases} x(M^{-1}n) & n \in \mathcal{L}(M) \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{L}(M)$ is the lattice generated by M [30, 11]. Finally, we will also adhere to the traditional abuse of notation and denote the Fourier transform $H(e^{\imath\omega})$ of the sequence $h(n)$ by $H(\omega)$.

2 M-band Orthonormal Wavelet Bases

M-band compactly supported ON wavelet bases are characterized by a sequence known as the unitary scaling vector [15, 12]. A unitary scaling vector h_0 of length N is characterized by the following linear and quadratic constraints:

$$\sum_{k=0}^{N-1} h_0(k) = \sqrt{M} \tag{1}$$

and

$$\sum_{k=0}^{N-1} h_0(k) h_0(k+Ml) = \delta(l). \tag{2}$$

N is generally of the form $N = MK$ (and can always be assumed to be of this form by padding with zeros). Given a unitary scaling vector, wavelet bases are constructed by the following recipe [15]. First, a *unique* scaling function $\psi_0(t)$ is constructed. Then a set of $(M-1)$ sequences h_i, $i \in \{1, \ldots, M-1\}$, (the *unitary wavelet vectors*) is constructed from h_0. When $(M = 2)$, h_0 uniquely determines h_1. In general the unitary wavelet vectors are determined by $\binom{M-1}{2}$ parameters. The wavelets are then constructed by taking linear combinations of translates of the scaling function, each combination determined by one unitary wavelet vector. The $(M-1)$ wavelets, their dilates by powers of M, and their translates form a wavelet tight frame. Most wavelet tight frames turn out to be ON wavelet bases. The length N of h_0 is related linearly to the support of the scaling functions and wavelets. In summary: a M-band wavelet basis is completely specified by the choice of the unitary scaling vector and the $\binom{M-1}{2}$ parameters determining the choice of the unitary wavelet vectors.

All scaling vectors of length $N = MK$ can be parameterized as described below. Let $h_{0,k}(n) = h_0(Mn + k)$ be the polyphase component sequences of h_0. That is,

$$H_0(z) = \sum_{k=0}^{M-1} z^{-k} H_{0,k}(z^M), \tag{3}$$

where $H_{0,k}(z)$ is the $\mathcal{Z}$-transform of $h_{0,k}(n)$. For each k, $H_{0,k}(z)$ is a polynomial of degree $K-1$ in z^{-1}. The quadratic constraint from Eqn. 2 can now be expressed in terms of the polyphase components as

$$\sum_{k=0}^{M-1} H_{0,k}(z) H_{0,k}(z^{-1}) = 1. \tag{4}$$

Using this fact one can obtain a parameterization of h_0 as follows [15].

Fact 1 (Scaling Vector Characterization) *Every unitary scaling vector h_0 of length $N = MK$ can be parameterized by $(K-1)$ unit vectors v_i (the Householder parameters of h_0).*

$$\begin{bmatrix} H_{0,0}(z) \\ H_{0,1}(z) \\ \dots \\ H_{0,M-1}(z) \end{bmatrix} = \frac{1}{\sqrt{M}} \prod_{i=1}^{K-1} \left[I - v_i v_i^T + z^{-1} v_i v_i^T \right] \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}. \tag{5}$$

The scaling function in turn determines a multiresolution analysis that is a sequence of successive approximation spaces for $L^2(\mathbb{R})$. Thus, choosing a scaling vector is equivalent to choosing a multiresolution analysis. Moreover, from the scaling function, one can define $M-1$ wavelets that give rise to a wavelet tight frame. By appropriate choice of Householder parameters one obtains a desired scaling vector which give rice to a unique scaling function. Furthermore, from the desired scaling vector on can obtain (non-unique) wavelet vectors and associated wavelets. The precise result is stated in the following fact proved in [15].

Fact 2 (Wavelet Tight Frames Theorem) *Given a length $N = MK$, M-band, unitary scaling vector h_0, there exists a unique, compactly supported scaling function, $\psi_0(t) \in L^2(\mathbb{R})$, with support in $[0, \frac{N-1}{M-1})$ determined by the following scaling recursion:*

$$\psi_0(t) = \sqrt{M} \sum_{k=0}^{N-1} h_0(k) \psi_0(Mt - k). \tag{6}$$

Moreover, there exist $(M-1)$ wavelet vectors h_i, all of the same length, N, that satisfy the equation

$$\sum_{k=0}^{N-1} h_i(k) h_j(k + Ml) = \delta(l)\delta(i-j). \tag{7}$$

The wavelet vectors are non-unique and parameterized by $\binom{M-1}{2}$ parameters. If for each wavelet vector, $h_i, i \in \{1, 2, \dots M-1\}$ we define the corresponding wavelets, $\psi_i(t)$, compactly supported in $[0, \frac{N-1}{M-1})$ by

$$\psi_i(t) = \sqrt{M} \sum_{k=0}^{N-1} h_i(k) \psi_0(Mt - k), \tag{8}$$

then $\{\psi_{i,j,k}(t)\}$, defined by

$$\psi_{i,j,k}(t) = M^{j/2} \psi_i(M^j t - k), \tag{9}$$

forms a tight frame for $L^2(\mathbb{R})$. In other words, for all $f \in L^2(\mathbb{R})$,

$$f(t) = \sum_{i=1}^{M-1} \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{i,j,k}(t) \rangle \, \psi_{i,j,k}(t). \tag{10}$$

Also,

$$f(t) = \sum_k \langle f, \psi_{0,0,k}(t)\rangle \, \psi_{0,0,k}(t) + \sum_{i=1}^{M-1} \sum_{j=1}^{\infty} \sum_{k\in\mathbb{Z}} \langle f, \psi_{i,j,k}(t)\rangle \, \psi_{i,j,k}(t). \tag{11}$$

Remark 1. Fact 2 in conjunction with Fact 1 gives a neat characterization of all WTFs in terms of the Householder parameters v_i (that determines h_0) and the $\binom{M-1}{2}$ parameters (determining h_i, $i \neq 0$). *No such clean parameterization exists if one desires ON wavelet bases instead of WTFs.* One usually imposes additional restrictions on the scaling vector to obtain ON wavelet bases. For example, when $M = 2$ one could use the Mallat condition ($|H_0(\omega)| > 0$ for $|\omega| \leq \frac{\pi}{2}$), or the Daubechies' regularity conditions [6]:

$$H_0(\omega) = \left(\frac{1+e^{-\imath\omega}}{2}\right)^K P(\omega) \quad \text{with} \quad |P(\omega)| \leq 2^{K-\frac{1}{2}}.$$

For the interested reader corresponding conditions for $M > 2$ can be found in [11, 27]. The Mallat and Daubechies' type conditions are sufficient but not necessary for orthonormality. Necessary and sufficient conditions for orthonormality when $M = 2$ have been obtained by Cohen and Lawton independently [4, 20], and since generalized for $M > 2$ [11, 27].

Remark 2. The scaling function constructed from h_0 is unique because of the following reason. The scaling recursion (Eqn. 6) in the Fourier transform domain can be written as the following infinite product.

$$\hat{\psi}_0(\omega) = H_0\left(\frac{\omega}{M}\right) \hat{\psi}_0\left(\frac{\omega}{M}\right) = \prod_{j=1}^{\infty} \left[\frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^j}\right)\right] \hat{\psi}_0(0) \tag{12}$$

By the uniqueness of the Fourier transform there can only be one such function $\psi_0(t)$ (provided one can take Fourier transforms). Fact 2 promises a compactly supported $\psi_0(t) \in L^2(\mathbb{R})$. Therefore, $\psi_0(t)$ is Fourier transformable and hence unique. There are many other ways to generate M-band scaling functions. For example, Madych [21] constructs scaling functions starting with a multiresolution analysis point of view. This approach will not in general give unique scaling functions. However, even in this approach there is a one-to-one correspondence between a given scaling function and its corresponding unitary scaling vector. The different scaling functions correspond to different unitary scaling vectors.

Remark 3. Fact 2 also defines a multiresolution analysis (sequence of approximation spaces) for $L^2(\mathbb{R})$. If we define the spaces

$$W_{i,j} = \operatorname*{Span}_k \{\psi_{i,j,k}\}, \tag{13}$$

then Fact 2 implies that

$$W_{0,j} = \bigoplus_{i=0}^{M-1} W_{i,j-1} \tag{14}$$

and

$$\lim_{j \to \infty} \overline{W_{0,j}} = L^2(\mathbb{R}). \tag{15}$$

The spaces $W_{0,j}$ (determined only by the scaling function or equivalently the scaling vector) form a multiresolution analysis for $L^2(\mathbb{R})$. *Given h_0, there exists a unique $\psi_0(t)$ and hence a unique multiresolution analysis $W_{0,j}$. However, given $W_{0,j}$ there may exist other scaling functions giving rise to $W_{0,j}$. Thus scaling functions uniquely determine the multiresolution analysis, while the converse is not true (see [21] for examples).* When $M = 2$, we get 2-band WTFs that include the orthonormal wavelet bases of Daubechies [6]. A fundamental difference between $M = 2$ and $M > 2$ is that in the former case the scaling function $\psi_0(t)$ uniquely determines the wavelet $\psi_1(t)$ (notice $\begin{pmatrix} M-1 \\ 2 \end{pmatrix} = 0$). When $M > 2$, the wavelets are non-unique and parameterized by $\begin{pmatrix} M-1 \\ 2 \end{pmatrix}$ parameters. This paper does not discuss the design of wavelets, only the design of the multiresolution analysis $W_{0,j}$, or equivalently the scaling vector h_0.

Most M-band WTFs are ON bases [12, 11]. Since some of the results in this paper are derived for ON wavelet bases we are interested in precise necessary and sufficient conditions for a WTF to be an ON basis. The precise result is stated in the following fact and a proof may be found in [11, 27].

