### The Projection onto the Finest Scale

One efficiently computes the DWT from the DST at scale $J_f$: $Wf(0, J_f, k)$. The DST is seldom computed exactly since inner products are involved. Moreover typically only the samples of $f(t)$ are available. From the samples by using local polynomial interpolation one can approximate $f(t)$ and then compute the DST by numerical quadrature. This procedure becomes the discrete convolution of the samples of $f(t)$ with the moments of $\psi_0(t)$ [33, 32]. Now the moments of $\psi_0(t)$ can be computed from the moments of the scaling vector. In fact if $\mu_{i,k}$ and $m_{i,k}$ denote the the $k^{th}$ moments of $h_i(n)$ and $\psi_i(t)$ respectively one can show that (see Section 5.3, Lemma 31 for details)

$$m_{i,k} = \frac{1}{M^{k+\frac{1}{2}}} \sum_{j=0}^{k} \binom{k}{j} \mu_{i,j} m_{0,k-j}.$$

Hence $m_{i,k}$ can be computed from $\mu_{i,k}$. In general with local polynomial approximation one can compute a sequence $e(k)$ so that $W(0, J_f, k) = e(k) * f(M^{-J_f} k)$.

In most applications one can take the samples of $f(t)$ to be the DST. In fact if the scaling function is $K$-regular with $K \geq 2$ one can show that (see Section 5.3) $f(M^{-J_f}(k + m_{0,1}))$ gives a third order approximation to $Wf(0, J_f, k)$.

### Analysis

From the DST at scale $j$ one can compute the DWT and DST coefficients at scale $j-1$ using a filter bank. Indeed from Eqn. 4.74 we get

$$Wf(i, j-1, n) = \sum_k h_i(k) Wf(0, j, Mn + k).$$

This corresponds to an analysis filter bank with filters $h_i(-n)$.

**Synthesis**

From the DST and DWT coefficients at scale $j-1$ the DST coefficients at scale $j$ can be computed using a synthesis filter bank with filters given by $g_i(-n)$.

$$Wf(0,j,n) = \sum_{i=0}^{M-1}\sum_k\sum_k g_i(Mk-n)Wf(i,j-1,k).$$

This follows from the fact that

$$\begin{aligned}
\sum_{i=0}^{M-1}\sum_k g_i(Mk-n)\psi_{i,j-1,k} &= \sum_{i=0}^{M-1}\sum_k g_i(Mk-n)\sum_l h_i(l)\psi_{0,j,Mk+l} \\
&= \sum_{i=0}^{M-1}\sum_k g_i(Mk-n)\sum_l h_i(-Mk+l)\psi_{0,j,l} \\
&= \sum_l \psi_{0,j,l}\left[\sum_{i=0}^{M-1}\sum_k g_i(Mk-n)h_i(-Mk+l)\right] \\
&= \sum_l \psi_{0,j,l}\delta(l-n) = \psi_{0,j,n}.
\end{aligned}$$

## 5.3 Moments of the Scaling Function and Wavelets

The moments of the scaling function and wavelets can be computed exactly from the moments of the scaling and wavelet vectors. For $K$-regular, multiplicity $M$ WTFs the moments of $h_0$ satisfy a set of *structural relationships* that imply a set of relationships between the moments of $\psi_0(t)$. One such relationship is that $m_{0,1}^2 = m_{0,2}$. This result implies that uniform samples of of a smooth function give a third order approximation to to the DST coefficients.

### 5.3.1 The Moments of $\psi_i(t)$ and $h_i(n)$

For $i \in \mathcal{R}(M)$ and $n \in \mathbb{N}$ let

$$m_{i,n} = \int_{\mathbf{R}} dt\, t^n \psi_l(t), \quad \text{and} \quad \mu_{i,n} = \sum_{k=0}^{N-1} k^n h_i(k). \tag{5.12}$$

**Lemma 31** The moments $\psi_i(t)$ and $h_i(n)$ are related as follows:

$$m_{i,n} = \frac{1}{M^{n+\frac{1}{2}}} \sum_{j=0}^{n} \binom{n}{j} \mu_{i,j} m_{0,n-j}. \tag{5.13}$$

**Proof:** From Eqn. 4.18

$$\begin{aligned}
m_{i,n} &= \sqrt{M} \sum_k h_i(k) \int_{\mathbf{R}} dt\, t^n \psi_0(Mt-k) \\
&= \frac{1}{M^{\frac{1}{2}}} \sum_k h_0(k) \int_{\mathbf{R}} dt \left(\frac{t+k}{M}\right)^n \psi_0(t)\, dt \\
&= \frac{1}{M^{n+\frac{1}{2}}} \sum_k h_i(k) \int_{\mathbf{R}} dt \sum_{j=0}^{n} \binom{n}{j} t^{n-j} k^j \psi_0(t) \\
&= \frac{1}{M^{n+\frac{1}{2}}} \sum_{j=0}^{n} \binom{n}{j} \left[\sum_k h_i(k) k^j\right] m_{0,n-j} \\
&= \frac{1}{M^{n+\frac{1}{2}}} \sum_{j=0}^{n} \binom{n}{j} \mu_{i,j} m_{0,n-j}
\end{aligned}$$