Fact 3 (Characterization of ON Bases) *A WTF is an ON basis iff $A(z) = 1$ is the unique polynomial solution to the equation*

$$[\downarrow M]\, H_0(z) H_0(z^{-1}) A(z) = A(z). \tag{16}$$

Notice that when $A(z) = 1$, Eqn. 16 is essentially Eqn. 2. Thus $A(z) = 1$ is always a solution of Eqn. 16. If no other solution exists then the WTF is an ON basis. A convenient numerical test for orthonormality is the following reformulation of Fact 3. Let $r(n)$ be the autocorrelation sequence of $h_0(n)$. Then, the WTF forms an orthonormal basis iff the *Lawton matrix*, Q, (after Lawton who constructed it for the 2-band case [20]) defined by

$$q_{i,j} = \begin{cases} r(M(i-1)) & \text{for } j = 1 \\ r(M(i-1)+j-1) + r(M(i-1)-j+1) & \text{for } 2 \le j \le N. \end{cases} \tag{17}$$

has a unique eigenvector of eigenvalue 1. Though, our theory of optimal design of h_0 will be derived for multiresolution analysis associated with ON bases, the numerical design will be over multiresolution analyses associated with WTFs (because the latter can be parameterized using the Householder parameters). Among all unitary scaling vectors of length N the subset that *does not* give rise to an ON basis is very sparse (being an algebraic subset satisfying Eqn. 16 it has measure zero!). Therefore, optimizing over all unitary scaling vectors will practically never violate the theory - yet, it is an important engineering fact to keep in mind. Another useful characterization of orthonormality of a WTF is that the scaling function and its translates form an orthonormal system, or equivalently in the Fourier transform domain [6, 15]

$$\sum_{k=-\infty}^{\infty} \left|\hat{\psi}_0(\omega + 2\pi k)\right|^2 = 1. \tag{18}$$

Since the wavelets are compactly supported the following wavelet expansion holds for any function $f(t) \in L^1_{loc}(\mathbb{R})$ and not just $f(t) \in L^2(\mathbb{R})$: Hence, recal from Fact 2 (Eqn. 10) that

$$f(t) = \sum_{i=1}^{M-1} \sum_{j,k\in\mathbb{Z}} \langle f, \psi_{i,j,k}\rangle \, \psi_{i,j,k}(t),$$

From the multiresolution analysis associated with the wavelet basis (Eqn. 14) is given by

$$\sum_{k\in\mathbb{Z}} \langle f, \psi_{0,J,k}\rangle \, \psi_{0,J,k}(t) = \sum_{i=1}^{M-1} \sum_{j=-\infty}^{J} \sum_{k\in\mathbb{Z}} \langle f, \psi_{i,j,k}\rangle \, \psi_{i,j,k}(t),$$

and hence

$$f(t) = \sum_{k\in\mathbb{Z}} \langle f, \psi_{0,J,k}\rangle \, \psi_{0,J,k}(t) + \sum_{i=1}^{M-1} \sum_{j=J+1}^{\infty} \sum_{k\in\mathbb{Z}} \langle f, \psi_{i,j,k}\rangle \, \psi_{i,j,k}(t).$$

A given function $f(t)$ can be expressed as a linear combination of wavelets $\{\psi_{i,j,k}\}$, $i \neq 0$, or as a linear combination of scaling functions at scale J, namely $\{\psi_{0,J,k}\}$, and wavelets for all scales $j > J$, namely $\{\psi_{i,j,k}\}$, $j > J$. A scale J approximation of $f(t)$ (i.e., approximation in $W_{0,J}$) depends only on the scaling function and hence h_0 (not on the wavelets!).

3 The Approximation Error

In practical wavelet analysis, signals are represented by their wavelet expansion coefficients, $\langle f, \psi_{i,j,k}\rangle$, where the scale index j and the time index k are allowed to run only over a finite set. If $f(t)$ is of finite duration, k runs over a finite set for any fixed j. If $f(t)$ is essentially bandlimited (i.e., all but a fraction ϵ of its energy is within some frequency band) then it is essentially scalelimited (as we will show later) and hence $\langle f, \psi_{i,j,k}\rangle$, $i \neq 0$ can be ignored for large j. Therefore, practically all functions $f(t)$ can be approximated at scale J (i.e., by a function in $W_{0,J}$) for some large J. That is

$$\begin{aligned} f(t) &\approx \sum_k \langle f, \psi_{0,J,k}\rangle\, \psi_{0,J,k}(t) \\ &= \sum_{i=1}^{M-1} \sum_{j=-\infty}^{J} \sum_{k\in\mathbb{Z}} \langle f, \psi_{i,j,k}\rangle\, \psi_{i,j,k}(t). \end{aligned} \tag{19}$$

An important problem in wavelet analysis is to find the best wavelet multiresolution analysis that approximates a given signal in some norm. Another related problem, especially when trying to study a class of signals, is to find the best wavelet multiresolution analysis that minimizes the worst case approximation error over all signals in a class. This multiresolution analysis, if it exists, will be referred to as the "robust wavelet multiresolution analysis". The former problem has been addressed by Tewfik, Sinha and Jorgensen (all errors being measured in the frequency domain) [29]. In that paper the authors derive upper bounds for the L^2 approximation error of a given signal approximated at a desired scale J in the special case when $M = 2$. In this paper, instead of an upper bound, we obtain explicit expressions for the approximation error in various norms. Moreover, we show that under a crucial bandlimitedness assumption on the signal being analyzed, it is possible to substantially simplify the error expressions. We then describe the design of optimal and robust (for classes of signals) wavelet multiresolution analysis of signals. The essential idea in the robust case is to model the error as the output of an error operator acting on the function, with robustness being achieved by minimizing the induced norm of the operator.

We now derive convenient expressions for the Fourier transforms of Pf (the approximation in $W_{0,J}$) and $Qf(t)$ (the approximation error). First define the following Fourier transform pair $a(t)$ and $\hat{a}(\omega)$:

$$a(t) = \int_{\mathbb{R}} d\lambda f(\lambda) M^J \psi_0(M^J\lambda - M^J t) = M^J(f(\cdot) * \psi_0(-M^J\cdot))(t) \tag{20}$$

$$\hat{a}(\omega) = \hat{f}(\omega)\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right), \tag{21}$$

then the scaling expansion coefficients $\langle f, \psi_{0,J,k}\rangle$ are samples of $a(t)$ given by

$$\langle f, \psi_{0,J,k}\rangle = \int_{\mathbb{R}} d\lambda f(\lambda) M^{J/2}\psi_0(M^J\lambda - k) = M^{-J/2} a(M^{-J}k). \tag{22}$$

The Fourier transform of the sequence $a(M^{-J}k)$ is given by (periodization of $\hat{a}(\omega)$)

$$M^J \sum_{k\in\mathbb{Z}} \hat{a}(\omega + 2\pi M^J k) = M^J \sum_{k\in\mathbb{Z}} \hat{f}(\omega + 2\pi M^J k)\psi_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right) \tag{23}$$

and the approximation of $f(t)$ is

$$Pf(t) = \sum_{k \in \mathbb{Z}} \langle f, \psi_{0,J,k} \rangle \, \psi_{0,J,k}(t) = \sum_k a(M^{-J}k)\psi_0(M^J(t - M^{-J}k)). \tag{24}$$

In the Fourier transform domain the above convolution (Eqn. 24) reduces to a product.

$$\begin{aligned} \widehat{Pf}(\omega) &= \left[M^J \sum_k \hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^* \left(\frac{\omega + 2\pi M^J k}{M^J} \right) \right] \left[\frac{1}{M^J} \hat{\psi}_0 \left(\frac{\omega}{M^J} \right) \right] \\ &= \hat{\psi}_0 \left(\frac{\omega}{M^J} \right) \sum_k \hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^* \left(\frac{\omega + 2\pi M^J k}{M^J} \right) \end{aligned} \tag{25}$$

Furthermore, if we denote by $Qf(t)$ the approximation error, then $Pf \perp Qf$ and we have

$$Qf(t) = f(t) - \sum_k a(M^{-J}k) M^J \psi_0(M^J(t - M^{-J}k)) \tag{26}$$

or equivalently in the transform domain,

$$\widehat{Qf}(\omega) = \hat{f}(\omega) \left(1 - \left| \hat{\psi}_0 \left(\frac{\omega}{M^J} \right) \right|^2 \right) - \hat{\psi}_0 \left(\frac{\omega}{M^J} \right) \sum_{k \neq 0} \hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^* \left(\frac{\omega + 2\pi M^J k}{M^J} \right). \tag{27}$$

Eqn. 27 gives the approximation error for an arbitrary signal $f(t)$ when approximated at scale J. The approximation error, Qf, depends only on $\psi_0(t)$ or equivalently on h_0. The Householder parameterization for h_0 (Eqn. 5) gives a finite-dimensional parameterization of the error Qf.

3.1 Optimum and Robust Multiresolution Analysis

Having derived the approximation error, Qf, we are now in a position to obtain objective functions for the optimal design. We will in particular derive objective functions for the L^p optimization problem for arbitrary p. Additionally, for the design of an optimal robust multiresolution analysis we derive the induced operator norms for both L^p to L^p and L^p to L^1. All objective functions are derived and computed in the frequency domain even though in many applications L^p error norms in the time-domain are more meaningful. (E.g., the time domain L^∞ error norm gives the maximum error in the time-domain.) However, there does not seem to be an easy way to obtain equivalent (simple) expressions for the objective functions in the time-domain. For $2 \leq p \leq \infty$ one can bound the time-domain L^p errors using the Hausdorff-Young inequality.

For $g \in L^p$, $1 \leq p \leq 2$, the Hausdorff-Young inequality ([19, p. 333] or [28]) states that $\hat{g} \in L^q$ $(2 \leq q \leq \infty)$, where p and q are Hölder conjugates (i.e., $\frac{1}{p} + \frac{1}{q} = 1$), and

$$\|\hat{g}\|_q \leq C \, \|g\|_p \, , \tag{28}$$

where $C = (2\pi)^{1/q}\frac{p^{1/p}}{q^{1/q}} \approx (2\pi)^{1/q}$.

Now consider the Fourier transform pair $(f, \hat{f})$. If $\hat{f}$ is in L^p, then the Hausdorff-Young inequality says that

$$\left\|\hat{\hat{f}}\right\|_q = \|f\|_q \leq C \left\|\hat{f}\right\|_p . \tag{29}$$

For example the time-domain L^∞ error is bounded by the frequency-domain L^1 error. This shows that if time-domain errors are crucial, then one can use the techniques in this paper to *minimize error bounds* rather than the error themselves. However, for frequency domain error norms, the results derived are exact for bandlimited signals.

3.1.1 Optimal Multiresolution Analysis - Transform Domain L^p Error

Each Householder parameter v_i, since it is a unit vector, can be parameterized by $(M-1)$ angle parameters $\theta_{i,j}$, $j \in \{1, \ldots, M-1\}$.

$$(v_i)_j = \begin{cases} \left\{\prod_{l=0}^{j-1} \sin(\theta_{i,l})\right\} \cos(\theta_{i,j}) & \text{for } j \in \{0, 1, \ldots, M-2\} \\ \left\{\prod_{l=0}^{M-1} \sin(\theta_{i,l})\right\} & \text{for } j = M-1 \end{cases} \tag{30}$$

Let Θ be the $(M-1)(K-1)$ length vector obtained by stacking $\theta_{i,j}$. Then Problem 1 for the L^p error norm takes one of the following two (*different*) forms.

1.
$$\min_{\Theta} \left[\frac{1}{2\pi}\int_{\mathbf{R}} d\omega \left|\widehat{Qf}\right|^p\right]^{\frac{1}{p}} = \min_{\Theta} \left\|\widehat{Qf}\right\|_p \tag{31}$$

2.
$$\max_{\Theta} \left[\frac{1}{2\pi}\int_{\mathbf{R}} d\omega \left|\widehat{Pf}\right|^p\right]^{\frac{1}{p}} = \max_{\Theta} \left\|\widehat{Pf}\right\|_p \tag{32}$$

One minimizes the p^{th} norm of the approximation error, while the other maximizes the p^{th} norm of the approximant. When $p = 2$, the basis is ON, $Pf \perp Qf$ and $\left\|\widehat{Pf}\right\|^2 + \left\|\widehat{Qf}\right\|^2 = \left\|\hat{f}\right\|^2 = \|f\|^2$, the problems are equivalent. This paper considers only the minimization of the approximation error.

Eqn. 27 yields a complicated expression for $\left\|\widehat{Qf}\right\|_p$. If $f(t)$ is bandlimited the expression for $\left\|\widehat{Qf}\right\|_p$ can be simplified. The basic idea is to split the frequency axis into bins (say $\Omega_l = \left\{\omega \mid lM^J\pi \leq |\omega| \leq (l+1)M^J\pi\right\}$, $l \in \mathbb{Z}$) and express the integral for $\left\|\widehat{Qf}\right\|_p^p$ as a sum of parts, one for each bin. If $f(t)$ is bandlimited to $\Omega \stackrel{\text{def}}{=} \Omega_0$, each part in the sum can be simplified. A similar approach can be used to obtain an expression for $\left\|\widehat{Pf}\right\|_p^p$.

Consider signals $f(t)$ bandlimited to Ω. That is $\hat{f}(\omega) = 0$ for $\omega \notin \Omega$. Then

$$\begin{aligned}
\left\|\widehat{Qf}\right\|_p^p &= \frac{1}{2\pi}\int_{\mathbf{R}} d\omega \left|\hat{f}(\omega)\left(1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right) - \hat{\psi}_0\left(\frac{\omega}{M^J}\right)\sum_{k\neq 0}\hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^p \\
&= \frac{1}{2\pi}\sum_l \int_{\Omega_l} d\omega \left|\hat{f}(\omega)\left(1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right) - \hat{\psi}_0\left(\frac{\omega}{M^J}\right)\sum_{k\neq 0}\hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^p \\
&= \frac{1}{2\pi}\int_{\Omega} d\omega \left|\hat{f}(\omega)\left(1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right)\right|^p + \int_{\Omega} d\omega \left|\hat{f}(\omega)\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right)\right|^p \left[\sum_{l\neq 0}\left|\hat{\psi}_0\left(\frac{\omega - 2\pi M^J l}{M^J}\right)\right|^p\right] \\
&= \frac{1}{2\pi}\int_{\Omega} d\omega \left|\hat{f}(\omega)\right|^p \left\{\left|1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right|^p + \left|\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right)\right|^p \sum_{k\neq 0}\left|\hat{\psi}_0\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^p\right\} \\
&\stackrel{\text{def}}{=} \frac{1}{2\pi}\int_{\Omega} d\omega \left|\hat{f}(\omega)\right|^p S_p(\omega) \qquad (33)
\end{aligned}$$

where for convenience one defines

$$S_p(\omega) = \left\{\left|1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right|^p + \left|\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right)\right|^p \sum_{k\neq 0}\left|\hat{\psi}_0\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^p\right\}.$$

By a similar procedure one can also obtain

$$\begin{aligned}
\left\|\widehat{Pf}\right\|_p^p &= \frac{1}{2\pi}\int_{\mathbf{R}} d\omega \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\left[\sum_k \hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right]\right|^p \\
&= \int_{\Omega} d\omega \left|\hat{f}(\omega)\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^p \sum_{k\in\mathbb{Z}}\left|\hat{\psi}_0\left(\frac{\omega - 2\pi M^J k}{M^J}\right)\right|^p \\
&\stackrel{\text{def}}{=} \int_{\Omega} d\omega \left|\hat{f}(\omega)\right|^p T_p(\omega), \qquad (34)
\end{aligned}$$

where we define

$$T_p(\omega) = \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^p \sum_{k\in\mathbb{Z}}\left|\hat{\psi}_0\left(\frac{\omega - 2\pi M^J k}{M^J}\right)\right|^p.$$

This gives the most general expressions for $\left\|\widehat{Qf}\right\|_p^p$ and $\left\|\widehat{Pf}\right\|_p^p$ for bandlimited signals. The terms $S_p(\omega)$ and $T_p(\omega)$ depend only upon the choice of the scaling function and p. Thus the objective functions for both forms of Problem 1 can be obtained by computing $S_p(\omega)$ or $T_p(\omega)$ and then implementing the integral in Eqn. 33 and Eqn. 34.

When $p = 2$ and *one has an ON basis (not a WTF)* $S_2(\omega)$ and $T_2(\omega)$ take particularly simple forms which have a nice interpretation. When one has an ON basis from Eqn. 18 we have

$$S_2(\omega) = \left|1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right|^2 + \left|\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right)\right|^2 \sum_{k\neq 0}\left|\hat{\psi}_0\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^2$$

$$
\begin{aligned}
&= 1 - 2\left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2 + \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2 \sum_{k\in\mathbb{Z}} \left|\hat{\psi}_0\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^2 \\
&= 1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2, \qquad (35)
\end{aligned}
$$

and similarly

$$
T_2(\omega) = \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2. \qquad (36)
$$

Therefore when $p = 2$, the general expressions for the p^{th} norms, namely $\left\|\widehat{Qf}\right\|_2^2$ and $\left\|\widehat{Pf}\right\|_2^2$ are:

$$
\left\|\widehat{Qf}\right\|_2^2 = \frac{1}{2\pi}\int_\Omega d\omega \left|\hat{f}(\omega)\right|^2 \left(1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right) \qquad (37)
$$

$$
\left\|\widehat{Pf}\right\|_2^2 = \int_\Omega d\omega \left|\hat{f}(\omega)\right|^2 \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2. \qquad (38)
$$

Based on the orthogonality between Pf and Qf when $p = 2$ one get, as one should expect, that $\left\|\widehat{Pf}\right\|_2^2 + \left\|\widehat{Qf}\right\|_2^2 = \left\|\hat{f}\right\|_2^2$ (this is easily checked by examining Eqn. 38 and Eqn. 37). In summary: given a bandlimited signal f, a scaling function $\psi_0(t)$, a scale J and some p, we have obtained explicit expressions for $\left\|\widehat{Pf}\right\|_p^p$ and $\left\|\widehat{Qf}\right\|_p^p$. This gives an unconstrained optimization scheme to compute the optimal multiresolution analysis. Section 4 gives the details of a numerical scheme for Problem 1 and give examples illustrating that for smooth signals, K-regular multiresolution analysis is nearly optimal.

3.1.2 Wavelet Sampling Theorem

Another important consequence of the analysis in the previous section is the following *wavelet sampling theorem.* Shannon's sampling theorem states that signals $f(t)$ bandlimited to $\Omega = \left\{\omega \mid |\omega| \leq M^J\pi\right\}$ are uniquely determined by the samples $f(M^{-J}k)$. Under mild restrictions on the scaling function of an ON basis it turns out that the scaling function expansion coefficients at $W_{0,J}$, namely $\{\langle f, \psi_{0,J,k}\rangle\}$, act as *generalized samples* of $f(t)$ bandlimited to Ω.

That the scaling expansion coefficients *act* as generalized samples of signals has already been reported in [15] where the arguments are based more on intuition than precise mathematical reasoning. First notice that if we choose the *sinc* wavelet basis for which the sinc function is the corresponding scaling function, Shannon's sampling theorem may also be interpreted as follows: the scaling expansion coefficients $\langle f, \psi_{0,J,k}\rangle = M^{-J/2}f(M^{-J}k)$ completely determine $f(t)$ (since they are the Nyquist rate samples!).

The wavelet sampling theorem, besides giving an interpretation to $\langle f, \psi_{0,J,k}\rangle$ also justifies an assumption that is used in practical signal analysis: essentially that $\langle f, \psi_{0,J,k}\rangle \approx M^{-J/2} f(M^{-J}k)$. Two other reasons for this assumption may also be found in the literature - the first based on the idea that for sufficiently large J, $M^J \psi_0(M^J t)$ approaches the Dirac measure $\delta(t)$ and therefore $\langle f, \psi_{0,J,k}\rangle \approx M^{-J/2} f(M^{-J}k)$, and the second based on the fact that the samples $f(M^{-J}k)$ give a third order approximation (i.e., exact for quadratics) to $\langle f, \psi_{0,J,k}\rangle$ [13].

It is an interesting fact that even though the scaling function is not bandlimited (for example when it is compactly supported) $W_{0,J}$ for large J can still completely represent bandlimited signals provided the hypotheses of the wavelet sampling theorem are satisfied. Most scaling functions (2-band Daubechies' wavelets and the M-band wavelets in [27, 12, 16, 26]) satisfy the conditions of the wavelet sampling theorem.

Theorem 1 *Let $f(t)$ be bandlimited to Ω (i.e., $\hat{f}(\omega) = 0$ for $\omega \notin \Omega$). Then $f(t)$ is uniquely determined by its scaling expansion coefficients at scale J (i.e., $\langle f, \psi_{0,J,k}\rangle$) with respect to a M-band orthonormal wavelet basis iff $\hat{\psi}_0(\omega)$ does not vanish on $[-\pi, \pi]$ (or equivalently that $H_0(\omega)$ does not vanish on $[-\pi/M, \pi/M]$. Moreover, in this case, there exists a function $c_{\psi_0}(t)$ such that*

$$f(t) = \sum_k \langle f, \psi_{0,J,k}\rangle \, c_{\psi_0}(t - M^{-J}k). \tag{39}$$

Proof: First notice that

$$\left|\hat{\psi}_0(\omega)\right| = \prod_{j=1}^{\infty} \left|\frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^j}\right)\right|. \tag{40}$$

Therefore, if $\hat{\psi}_0(\omega)$ is non-zero on $[-\pi, \pi]$, in particular the first term $H_0(\frac{\omega}{M})$ is non-zero on $[-\pi, \pi]$. Equivalently, $H_0(\omega)$ is non-zero on $[-\pi/M, \pi/M]$. Conversely, if $H_0(\omega)$ is non-zero on $[-\pi/M, \pi/M]$, $H_0(\omega/M^j)$ is non-zero on $[-\pi, \pi]$ for all $j \geq 1$. Then from Theorem 15.5 in [25] it follows that $\hat{\psi}_0(\omega)$ is non-zero on $[-\pi, \pi]$.