□

Eqn. 5.13 gives a recursive formula to compute $m_{i,k}$. Since $\mu_{i,0} = \sum_n h_i(n) = \sqrt{M}\delta(i)$ and $m_{0,0} = 1$,

$$m_{i,0} = \frac{1}{\sqrt{M}} \mu_{i,0} m_{0,0} = \frac{1}{\sqrt{M}} \sqrt{M}\delta(i) = \delta(i).$$

If we define the *scaled discrete moments* $d_{i,n} = \mu_{i,n}/\sqrt{M}$, Eqn. 5.13 becomes

$$m_{i,n} = \frac{1}{M^n} \sum_{j=0}^{n} \binom{n}{j} d_{i,j} m_{0,n-j}, \tag{5.14}$$

with $d_{i,0} = m_{i,0} = \delta(i)$.

**Lemma 32** Given $\psi_0(t)$ and an integer $k \geq 0$ the following statements are equivalent:

1. For all $n$, $0 \leq n \leq k$, $m_{0,n} = (m_{0,1})^n$.

2. For all $n$, $0 \le n \le k$, $d_{0,n} = (d_{0,1})^n$.

If either condition is satisfied $d_{0,n} = (d_{0,1})^n = (M-1)^n (m_{0,1})^n$ for all non-negative $n$.

**Proof:** $d_{0,0} = m_{0,0}$. From Eqn. 5.14 $Mm_{0,1} = d_{0,0}m_{0,1} + d_{0,1}m_{0,0}$ and hence $d_{0,1} = (M-1)m_{0,1}$. For $0 \le n \le k$, let $d_{0,n} = (d_{0,1})^n$. By the induction hypothesis $m_{0,n} = (m_{0,1})^n$ and $d_{0,n} = (M-1)^n (m_{0,1})^n$. Now invoking Eqn. 5.14 for $l = 0$ and $n+1$, and using the fact that $d_{0,n+1} = (d_{0,1})^{n+1}$ we get

$$\begin{aligned} M^{n+1} m_{0,n+1} &= \sum_{i=0}^{n+1} \binom{n+1}{i} d_{0,i} m_{0,n+1-i} \\ &= \sum_{i=0}^{n+1} \binom{n+1}{i} (d_{0,1})^i (m_{0,1})^{n+1-i} + m_{0,n+1} - m_{0,1}^{n+1} \\ &= (d_{0,1} + m_{0,1})^{n+1} + m_{0,n+1} - m_{0,1}^{n+1} \\ &= M^{n+1} (m_{0,1})^{n+1} + m_{0,n+1} - (m_{0,1})^{n+1} \end{aligned}$$

and hence the result follows. The converse also follows similarly. In particular, $m_{0,2} = m_{0,1}^2$ if and only if $d_{0,2} = d_{0,1}^2$. □

Under the conditions above the first $k$ moments of $\breve{\psi}_0(t) = \psi_0(t + m_{0,1})$ are zero. Indeed if $\breve{m}_{0,n}$ is the $n^{th}$ moment of $\breve{\psi}_0(t)$, then

$$\breve{m}_{0,n} = \int_{\mathbf{R}} dt\, t^n \psi_0(t + m_{0,1}) = \int_{\mathbf{R}} dt\, (t - m_{0,1})^n \psi_0(t) = \sum_{i=0}^{n} \binom{n}{i} (-1)^i m_{0,i} m_{0,1}^{(n-i)}. \tag{5.15}$$

From the above equation we have the following result:

**Lemma 33** For $1 \le n \le k$, let $d_{0,n} = d_{0,1}^n$. Then $m_{0,n} = 0$.

**Proof:** From Lemma 32 for $n \le k$, $m_{0,n} = m_{0,1}^n$. Now from Eqn. 5.15,

$$\breve{m}_{0,n} = \sum_{i=0}^{n} \binom{n}{i} (-1)^i m_{0,1}^i m_{0,1}^{(n-i)} = m_{0,1}^n \sum_{i=0}^{n} \binom{n}{i} (-1)^i = 0.$$

□

### 5.3.2 The Fourier Transform and Discrete Moments

The moments of a sequence are related to the behavior of its Fourier transform in a neighborhood of $\omega = 0$. For $K$-regular WTFs $H_0(\omega)$ behaves like $\sqrt{M} + O(|\omega|^{2K})$ for small $\omega$. From this fact one can infer a set of relationships between $\mu_{0,n}$. Since

$$|H_0(\omega)|^2 = \sum_{k,l} h_0(k)h_0(l)e^{\imath(k-l)\omega},$$

we get

$$\left[\left(\frac{d}{d\omega}\right)^n |H_0(\omega)|^2\right]_{\omega=0} \stackrel{\text{def}}{=} a(n) = \imath^n \sum_{k,l} h_0(k)h_0(l)(k-l)^n.$$