First we show that if $\hat{\psi}_0(\omega_0) = 0$ for $\omega_0 \in [-\pi, \pi]$, then there exist bandlimited functions that cannot be recovered. Take, for instance, a pure tone at $M^J\omega_0$. Then $a(t)$ in Eqn. 20 is zero and hence $\langle f, \psi_{0,J,k}\rangle$ is zero for all k. Therefore, one cannot have any $c_{\psi_0}(t)$ such that Eqn. 39 holds.

Now let $\hat{\psi}_0(\omega)$ be non-zero on $[-\pi, \pi]$. To prove that Eqn. 39 holds the following idea is useful. The Fourier transform of $\langle f, \psi_{0,J,k}\rangle$ considered as an impulse train is the periodization of the Fourier transform of $a(t)$ in Eqn. 20 (i.e periodization of $\hat{f}(\omega)\hat{\psi}_0^*(\frac{\omega}{M^J})$). Therefore, in order to recover $f(t)$ we have to be able to recover $\hat{f}$ from the periodization. So we define,

$$\hat{c}_{\psi_0}(\omega) = \begin{cases} \left[M^{J/2}\hat{\psi}_0^*(\frac{\omega}{M^J})\right]^{-1} & \text{for } \omega \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

This function is well-defined because $\hat{\psi}_0(\omega)$ does not vanish on $[-\pi, \pi]$. Now the Fourier transform of $\sum_k \langle f, \psi_{0,J,k} \rangle \, c_\psi(t - M^{-J}k)$ (because of the bandlimitedness of c_{ψ_0}) is only affected by the first period of the periodization and is given by

$$\left[M^{J/2}\hat{f}(\omega)\hat{\psi}_0^*(\frac{\omega}{M^J})\right]\left[M^{J/2}\hat{\psi}_0^*(\frac{\omega}{M^J})\right]^{-1} = \hat{f}(\omega).$$

□

The theorem states that for a bandlimited signal, knowing Pf, which is *not* bandlimited, is adequate. Notice that if $\psi_0(t)$ is real, then $c_{\psi_0}(t)$ is also real.

3.1.3 Robust Multiresolution Analysis - L^p to L^p

When $f(t)$ is bandlimited, and its Fourier transform is in L^p, we obtain trivially from Eqn. 33 that

$$\left\|\widehat{Qf}\right\|_p^p \leq \left\|\hat{f}(\omega)\right\|_p^p \sup_{\omega\in\Omega} S_p(\omega). \tag{42}$$

Therefore, for the entire class of bandlimited signals with $\left\|\hat{f}(\omega)\right\|_p^p \leq 1$, the worst case L^p approximation error is minimized if $\psi_0(t)$ is such that $\sup_{\omega\in\Omega} S_p(\omega)$ is minimized. In other words, for this class of signals, the *optimal robust* multiresolution analysis is determined by the $\psi_0(t)$ that solves the problem

$$\min_{\Theta}\left[\sup_{\omega\in\Omega} S_p(\omega)\right]. \tag{43}$$

For orthonormal $\psi_0(t)$, and for the L^2 norm, we now show that the optimal robust $\psi_0(t)$ approaches the *sinc* function. This is not surprising since $f(t)$ is bandlimited. Indeed, for $p = 2$ and the wavelet basis orthonormal, if we take $\psi_0(t)$ to be the *sinc* wavelet, we have from Eqn. 35 that $S_2(\omega) = 0$! Therefore, for *sinc* wavelet, the error Qf is always zero.

Eqn. 42 also has the following important consequence. It says that bandlimited signals are *essentially scale-limited* [23].

Definition 1 *A signal $f(t)$ is essentially ϵ scale-limited to scale J, if for all $T \in \mathbb{R}$*

$$\|Qf(t-T)\|_2 \leq \epsilon \|f\|_2 \,. \tag{44}$$

For any given $\psi_0(t)$ if we define $\epsilon = \sup_{\omega\in\Omega} S_2(\omega)$, then we immediately see that bandlimited signals are essentially ϵ-scale-limited.

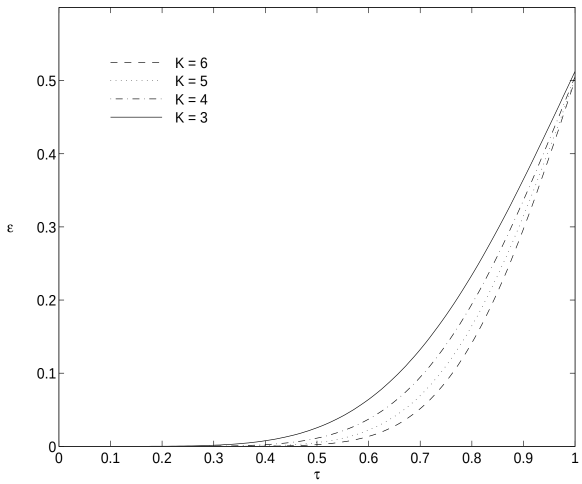
The above definition is meaningful only if ϵ can be made arbitrarily small for a given bandlimited signal and for an appropriate choice of scaling function and scale J. Instead of considering

$f(t)$ bandlimited to Ω and increasing the scale at which the signal is being expanded, we will assume that we are studying a function $f(t) \in W_{0,J}$ and assume that it is bandlimited to $\tau\Omega$, where $0 < \tau \leq 1$. Then, $f(t)$ is essentially ϵ_τ-scale-limited to scale J with

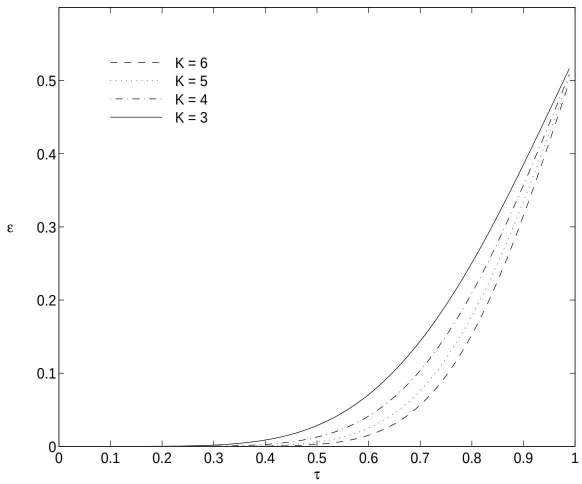
$$\epsilon_\tau = \sup_{\omega \in \tau\Omega} S_2(\omega).$$

For any scaling function $\hat{\psi}_0(0) = 1$ and therefore $S_2(0) = 0$. This shows that $\lim_{\tau \to 0} \epsilon_\tau = 0$ independent of the wavelet basis. However, we are interested in finding a $\psi_0(t)$ for which ϵ_τ decays "fast".

In summary, given a scaling function and an arbitrary ϵ, there always exists a scale such that at most a fraction, ϵ, of the energy of any bandlimited signal (and translates thereof) is above scale J. ϵ_τ as a function of τ for K-regular 2-band [6] and 3-band [27] wavelet bases for different values of K are shown in Fig. 1. For a fixed τ, choosing a more regular (i.e., increasing K) M-band wavelet basis reduces ϵ_τ. Section 4.3 gives designs (Example 5) of the best M-band scaling function for a fixed τ (i.e., the one that minimizes ϵ_τ).



(a)



(b)

Figure 1: ϵ_τ as a function of τ for K-regular (a) 2-band wavelet bases and (b) 3-band wavelet bases.

3.1.4 Robust Multiresolution Analysis - L^p to L^1

Sometimes, in an approximation the maximum error or L^∞ error in the time-domain is important. This error is bounded by the L^1 error in the frequency domain. We now show how optimal robust multiresolution analysis for L^1 error in the frequency domain can be designed. The results are a direct consequence of Hölder's inequality which states that for $f \in L^p(\mathbb{R})$, and $g \in L^q(\mathbb{R})$, where p and q are Hölder conjugates (i.e., $\frac{1}{p} + \frac{1}{q} = 1$) and $1 \leq p, q \leq \infty$,

$$\int_{\mathbb{R}} dt f(t) g(t) \leq \|f\|_p \|g\|_q . \tag{45}$$

Notice that for $p = 2$, Hölder's inequality is equivalent to the Cauchy-Schwartz inequality. From Eqn. 27, for bandlimited $f(t)$ ($\hat{f}(\omega) = 0$ for $\omega \notin \Omega$), the L^1 norm of the error is given by

$$\begin{aligned} \left\| \widehat{Qf} \right\|_1 &= \int_{\mathbb{R}} d\omega \left| \widehat{Qf}(\omega) \right| \\ &= \int_{\Omega} d\omega \left| \hat{f}(\omega) \right| S_p(\omega) \\ &\leq \left\| \hat{f} \right\|_p \|S_p\|_q \qquad \text{From Eqn. 45.} \end{aligned} \tag{46}$$

Therefore, the design of the optimal robust multiresolution analysis for L^p classes of signals, with the L^1 error norm is equivalent to the problem

$$\min_\Theta \|S_p\|_q^q = \min_\Theta \left\{ \int_\Omega d\omega \left(\left| 1 - \left| \hat{\psi}_0 \left(\frac{\omega}{M^J} \right) \right|^2 \right|^p + \left| \hat{\psi}_0^* \left(\frac{\omega}{M^J} \right) \right|^p \sum_{k \neq 0} \left| \hat{\psi}_0 \left(\frac{\omega + 2\pi M^J k}{M^J} \right) \right|^p \right) \right\}^q . \tag{47}$$

4 Numerical Design Examples

This section gives design examples based on the theory derived. We will focus on the L^2 case due to its general importance in engineering and mathematics. We will also solve the more general L^p optimization problem as derived in the paper.

4.1 L^2 designs

This section describes the design of optimal multiresolution analyses for $p = 2$. When $p = 2$, by Parseval's theorem, the frequency domain design technique actually minimizes the energy of the approximation error. From Eqn. 38, the design objective is

$$\max_\Theta \frac{1}{2\pi} \int_\Omega d\omega \left| \hat{f}(\omega) \right|^2 \left| \hat{\psi}_0 \left(\frac{\omega}{M^J} \right) \right|^2 . \tag{48}$$

Since $f(t)$ is bandlimited to Ω, it is given by

$$f(t) = \sum_n f(M^{-J}n) \frac{\sin(2\pi(M^J t - k))}{2\pi(M^J t - k)}. \tag{49}$$

It is also true that for any $L \geq 1$,

$$f(t) = \sum_n f(M^{-(J+L)}n) \frac{\sin(2\pi(M^{J+L} t - k))}{2\pi(M^{J+L} t - k)}. \tag{50}$$

Because of the flatness of $\left|\widehat{\psi}_0(\omega)\right|$ close to the origin, for large enough L,

$$\widehat{\psi}_0\left(\frac{\omega}{M^{l+J}}\right) \approx 1 \text{ on } \Omega \quad \forall\, l \geq L. \tag{51}$$

From Eqn. 12 we now have

$$\left|\widehat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2 \approx \left|\frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^{J+1}}\right)\right|^2 \cdots \left|\frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^{J+L}}\right)\right|^2 \text{ for } \omega \in \Omega \tag{52}$$

and Eqn. 48 can be approximated by

$$\max_{\Theta} \frac{1}{2\pi} \int_{\Omega} d\omega \left|\hat{f}(\omega)\right|^2 \left|\frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^{J+1}}\right)\right|^2 \cdots \left|\frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^{J+L}}\right)\right|^2 \tag{53}$$

or equivalently,

$$\max_{\Theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} d\omega \left|\hat{f}\left(M^{J+L}\omega\right)\right|^2 \left|H_0(M^{L-1}\omega)\right|^2 \cdots \left|H_0(\omega)\right|^2. \tag{54}$$

From Eqn. 54 we can design optimal scaling functions for a given signal. However, we can further simplify the design by realizing that if we define

$$b(n) = f(M^{-(J+L)}n) * f(-M^{-(J+L)}n), \tag{55}$$

then

$$B(\omega) = \left|\hat{f}\left(M^{J+L}\omega\right)\right|^2. \tag{56}$$

Furthermore, let $r(n)$ be the sequence whose (discrete time) Fourier transform is,

$$R(\omega) = \text{DTFT}\{r(n)\} = \left|H_0(M^{L-1}\omega)\right|^2 \cdots \left|H_0(\omega)\right|^2, \tag{57}$$

then the optimal design problem reduces to the following inner product of sequences

$$\max_{\Theta} \sum_n b(n) r(n). \tag{58}$$

Since $b(n)$ and $r(n)$ (being related to autocorrelation of sequences) are symmetric this is equivalent to

$$\max_{\Theta} \left\{\frac{1}{2} b(0) r(0) + \sum_{n=1}^{\infty} b(n) r(n)\right\}. \tag{59}$$

Furthermore, since $h_0(n)$ is a finite length sequence $r(n)$ is finite, and therefore the above sum is also finite. Thus, the objective function is a simple (discrete) inner product between the autocorrelation of the samples of the function and the autocorrelation of the samples of the scaling function.

Having developed the necessary theory for efficient design of an optimal L^2 multiresolution we will now consider four different examples. The first example designs the optimal 2-band, length $N = 8$ scaling vector for a smooth sinusoidal signal. In the second example the desired signal, for which an optimal multiresolution is designed, will be high frequency. To get a better understanding of the optimization the third example will be for a 2-band, length $N = 6$ scaling vector. For the length $N = 6$ scaling vector the parameter space is two dimensional and hence we can visually analyze the error surface. Furthermore, we can now compare the error surface obtained with the approximate method (Eqn. 59) and the exact implementation of the projection error (Eqn. 38). Finally, in the fourth example we design an optimal multiresolution for a 3-band, length $N = 9$ scaling vector for voiced and unvoiced segments of sample speech.

To measure the performance we will list the actual projection errors (i.e., $\|Qf\|$) for both the optimal solution and the corresponding M-band, K-regular solution (which for $M = 2$ reduces to the Daubechies scaling vector).

Example 1 Given the sinusoidal signal sampled at $t = 0, 1, \ldots, 511$

$$f_1(t) = \cos\left(\frac{\pi t}{20}\right) + \frac{2}{5}\cos\left(\frac{\pi t}{5}\right), \tag{60}$$

the corresponding optimal scaling function $\psi_0(t)$ is shown in Fig. 2. The optimal scaling vector,

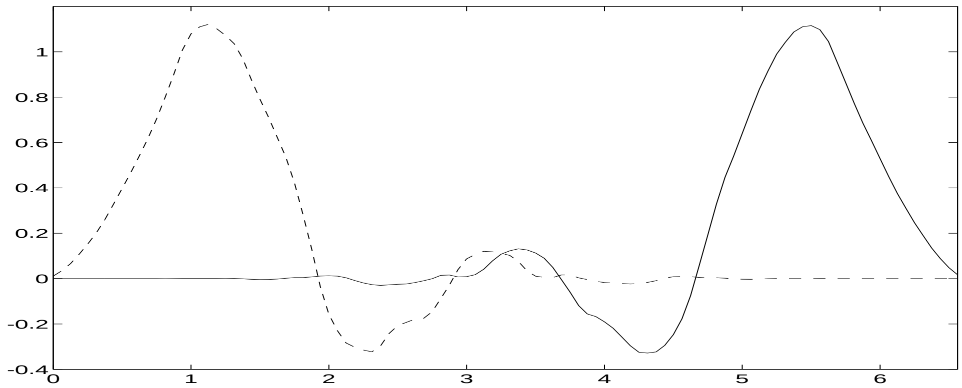


Figure 2: Optimal $\psi_0(t)$ for $M = 2$, $N = 8$. The solid line shows the optimal scaling function and the dashed line shows the corresponding Daubechies 4-regular, 2-band scaling function.

h_0, and the corresponding Daubechies' scaling vector are given in Table 1 with the corresponding projection errors. Hence for this particular signal (smooth at the desired scale) the gain in using

Table 1: Optimal and 4-regular scaling vector, h_0, for $f_1(t)$, $M = 2$ and $N = 8$.

Optimal $h_0(n)$	$\|Qf\|_2$	4-regular $h_0(n)$	$\|Qf\|_2$
-0.01396127556651		0.23037781330890	
0.03980552215623		0.71484657055292	
0.03354698844487		0.63088076792986	
-0.19222802338081	0.09124	-0.02798376941686	0.09258
-0.03914132416347		-0.18703481171909	
0.60466178273853		0.03084138183556	
0.72666239247167		0.03288301166689	
0.25486749967260		-0.01059740178507	

the optimal scaling function is minimal and the relative improvement (normalized by the error of the Daubechies solution) is only 1.5%.

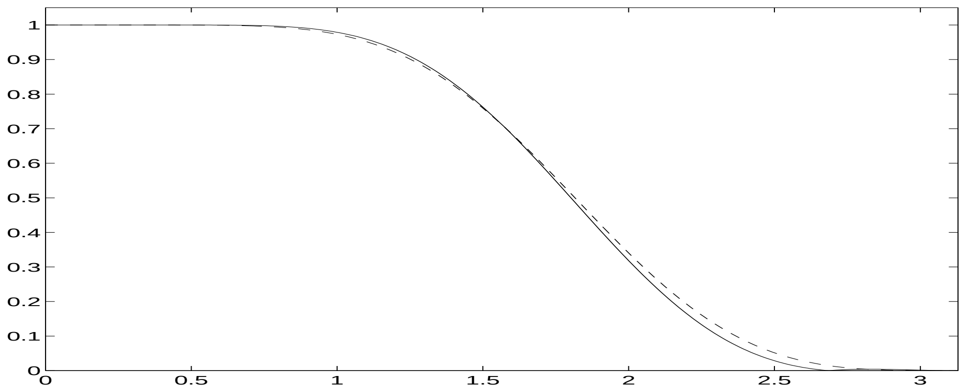
Notice, however, with different values of initial starting points the unconstrained nonlinear numerical optimization scheme yields several solutions, apparently different (we here only refer to distinct solutions with identical error norms and hence exclude possible local minima that one might converge to). However, one easily observe that the scaling vectors corresponding to the various optimal solutions have the same *autocorrelations*, which is what is being optimized (i.e., they are distinct spectral factors). For a more detailed analysis of the optimization problem see Example 3.

To further appreciate the optimal solution it is useful to consider the frequency response of the filters, h_0, (see Fig. 3a) and the Fourier magnitude of the scaling function ψ_0 (see Fig. 3b). From Fig. 3a we observe that the filter response of the half-band optimal solution behaves better in the transition band (i.e., a sharper filter). Similarly from Fig. 3b we also observe that the Fourier magnitude, $\left|\hat{\psi}_0\right|$, for the optimal solution has smaller sidelobes than the 2-band 4-regular Daubechies' scaling function.

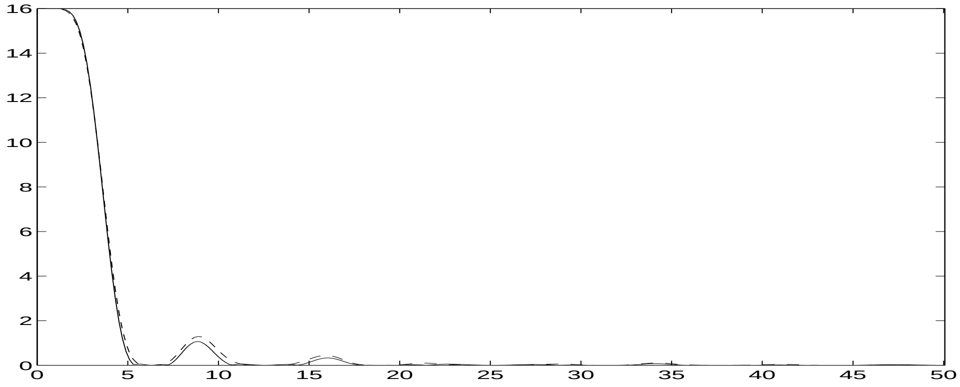
Example 2 Now consider the following high frequency (relative to $f_1(t)$ in the previous example) signal (sampled at $t = 0, 1, \ldots, 511$).

$$f_2(t) = \cos\left(\frac{\pi t}{2}\right) + \frac{2}{5}\cos\left(\frac{\pi t}{4}\right). \tag{61}$$

The scaling function corresponding to the optimal design is plotted in Fig. 4. The optimal scaling vector, h_0, and the corresponding Daubechies scaling vector are given in Table 2 with the corresponding projection errors. For this particular signal (not so smooth at the desired scale), the gain



(a)



(b)

Figure 3: (a) DTFT(h_0) (b) DTFT(ψ_0) In both figures the solid line represents the optimal solution and the dashed line shows the Daubechies 4-regular, 2-band solution.