For odd $n$ from symmetry it is clear that the right hand side evaluates to zero. Therefore all odd derivatives of $|H_0(\omega)|^2$ are zero. The even derivatives are related to the the discrete moments of $h_0$. Indeed for $n = 2p$

$$\begin{aligned} a(2p) &= \imath^{2p} \sum_{k,l} h_0(k)h_0(l) \sum_{j=0}^{2p} \binom{2p}{j} k^{(2p-j)}(-l)^j \\ &= (-1)^p \sum_{j=0}^{2p} \binom{2p}{j} \left(\sum_k h_0(k)k^{2p-j}\right)\left(\sum_l h_0(l)(-l)^j\right) \\ &= (-1)^p \sum_{j=0}^{2p} \binom{2p}{j} (-1)^j \mu_{0,2p-j}\mu_{0,j}. \end{aligned} \quad (5.16)$$

**Lemma 34** For a multiplicity $M$, $K$-regular WTF

$$|H_0(\omega)|^2 = M + O(|\omega|^{2K}) \quad (5.17)$$

**Proof:** For $K$-regularity (from Eqn. 4.48)

$$H_0(\omega) = e^{-\imath(M-1)K\omega/2}\left(\frac{\sin(M\omega/2)}{\sin(\omega/2)}\right)^K R(\omega) \quad (5.18)$$

Since $R(\omega)$ cannot have a factor $\left(\frac{1+e^{-\imath\omega}+\ldots+e^{-\imath(M-1)\omega}}{M}\right)^K$ (otherwise $h_0$ would be $K+1$ regular), for some $k \in \{1,\ldots,M-1\}$ $R\left(\frac{2\pi k}{M}\right) \neq 0$. But

$$|H_0(\omega)|^2 = \left(\frac{\sin(M\omega/2)}{M\sin(\omega/2)}\right)^{2K} |R(\omega)|^2 .$$

and therefore

$$\begin{aligned}\sum_{k=1}^{M-1} \left|H_0(\frac{\omega+2\pi k}{M})\right|^2 &= \sum_{k=1}^{M-1} \left(\frac{\sin(M\omega/2)}{M\sin(\frac{M\omega+2\pi k}{2M})}\right)^{2K} \left|R(\omega+\frac{2\pi k}{M})\right|^2 \\ &= \frac{\sin^{2K}(M\omega/2)}{M^{2K}} \sum_{k=1}^{M-1} \frac{|R(\omega+2\pi k/M)|^2}{\sin^{2K}(\frac{M\omega+2\pi k}{2M})}.\end{aligned}$$

For $k \in \{1,\ldots,M-1\}$, $\sin(\frac{\pi k}{M})$ is not zero. There exists a compact neighborhood of $\omega = 0$ on which $\sin(\frac{M\omega+2\pi k}{2M})$ is bounded away from zero. For sufficiently small $\epsilon$, therefore, there exists a constant $C$ such that for all $|\omega| < \epsilon$

$$\sum_{k=1}^{M-1} \left|H_0\left(\frac{\omega+2\pi k}{M}\right)\right|^2 = \left(\frac{\sin(M\omega/2)}{M}\right)^{2K} [C + O(|\omega|)]$$

or equivalently

$$\sum_{k=1}^{M-1} \left|H_0\left(\frac{\omega+2\pi k}{M}\right)\right|^2 = O(|\omega|^{2K}) \tag{5.19}$$

Now from the transmultiplexer PR property (Eqn. 3.20) we get

$$|H_0(\omega)|^2 = M - \sum_{k=1}^{M-1} \left|H_0\left(\frac{\omega+2\pi k}{M}\right)\right|^2. \tag{5.20}$$

The result follows from Eqn. 5.19 and Eqn. 5.20. □

For $K \geq 1$ and $p \in \{0,1,2,\ldots,K-1\}$

$$\left[\left(\frac{d}{d\omega}\right)^{2p} |H_0(\omega)|^2\right]_{\omega=0} = M\delta(p)$$

This is a set of $K$ equations relating the first $2K-1$ moments of $h_0$. This information is not sufficient to know all of the first $2K-1$ moments. For $K \geq 2$, the maximum value of $p$ is greater than or equal to 1. When $p = 1$, $2\mu_{0,2}\mu_{0,0} - 2\mu_{0,1}\mu_{0,1} = 0$ and hence $\mu_{0,2} = \mu_{0,1}^2/\sqrt{M}$ and $d_{0,2}^2 = d_{0,1}^2$.

**Theorem 37** For compactly supported, multiplicity $M$, $K$-regular, WTFs with $K \geq 2$ (i.e., except for the Haar case), the moments of the scaling function satisfy $m_{0,2} = (m_{0,1})^2$.

Tables 5.1-5.3 give the moments of the scaling functions and scaling vectors of $K$-regular, minimal length, multiplicity $M$ orthonormal wavelet bases.