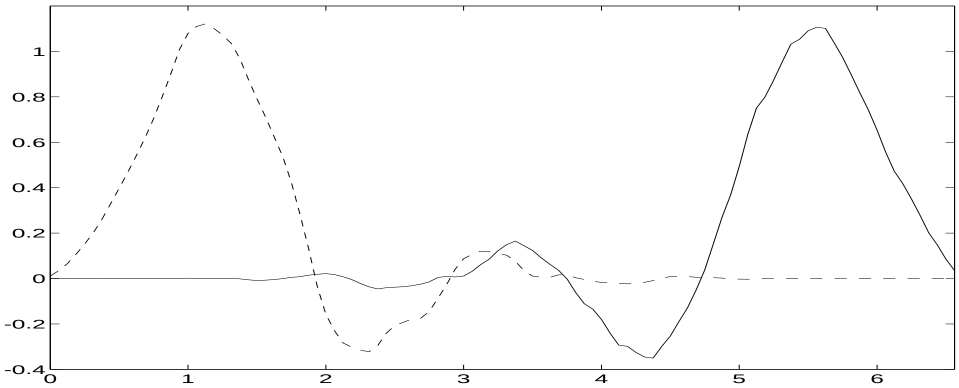


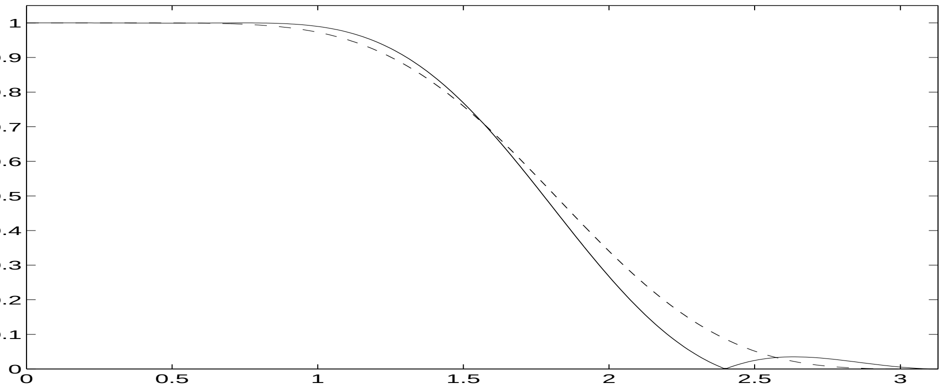
Figure 4: (a) Smooth signal, $f_2(t)$ to approximate. (b) $\psi_0(t)$ with $M = 2$, $N = 8$ (where the solid line shows the optimal design and the dashed line shows the Daubechies 4-regular, 2-band scaling function).

Table 2: Optimal and 4-regular scaling vector, h_0, for $f_2(t)$, $M = 2$ and $N = 8$.

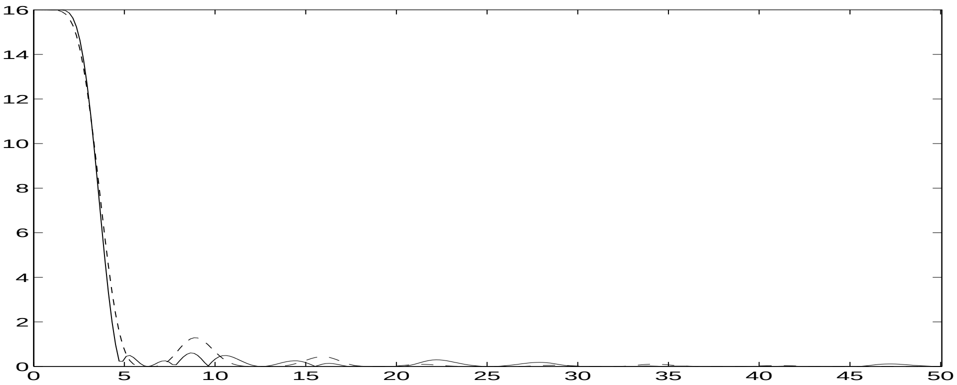
Optimal		4-regular	
$h_0(n)$	$\|Qf\|_2$	$h_0(n)$	$\|Qf\|_2$
-0.02288249926024		0.23037781330890	
0.05580921992427		0.71484657055292	
0.04049121084437		0.63088076792986	
-0.20315246692216	0.18691	-0.02798376941686	0.31341
-0.05596523913400		-0.18703481171909	
0.54880046644809		0.03084138183556	
0.74546330873642		0.03288301166689	
0.30564956173635		-0.01059740178507	

in using the optimal scaling function is substantial and the relative improvement is 40%.

The frequency responses of the scaling vectors and scaling functions corresponding to the optimal solution and the Daubechies' solution are given in Fig. 5a and Fig. 5b. From Fig. 5a we



(a)



(b)

Figure 5: (a) DTFT(h_0) (b) DTFT(ψ_0) In both figures the solid line represents the optimal solution and the dashed line shows the Daubechies 4-regular, 2-band solution.

again observe that the optimal solution result in slightly sharper filters. The tradeoff in this case is large sidelobe in the stop band. Similarly from Fig. 5b we see that the Fourier magnitude of the scaling function, ψ_0, for the optimal solution has smaller sidelobes compare to the Daubechies' solution.

Example 3 Having demonstrated with two examples the optimal design we will now address various issues regarding the numerical aspect of the optimization. This is, as mentioned before, an unconstrained nonlinear optimization. Without going in any depth about the optimization algorithm beyond mentioning that it is based the BFGS update (Broyden, Fletcher, Goldfard and Shanno) [8] and that the code is available if desired (see acknowledgments) we will instead discuss the various solutions as it applies to our problem. Hence, we will in particular consider the error surface for the test signal, $f_2(t)$, from Example 2. To be able to visualize the error surface we will here consider a length $N = 6$, 2-band solution (i.e., a 2-d parameter space). In Fig. 6 we have plotted the negative of the corresponding error surface (hence peaks on the surface corresponds to solutions to our problem). As can be seen from Fig. 6 the error surface is periodic and hence

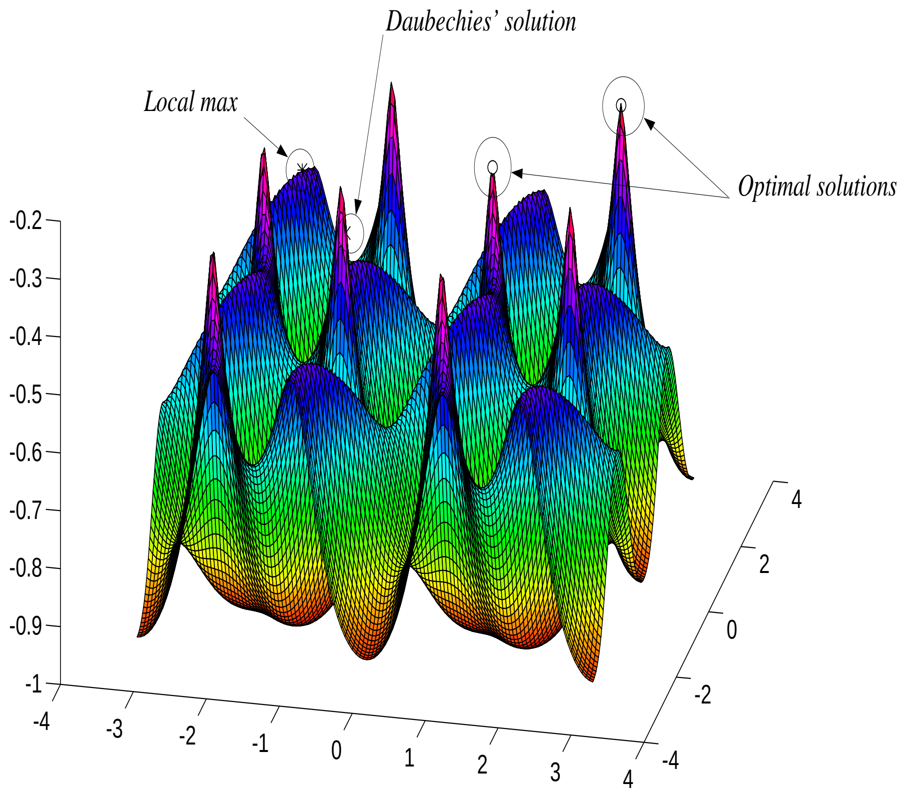


Figure 6: Error surface for $f_2(t)$.

the optimum is global. In Fig. 6, two optimal solutions are marked by $\circ$ and they both yield the same L^2 error but corresponds to either a "time reversed" on not (with the Daubechies solution as a reference). However, depending on the initial starting point in the nonlinear optimization one might converge to a local minimum as illustrated in Fig. 6 by a '*'. For reference we have also marked the Daubechies' solution in Fig. 6 by the symbol $\times$.

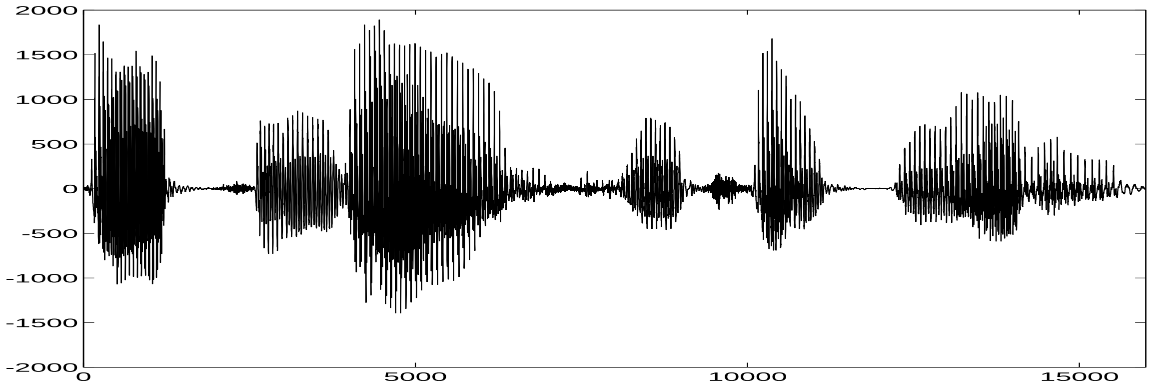
Example 4 In this example consider sampled speech (16000 samples at 8kHz) as shown in Fig 7a. We are in particular interested in two distinct features of the speech signal 1) voiced speech and 2) unvoiced speech. Voiced speech has distinct harmonic components due to vibrations of the vocal cord (e.g., Fig. 7b). Unvoiced speech on the other hand are characterized by the absence of harmonic components and have considerably lower amplitude (e.g., Fig. 7c). In the phrase "It's time" the $/t/$ in "It's" and "time" would be an unvoiced sounds while the $/s/$ in the release of "It's" and $/a^y/$ in "time" would be voiced sounds [24]. Using $M = 3$ we designed the optimal scaling function for representing the two speech segments given in Fig. 7b and Fig. 7c respectively. The optimal scaling functions are shown in Fig. 8 for $J = 0, 1, 2$. The "absolute" approximation error improves slightly as J increases for the voiced segment while for the unvoiced segment the error improves with several orders of magnitude as J increases.