Table 5.1: The Moments of $\psi_0(t)$ : $M = 2$

$M = 2$ and $N = MK$

| $N$ | $k$ | $m_{0,k}$ | $d_{0,k}$ |
|---|---|---|---|
| 4 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | →1 | 6.3397460e-01 | 6.3397460e-01 |
| | →2 | 4.0192379e-01 | 4.0192379e-01 |
| | 3 | 1.3109156e-01 | -6.1121593e-01 |
| | 4 | -3.0219333e-01 | -4.2846097e+00 |
| | 5 | -1.0658728e+00 | -1.6572740e+01 |
| 6 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | →1 | 8.1740117e-01 | 8.1740117e-01 |
| | →2 | 6.6814467e-01 | 6.6814467e-01 |
| | 3 | 4.4546004e-01 | -1.5863308e-01 |
| | 4 | 1.1722635e-01 | -1.8579194e+00 |
| | 5 | -4.6651091e-02 | 3.7516197e+00 |
| 8 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | →1 | 1.0053932e+00 | 1.0053932e+00 |
| | →2 | 1.0108155e+00 | 1.0108155e+00 |
| | 3 | 9.0736037e-01 | 2.5392023e-01 |
| | 4 | 5.8377181e-01 | -2.0440853e+00 |
| | 5 | 6.3077524e-02 | -2.4420547e+00 |
| 10 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | →1 | 1.1939080e+00 | 1.1939080e+00 |
| | →2 | 1.4254164e+00 | 1.4254164e+00 |
| | 3 | 1.5802598e+00 | 8.5092254e-01 |
| | 4 | 1.4513041e+00 | -2.0317424e+00 |
| | 5 | 8.1371053e-01 | -5.9644946e+00 |

Table 5.2: The Moments of $\psi_0(t)$ : $M = 3$

$M = 3$ and $N = MK$

| $N$ | $k$ | $m_{0,k}$ | $d_{0,k}$ |
|---|---|---|---|
| 6 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | $\rightarrow$ 1 | 6.2084713e-01 | 1.2416943e+00 |
| | $\rightarrow$ 2 | 3.8545116e-01 | 1.5418046e+00 |
| | 3 | 1.1024925e-01 | -1.4410320e+00 |
| | 4 | -3.3859274e-01 | -2.7622103e+01 |
| 9 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | $\rightarrow$ 1 | 7.8515128e-01 | 1.5703026e+00 |
| | $\rightarrow$ 2 | 6.1646253e-01 | 2.4658501e+00 |
| | 3 | 3.8154196e-01 | 1.2077966e+00 |
| | 4 | 5.8194455e-02 | -1.0654826e+01 |
| 12 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | $\rightarrow$ 1 | 9.5286399e-01 | 1.9057280e+00 |
| | $\rightarrow$2 | 9.0794979e-01 | 3.6317991e+00 |
| | 3 | 7.5580853e-01 | 4.0782740e+00 |
| | 4 | 4.0761249e-01 | -8.4815717e+00 |

Table 5.3: The Moments of $\psi_0(t)$ : $M = 5$

$M = 5$ and $N = MK$

| $N$ | $k$ | $m_{0,k}$ | $d_{0,k}$ |
|---|---|---|---|
| 10 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | $\rightarrow$1 | 6.0961180e-01 | 2.4384472e+00 |
| | $\rightarrow$2 | 3.7162654e-01 | 5.9460247e+00 |
| | 3 | 9.3544517e-02 | -1.9933553e+00 |
| | 4 | -3.6313857e-01 | -2.3590840e+02 |
| 15 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | $\rightarrow$1 | 7.5803488e-01 | 3.0321395e+00 |
| | $\rightarrow$2 | 5.7461687e-01 | 9.1938700e+00 |
| | 3 | 3.3138863e-01 | 1.4957413e+01 |
| | 4 | 1.4262918e-02 | -7.2169885e+01 |
| 20 | 0 | 1.0000000e+00 | 1.0000000e+00 |
| | $\rightarrow$1 | 9.0920717e-01 | 3.6368287e+00 |
| | $\rightarrow$2 | 8.2665767e-01 | 1.3226523e+01 |
| | 3 | 6.4125671e-01 | 3.4419647e+01 |
| | 4 | 2.8205206e-01 | -2.4109276e+01 |

### 5.3.3 Sample Approximation of $Wf(0, J_f, k)$

For a compactly supported WTF $\psi_{0,J,k}(t)$ is concentrated around $M^{-J}k$. In a neighborhood of this point a function $f(t)$ may be approximated by a Taylor series for the computation of $\langle \psi_{0,J,k}, f\rangle$. Since $\psi_0$ is supported in $[0, \frac{N-1}{M-1}]$, $\psi_{0,J,k}$ is supported in $[M^{-J}k, M^{-J}(k + \frac{N-1}{M-1})]$. Consider the Taylor series expansion of $f(t)$ around the first moment of $\psi_{0,J,k}$:

$$f(M^{-J}(k+t)) = f\left(\frac{k+m_{0,1}}{M^J}\right) + \left(\frac{t-m_{0,1}}{M^J}\right) f^{(1)}\left(\frac{k+m_{0,1}}{M^J}\right) + \ldots$$

If $h_0(n)$ is $K$-regular, $K \geq 2$, then from Theorem 37 $m_{0,2} = m_{0,1}^2$ and hence

$$\begin{aligned}
\langle \psi_{0,J,k}, f\rangle &= \int_{\mathbf{R}} dt f(t) M^{J/2}\psi_0(M^J t - k) \\
&= M^{-J/2}\left\{\int_R dt f\left(\frac{t+k}{M^J}\right)\psi_0(t)\, dt\right\} \\
&= M^{-J/2}\left\{f\left(\frac{k+m_{0,1}}{M^J}\right) + O\left((1/M^J)^3\right).\right\}
\end{aligned}$$

The last step is obtained by invoking the Taylor series expansion and using the relationships between the moments. Hence the samples $f(M^{-J}(k+m_{0,1}))$ (appropriately scaled) themselves give a third order approximation to the scaling expansion coefficients. Increasing the sampling rate by a factor of $M$ reduces the error by a factor of $M^3$.

Consider an application in which one chooses $Wf(0, J, k) = f(M^{-J}k)$ (this is what is usually done in practice). From this we can compute $\breve{f}(t) = \sum_k f(M^{-J}k)\psi_{0,J,k}(t)$, which is an approximation to $f(t)$. Fig. 5.10 shows an example function $f(t)$ and the corresponding reconstructed function $\breve{f}(t)$. In this example a multiplicity 2, 4-regular, length $N = 8$ WTF is used and the approximation is done at scale $J = 4$. One notices that $f(t)$ and $\breve{f}(t)$ are time-shifts of each other This time-shift is roughly $M^{-J}m_{0,1}$. This phenomenon occurs because the samples $f(M^{-J}k)$ give a third order approximation to the DST coefficients of the function $f(t + M^{-J}m_{0,1})$ and hence the reconstructed function $\breve{f}(t)$ is a very good approximation to $f(t + M^{-J}m_{0,1})$. An

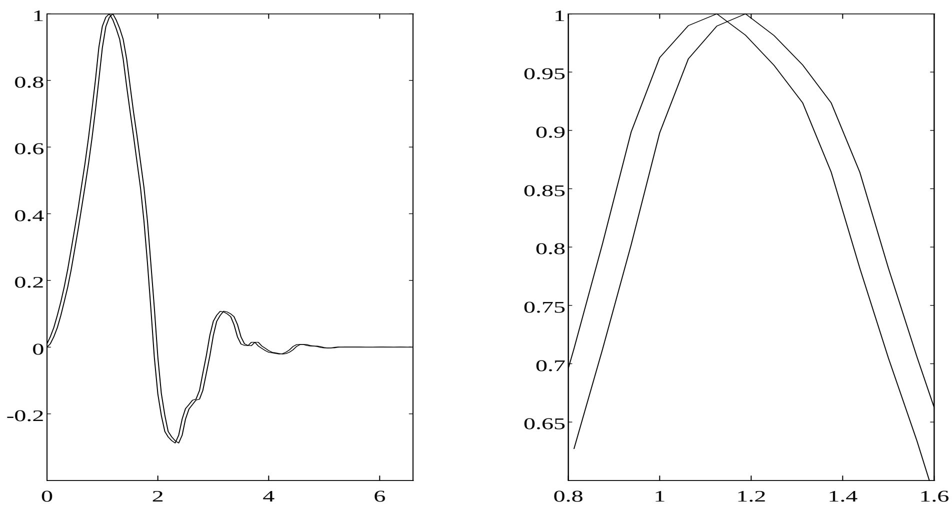


Figure 5.10: Reconstruction of $f(t)$ : $Wf(0,4,k) \approx f(2^{-4}k)$

interesting question is whether $R(\omega)$ in Eqn. 5.18 can be chosen so the *third* moment of the scaling function is the third power of the first moment etc? If it can be done, then the samples of a smooth function will give even higher (than 3) order approximation of $Wf(0, J_f, k)$.

## 5.4 Optimal Wavelets and The Wavelet Sampling Theorem

This section 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 of resolution $J$ depends only on the scaling function $\psi_0(t)$.

2. Given an 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 ([81]) for the special case $M = 2$ (i.e., for Daubechies' orthonormal wavelet bases). The approach in [81] is to obtain upper and lower bounds on the approximation error and to numerically optimize this 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 is used to obtain a simple expression for the approximation error. Using this expression we develop an efficient numerical scheme to solve Problem 1 and illustrate it with examples. 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.