Furthermore, notice that the optimal solution for $J = 1$ and 2 for both the voiced and unvoiced segment (Figs. 8c, d, e and f) are similar (modulo time reversal) to the 3-band, 3-regular scaling function. Intuitively the reason for this is that at these scales both signals can be considered sufficiently smooth and hence the regularity condition will give good results. The optimal solution for $J = 0$ on the other hand (see Figs. 8a and b) behaves differently for voiced and unvoiced. This is due to the fact that the voiced segment is still a smooth signal at this scale while the unvoiced speech segment might not be considered as a smooth signal at this scale.

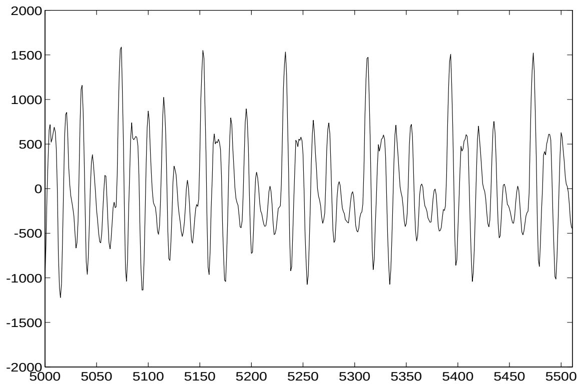
4.2 L^p design

For the general L^p design there is no efficient approximation for computing $\left\|\widehat{Qf}\right\|_p$ and hence we have to implement Eqn. 33. In the following example we will demonstrate one possible implication of the general L^p theory by considering an example for optimizing in the frequency domain over L^1 and evaluate the time domain error as described by Eqn. 29.

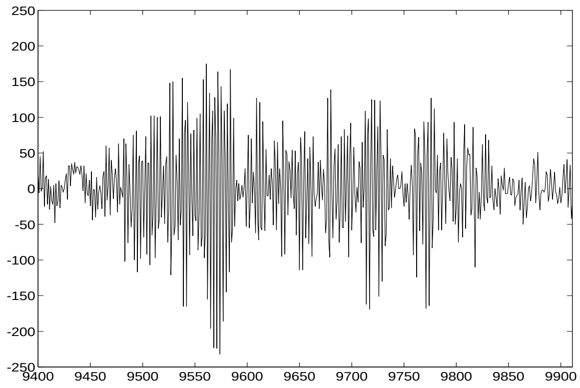
Example 5 In this example we illustrate an L^∞ time domain design by minimizing the L^1 frequency domain approximation error. As a desired signal we use $f_2(t)$ from Example 2. The optimal solution is plotted in Fig. 9. In Fig. 10a we can again observe that the half band filter frequency response has a sharper transition while from Fig. 10b we observe (as in Examples 1 and 2) that the sidelobes



(a)



(b)



(c)

Figure 7: Speech signal (a) "Cats and dogs each hate the other" sampled at $8kHz$. (b) 512 samples of voiced speech and (c) 512 samples of unvoiced speech.

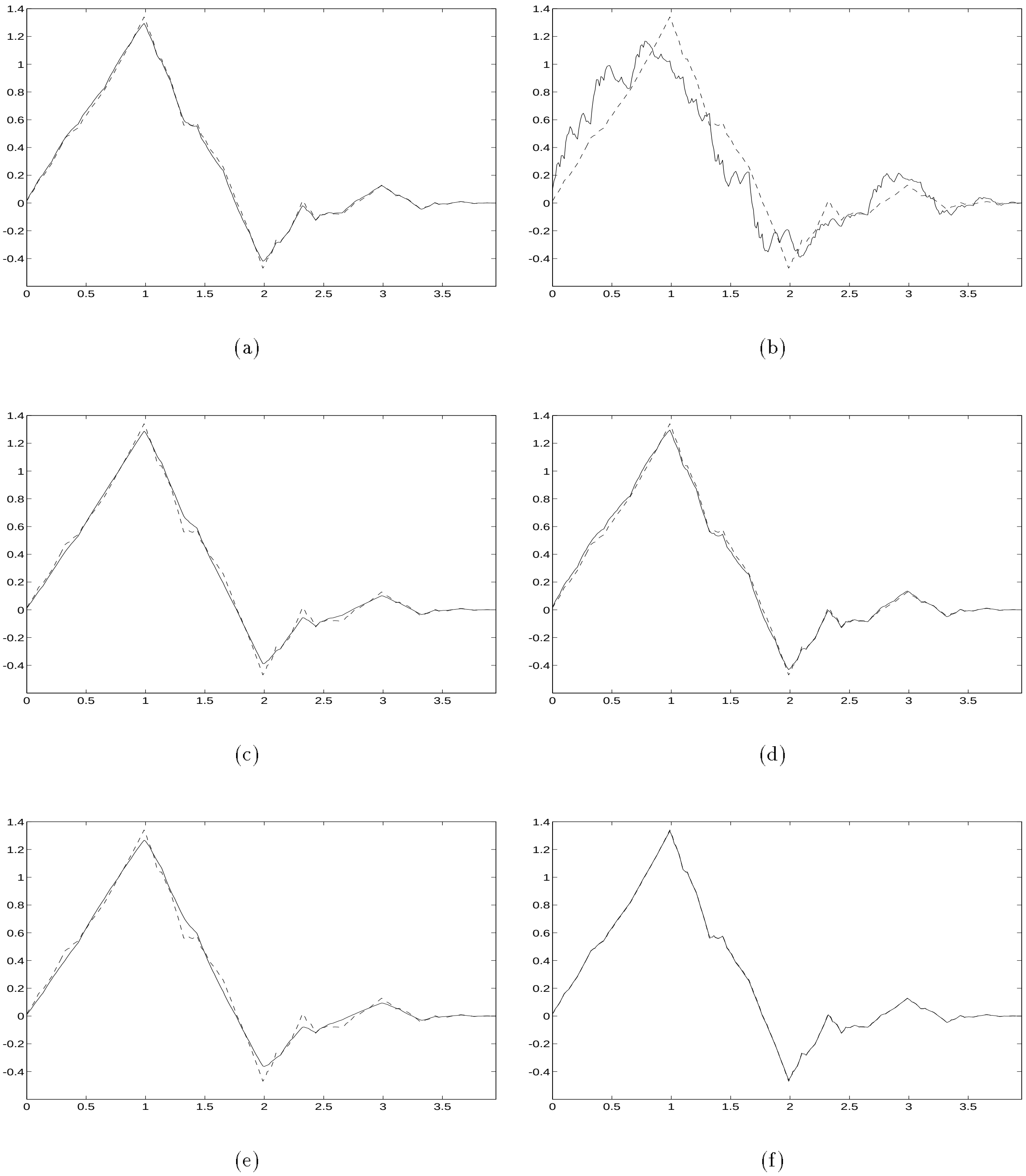


Figure 8: (a), (c) and (e) Optimal scaling functions for voiced speech, (b), (d) and (f) optimal scaling function for unvoiced speech. (a) and (b) corresponds to analysis at scale $J = 0$, (c) and (d) at scale $J = 1$ and, (e) and (f) at scale $J = 2$. The solid line plots the optimal solution and the dashed line plots the 3-band 3-regular solution.

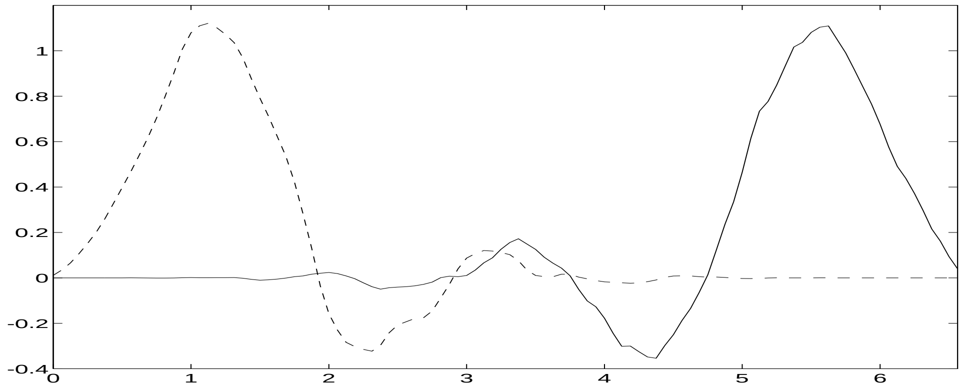


Figure 9: Optimal L^1 $\psi_0(t)$ for $M = 2$, $N = 8$. The solid line shows the optimal design and the dashed line shows the corresponding Daubechies 4-regular, 2-band scaling function).

of the Fourier magnitude, $\left|\hat{\psi}_0\right|$, is less for the optimal solution. Finally, in Fig. 11 we show the reconstruction error for both the 4-regular 2-band scaling function and the L^1 frequency domain optimal scaling function. While the L^1 error improves by about 24%, the time domain L^∞ error improves only by 14%. The fact that the time domain L^∞ error improves less than the L^1 frequency domain error is no surprise (Eqn. 29) since optimizing the L^1 frequency domain error only yields an upper bound for the time domain L^∞ error.

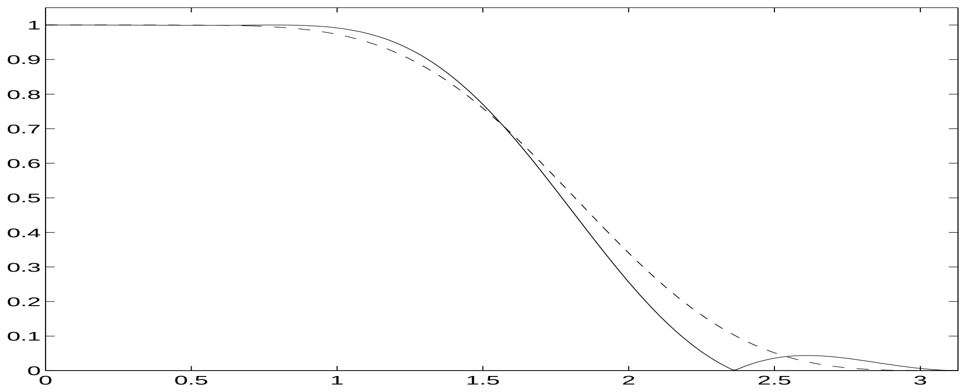
4.3 L^2 robust design

We now design a multiresolution analysis that best represents all finite energy bandlimited signals (i.e., minimizes the maximum L^2 error over this class). The minimization is over all M-band scaling vectors of a given length. We compare our result to the K-regular M-band Daubechies' scaling function.