### 5.4.1 The Approximation Error

If a function $f(t)$ is approximated at scale $J$

$$\begin{aligned} f(t) &\approx \sum_k Wf(0,J,k)\psi_{0,J,k}(t) \\ &= \sum_{i=1}^{M-1} \sum_k \sum_{j=-\infty}^{J-1} Wf(i,j,k)\psi_{i,j,k}(t). \end{aligned} \tag{5.21}$$

For fixed $J$ the approximation depends only on $\psi_0(t)$ (and not on the wavelets). We now derive convenient expressions for the Fourier transforms of $Pf(t)$ (the approximation in $W_{0,J}$) and $Qf(t)$ (the approximation error). First define the following Fourier transform pair $a(t)$ and $\hat{a}(\omega)$:

$$a(t) = \int_{\mathbf{R}} d\lambda f(\lambda) M^J \psi_0(M^J\lambda - M^J t). = M^J (f(\cdot) * \psi_0(-M^J \cdot))(t), \tag{5.22}$$

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

The scaling expansion coefficients $Wf(0,J,k)$ are samples of $a(t)$.

$$Wf(0,J,k) = \langle f, \psi_{0,J,k}\rangle = \int_{\mathbf{R}} d\lambda M^{J/2}\psi_0(M^J\lambda - k) = M^{-J/2}a(M^{-J}k). \tag{5.23}$$

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

$$M^J\sum_{k\in\mathbf{Z}}\hat{a}(\omega + 2\pi M^J k) = M^J\sum_{k\in\mathbf{Z}}\hat{f}(\omega + 2\pi M^J k)\psi_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right). \tag{5.24}$$

The approximation of $f(t)$ is given by

$$Pf(t) = \sum_{k\in\mathbf{Z}} Wf(0,J,k)\psi_{0,J,k}(t) = \sum_k a(M^{-J}k)\psi_0(M^J(t - M^{-J}k)).$$

In the Fourier transform domain the above convolution becomes a product.

$$\begin{aligned}\widehat{Pf}(\omega) &= \left[M^J\sum_k \hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right]\left[\frac{1}{M^J}\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right] \\ &= \hat{\psi}_0\left(\frac{\omega}{M^J}\right)\sum_k \hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right). \end{aligned}\tag{5.25}$$

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

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

or equivalently in the transform domain,

$$\widehat{Qf}(\omega) = \hat{f}(\omega)\left(1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right) - \hat{\psi}_0\left(\frac{\omega}{M^J}\right)\sum_{k\neq 0}\hat{f}(\omega + 2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega + 2\pi M^J k}{M^J}\right). \tag{5.26}$$

Eqn. 5.26 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. 4.15) gives a finite-dimensional parameterization of the error $Qf$.

### 5.4.2 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 errors are measured in the frequency domain. In many applications $L^p$ error norms in the time-domain are more meaningful. For example the time domain $L^\infty$ error norm gives the maximum error in the time-domain. Time-domain equivalents of results obtained in this section seem impossible to obtain. However, 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 ([48, p. 333] or [79]) states that that $\hat{g} \in L^q$, where $p$ and $q$ are Hölder conjugates (i.e., $\frac{1}{p} + \frac{1}{q} = 1$), and

$$\|\hat{g}\|_q \leq C \|g\|_p , \tag{5.27}$$

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

$$\left\|\hat{\hat{f}}\right\|_q = \|f\|_q \leq C \left\|\hat{f}\right\|_p . \tag{5.28}$$

For example the time-domain $L^\infty$ 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 above to *minimize error bounds* rather than the error themselves. However, for frequency domain error norms, the results derived are exact for bandlimited signals.

### Optimal Multiresolution Analysis - Transform Domain $L^p$ Error

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

$$(v_i)_j = \begin{cases} \left\{ \prod_{l=0}^{j-1} \sin(\theta_{i,l}) \right\} \cos(\theta_{i,j}) & \text{for } j \in \{0, 1, \ldots, M-2\} \\ \left\{ \prod_{l=0}^{M-1} \sin(\theta_{i,l}) \right\} & \text{for } j = M-1. \end{cases} \tag{5.29}$$

Let $\Theta$ 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.
$$\min_{\Theta} \left[ \frac{1}{2\pi} \int_{\mathbf{R}} d\omega \left| \widehat{Qf} \right|^p \right]^{\frac{1}{p}} = \min_{\Theta} \left\| \widehat{Qf} \right\|_p . \tag{5.30}$$

2.
$$\max_{\Theta} \left[ \frac{1}{2\pi} \int_{\mathbf{R}} d\omega \left| \widehat{Pf} \right|^p \right]^{\frac{1}{p}} = \max_{\Theta} \left\| \widehat{Pf} \right\|_p . \tag{5.31}$$

One minimizes the $p^{th}$ norm of the approximation error, while the other maximizes the $p^{th}$ norm of the approximant. When $p = 2$, and the basis is ON, $Pf(t) \perp Qf(t)$,

$$\left\| \widehat{Pf} \right\|^2 + \left\| \widehat{Qf} \right\|^2 = \left\| \hat{f} \right\|^2 = \|f\|^2, \tag{5.32}$$

and the problems are equivalent. We consider only the minimization of the approximation error.

Eqn. 5.26 yields a complicated expression for $\left\| \widehat{Qf} \right\|_p$. If $f(t)$ is bandlimited *and* the basis in ON (not just a WTF) the expression for $\left\| \widehat{Qf} \right\|_p$ can be simplified. One splits the frequency axis into bins (say $\Omega_l = \left\{ \omega \mid lM^J\pi \leq |\omega| \leq (l+1)M^J\pi \right\}$, $l \in \mathbf{Z}$) and expresses 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 relatively simplified. A similar approach can be used to obtain an expression for $\left\| \widehat{Pf} \right\|_p^p$ also.

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

$$\begin{aligned}
\left\|\widehat{Qf}\right\|_p^p &= \frac{1}{2\pi}\int_{\mathbf{R}} d\omega \left|\hat{f}(\omega)\left(1-\left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right)\right. \\
&\qquad \left. -\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\sum_{k\neq 0}\hat{f}(\omega+2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega+2\pi M^J k}{M^J}\right)\right|^p \\
&= \frac{1}{2\pi}\sum_l \int_{\Omega_l} d\omega \left|\hat{f}(\omega)\left(1-\left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right)\right. \\
&\qquad \left. -\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\sum_{k\neq 0}\hat{f}(\omega+2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega+2\pi M^J k}{M^J}\right)\right|^p \\
&= \frac{1}{2\pi}\int_{\Omega} d\omega \left|\hat{f}(\omega)\left(1-\left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right)\right|^p \\
&\qquad +\frac{1}{2\pi}\int_{\Omega} d\omega \left|\hat{f}(\omega)\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right)\right|^p \left[\sum_{l\neq 0}\left|\hat{\psi}_0\left(\frac{\omega-2\pi M^J l}{M^J}\right)\right|^p\right] \\
&= \frac{1}{2\pi}\int_{\Omega} d\omega \left|\hat{f}(\omega)\right|^p S_p(\omega). \qquad (5.33)
\end{aligned}$$

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 the following expression for $\left\|\widehat{Pf}\right\|_p^p$.

$$\begin{aligned}
\left\|\widehat{Pf}\right\|_p^p &= \frac{1}{2\pi}\int_{\mathbf{R}} d\omega \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\left[\sum_k \hat{f}(\omega+2\pi M^J k)\hat{\psi}_0^*\left(\frac{\omega+2\pi M^J k}{M^J}\right)\right]\right|^p \\
&= \int_{\Omega} d\omega \left|\hat{f}(\omega)\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^p \sum_{k\in\mathbf{Z}}\left|\hat{\psi}_0\left(\frac{\omega-2\pi M^J k}{M^J}\right)\right|^p \\
&\stackrel{\text{def}}{=} \int_{\Omega} d\omega \left|\hat{f}(\omega)\right|^p T_p(\omega), \qquad (5.34)
\end{aligned}$$

where one defines

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

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. 5.33 and Eqn. 5.34.

When $p = 2$ and *one has an ON basis (not a WTF)*, $S_2(\omega)$ and $T_2(\omega)$ take particularly simple forms which can be interpreted easily. When one has an ON basis from Eqn. 4.35 we have

$$\begin{aligned} S_2(\omega) &= \left|1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right|^2 + \left|\hat{\psi}_0^*\left(\frac{\omega}{M^J}\right)\right|^2 \sum_{k\neq 0}\left|\hat{\psi}_0\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^2 \\ &= 1 - 2\left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2 + \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2 \sum_{k\in\mathbf{Z}}\left|\hat{\psi}_0\left(\frac{\omega + 2\pi M^J k}{M^J}\right)\right|^2 \\ &= 1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2, \end{aligned} \tag{5.35}$$

and

$$T_2(\omega) = \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2. \tag{5.36}$$

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

$$\left\|\widehat{Qf}\right\|_2^2 = \frac{1}{2\pi}\int_\Omega d\omega \left|\hat{f}(\omega)\right|^2 \left(1 - \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2\right). \tag{5.37}$$

$$\left\|\widehat{Pf}\right\|_2^2 = \int_\Omega d\omega \left|\hat{f}(\omega)\right|^2 \left|\hat{\psi}_0\left(\frac{\omega}{M^J}\right)\right|^2. \tag{5.38}$$

Also, when $p = 2$, from the orthogonality between $Pf(t)$ and $Qf(t)$, $\left\|\widehat{Pf}\right\|_2^2 + \left\|\widehat{Qf}\right\|_2^2 = \left\|\hat{f}\right\|_2^2$. Indeed this is checked by examining Eqn. 5.38 and Eqn. 5.37. In summary, given a bandlimited signal $f(t)$, 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 5.4.3 gives the details of numerical schemes for Problem 1 and gives examples illustrating that for smooth signals, $K$-regular multiresolution analysis is nearly optimal.

### 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 $\Omega = \{\omega \mid |\omega| \leq M^J\pi\}$ 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 $\{Wf(0,J,k)\}$, act as *generalized samples* of $f(t)$ bandlimited to $\Omega$.