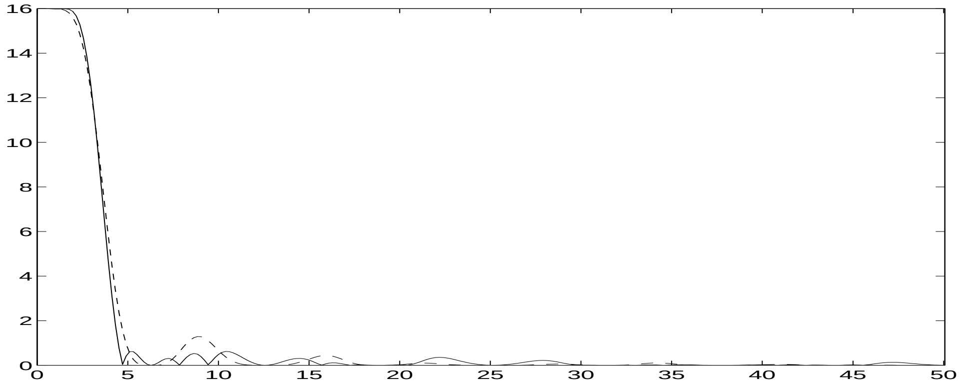
Example 6 For signals in $L^2(\mathbb{R})$, and bandlimited to $\tau\Omega$ ($\tau \in (0, 1]$) we design the scaling function that gives the robust multiresolution analysis. Table 3 compares the worst case L^2 error using our robust designs and the Daubechies' type K-regular, M-band scaling functions. In this example $M = 2$ and $N = 8$ (i.e., corresponds to 4-regular 2-band Daubechies' wavelets). As can be seen from Table 3, for a fixed τ there is a significant improvement in the worst case L^2 error.

5 Conclusion

In this paper we considered the problem of optimal representation of a given signal or class of signals in compactly supported orthonormal wavelet bases. By an optimal representation, we mean

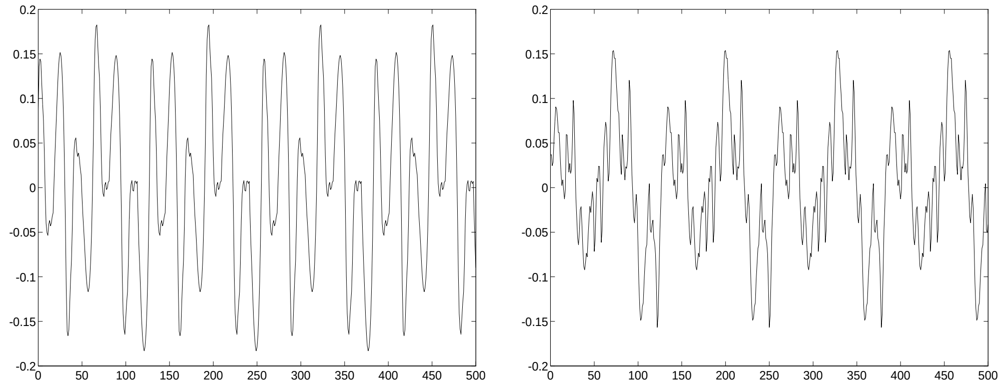


(a)



(b)

Figure 10: (a) DTFT(h_0) (b) DTFT(ψ_0) In both figures the solid line represents the optimal solution and the dashed line shows the Daubechies 4-regular, 2-band solution.



(a) (b)

Figure 11: This shows the difference between the original signal and the signal expanded with the scaling function at some chosen scale J. Only a small part of the time axis is shown for each signal since this is a difference of two periodic signals. (a) Shows the signal difference using a 4-regular 2-band scaling function and (b) used the L^1 frequency domain optimal solution.

Table 3: Comparison of robust L^2 design and 4-regular Daubechies scaling functions.

τ	Optimal ϵ_τ	4-regular ϵ_τ	% improvement
0.1	4.3327e-11	1.1739e-09	96.3
0.2	7.7292e-08	1.7088e-06	95.5
0.3	3.6115e-06	6.7238e-05	94.6
0.4	1.0688e-04	1.4394e-03	92.6
0.5	7.2809e-04	7.1475e-03	89.8
0.6	3.7668e-03	2.4575e-02	84.7
0.7	2.6146e-02	8.5206e-02	69.3
0.8	8.9117e-02	1.7219e-01	48.2
0.9	2.2951e-01	2.9766e-01	22.9

the best multiresolution analysis that describes the signal at a pre-determined scale J. We obtained explicit formulae for the p^{th} norm of the approximation error in the Fourier transform domain, in terms of the Fourier transform of the scaling function associated with the multiresolution analysis. In general the error can only be modeled as the output of a linear operator driven by the given signal. In the special case of bandlimited signals, it turns out that this *error operator* is linear and translation-invariant. Thus the error norm can easily be computed for a given signal and scaling function. By using this fact, and the explicit parameterization of all M-band, wavelet tight frames [15], it is possible to numerically design the "optimal" scaling function for the representation. Again, exploiting the fact that the error is the output of an error operator, we have derived necessary objective functions for the design of a "robust" multiresolution analysis by minimizing the induced error operator norm for $L^p(\mathbb{R})$ classes of signals. Numerical examples for design of optimal wavelets in the least-squared sense and the Chebychev error sense, and robust wavelets in the least-square error sense are given.

The optimal design we have developed will satisfy the regularity condition or equivalently the vanishing moment conditions set forth by Daubechies in her original paper [6]. However, recently there has been some work on characterizing Sobolev type smoothness of the scaling function [18, 10]. The tools developed in this paper can immediately be applied to the design of optimally regular functions in a Sobolev sense.

Acknowledgment

We would like to thank members of the DSP group at Rice University for their suggestions and comments. In particular Dr. Lang contributed greatly to enhance the readability of the paper. We would also like like to thank Dr. Dennis at Rice University and Dr. Schnabel at University of Colorado for providing us with their nonlinear optimization code [8]. The computer programs used in this paper is available in a Matlab toolbox, "rice-wlet-tools," on anonymous ftp from cml.rice.edu in the directory /pub/software.

References

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using the wavelet transform. *IEEE Trans. on Image Processing*, 2(2), April 1992.

[2] A. Barbé. A level-crossing based scaling-dimensionality transform applied to stationary gaussian processes. *IEEE Trans. Inform. Theory*, 38(2):814–823, March 1992.

[3] M. Basseville, A. Bienveniste, and K. C. Chou. Modeling and estimation of multiresolution stochastic process. *IEEE Trans. Inform. Theory*, 38(2):766–784, March 1992.

[4] A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Applied Math.*, 1992.

[5] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory*, 38(2):1713–1716, 1992.

[6] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Applied Math.*, XLI(41):909–996, November 1988.

[7] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.

[8] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1st edition, 1983.

[9] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Inform. Theory*, 38(2):719–746, March 1992.

[10] T. Eirola. Sobolev characterization of solutions of dilation equations. *SIAM J. of Math.*, 23(4):1015–1030, July 1992.

[11] R. A. Gopinath. *Wavelet and Filter Banks - New Results and Applications.* PhD thesis, Rice University, Houston, TX-77251, August 1992.

[12] R. A. Gopinath and C. S. Burrus. State-space approach to multiplicity M orthonormal wavelet bases. Technical Report CML TR91-22, Computational Mathematics Laboratory, Rice University, 1991. Submitted to IEEE Trans. on SP.

[13] R. A. Gopinath and C. S. Burrus. On the moments of the scaling function ψ_0. In *Proc. of the IEEE ISCAS*, volume 2, pages 963–966, San Diego, CA, May 1992. Also Tech. Report CML TR91-05.

[14] R. A. Gopinath and C. S. Burrus. Wavelet-based lowpass/bandpass interpolation. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, volume 4, pages IV385–IV388, San Francisco, CA, March 1992. IEEE. Also Tech. Report CML TR91-06.

[15] R. A. Gopinath and C. S. Burrus. Wavelets and filter banks. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 603–654. Academic Press, San Diego, CA, 1992. Also Tech. Report CML TR91-20, September 1991.

[16] P. Heller. Regular M-band wavelets. Technical Report AD920608, Aware Inc., Cambridge, MA 02142, 1992.

[17] P. Heller, H. W. Resnikoff, and R. O. Wells, Jr. Wavelet matrices and representation of discrete functions. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 15–51. Academic Press, 1992.

[18] P. Heller and R. O. Wells, Jr. Spectral theory of multiresolution operators and applications. In *Proceedings of the International Conference on Wavelets*, Taormina, Italy, 1993. Also Rice University Tech. Report CML TR93-09.

[19] F. Jones. *Lebesgue Integration on Euclidean Space.* Jones and Bartlett, Boston, MA, 1993.

[20] W. M. Lawton. Necessary and sufficient conditions for constructing orthonormal wavelet bases. *Journal of Math. Physics*, 32(1):57–61, January 1991. Also Aware, Inc. Tech. Report AD900402, April 1990.

[21] W. R. Madych. Some elementary properties of multiresolution analyses of $\ell^2(\mathbb{R}^n)$. In C. K. Chui, editor, *Wavelets: A Tutorial in Theory and Applications*, pages 259–294. Academic Press, 1992.

[22] P. Moulin, J. A. O'Sullivan, and D. L. Snyder. A method of sieves for multiresolution spectrum estimation and radar imaging. *IEEE Trans. Inform. Theory*, 38(2):801–813, March 1992.

[23] J. E. Odegard, R. A. Gopinath, and C. S. Burrus. Optimal wavelets for signal decomposition and the existence of scale limited signals. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, volume 4, pages IV 597–600, San Francisco, CA, March 1992. IEEE. Also Tech. Report CML TR91-07.

[24] L. Rabiner and B-H. Juang. *Fundamentals of Speech Recognition.* Prentice-Hall, 1993.

[25] W. Rudin. *Real and Complex Analysis.* McGraw-Hill., 3 edition, 1987.

[26] P. Steffen. Closed form derivation of orthogonal wavelet bases for arbitrary integer scales. Technical report, University of Erlangen-Nürnberg, Germany, 1992.

[27] P. Steffen, P. Heller, R. A. Gopinath, and C. S. Burrus. Theory of regular M-band wavelet bases. *IEEE Trans. SP*, 41(12), December 1993. Special Transaction issue on wavelets; Rice contribution also in Tech. Report No. CML TR-91-22, November 1991.

[28] E. M. Stein and G. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press, Princeton, NJ, 1975.

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

[30] M. Vetterli. A theory of multirate filter banks. *IEEE Trans. ASSP*, 35(3):356–372, March 1987.

[31] G. W. Wornell and A. V. Oppenheim. Wavelet-based representations for a class of self-similar signals with application to fractal modulation. *IEEE Trans. Inform. Theory*, 38(2):785–800, March 1992.

[32] H. Zou and A. H. Tewfik. Discrete orthogonal M-band wavelet decompositions. In *Proc. Int. Conf. Acoust., Speech, Signal Processing*, volume 4, pages IV–605–IV–608, San Francisco, CA, 1992. IEEE.