That the scaling expansion coefficients *act* as generalized samples of signals has already been reported in [35] where the arguments are based more on intuition than precise mathematical reasoning. First notice that if we choose the *sinc* wavelet basis the scaling function corresponding to which is the sinc function, Shannon's sampling theorem may also be interpreted as follows: the scaling expansion coefficients $Wf(0,J,k) = M^{-J/2}f(M^{-J}k)$ completely determine $f(t)$ (since they they are the Nyquist rate samples!).

The wavelet sampling theorem, besides giving an interpretation to $Wf(0,J,k)$ also justifies an assumption that is used in practical signal analysis: essentially that $Wf(0,J,k) \approx M^{-J/2}f(M^{-J}k)$. Two other reasons for this assumption may also be found in the literature - the first based on the idea that for sufficiently large $J$, $M^J\psi_0(M^Jt)$ approaches the Dirac measure $\delta(t)$ and therefore $Wf(0,J,k) = \langle f, \psi_{0,J,k}\rangle \approx M^{-J}2f(M^{-J}k)$, 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 $Wf(0,J,k)$ [37].

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. All $K$-regular multiplicity $M$ wavelet bases satisfy the conditions of the wavelet sampling theorem.

**Theorem 38** Let $f(t)$ be bandlimited to $\Omega$ (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., $Wf(0, J, k)$) with respect to a multiplicity $M$ ON 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

$$f(t) = \sum_k Wf(0, J, k) c_{\psi_0}(t - M^{-J}k). \tag{5.39}$$

**Proof:** First notice that

$$\left|\hat{\psi}_0(\omega)\right| = \prod_{j=1}^{\infty} \left| \frac{1}{\sqrt{M}} H_0\left(\frac{\omega}{M^j}\right) \right|. \tag{5.40}$$

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 $[-\frac{\pi}{M}, \frac{\pi}{M}]$, $H_0(\omega/M^j)$ is non-zero on $[-\pi, \pi]$ for all $j \geq 1$. Then from Theorem 15.5 in [74] 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 exists bandlimited functions that cannot be recovered. Take, for instance, a pure tone at $M^J\omega_0$. Then $a(t)$ (in Eqn. 5.22) is zero and hence $Wf(0, J, k)$ is zero for all $k$. Therefore, one cannot have any $c_{\psi_0}(t)$ such that Eqn. 5.39 holds.

Now let $\hat{\psi}_0(\omega)$ be non-zero on $[-\pi, \pi]$. To prove that Eqn. 5.39 holds the following idea is useful. The Fourier transform of $Wf(0, J, k)$ considered as an impulse train is the periodization of the Fourier transform of $a(t)$ (in Eqn. 5.22) (i.e periodization of $\hat{f}(\omega)\hat{\psi}_0^\star(\frac{\omega}{M^J})$). Therefore, in order to recover $f(t)$ we have to be able $\hat{f}$ from the periodization. So we define,

$$\hat{c}_{\psi_0}(\omega) = \begin{cases} \left[M^{J/2}\hat{\psi}_0(\frac{\omega}{M^J})\right]^{-1} & \text{for } \omega \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{5.41}$$

This function is well-defined because $\hat{\psi}_0(\omega)$ does not vanish on $[-\pi, \pi]$. Now the Fourier transform of $\sum_k Wf(0, J, k)c_\psi(t - M^{-J}k)$ (because of the bandlimitedness of $c_{\psi_0}$) is only affected by the first period of the periodization and is given by

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

□

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.

**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. 5.33 that

$$\left\|\widehat{Qf}\right\|_p^p \leq \left\|\hat{f}(\omega)\right\|_p^p \sup_{\omega\in\Omega} S_p(\omega). \tag{5.42}$$

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 that $\psi_0(t)$ that solves the problem

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

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. 5.35 that $S_2(\omega) = 0$! Therefore, for *sinc* wavelet, the error $Qf$ is always zero.

Eqn. 5.42 also has the following important consequence. It says that bandlimited signals are *essentially scale-limited* [68].

**Definition 15** A signal $f(t)$ is essentially $\epsilon$ scale-limited to scale $J$, if for all $T \in \mathbb{R}$

$$\|Qf(t-T)\|_2 \leq \epsilon \|f\|_2 .$$

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 $\epsilon$-scale-limited.

The above definition is meaningful only if $\epsilon$ can be made to be 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 $\Omega$ and increasing the scale at which the signal is being expanded, we will assume that we are studying a function in $W_{0,J}$ and assume that it is bandlimited to a $\tau\Omega$, where $0 < \tau \leq 1$. Then, $f(t)$ is essentially $\epsilon_\tau$-scale-limited to $J$ 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).

Given a scaling function and arbitrary $\epsilon$, there always exists a scale such that at most a fraction, $\epsilon$, of the energy of any bandlimited signal (and translates thereof) is above scale $J$. The values of $\epsilon_\tau$ as a function of $\tau$ for $K$-regular multiplicity 2 and multiplicity 3 orthonormal wavelet bases are shown in Fig. 5.11. For a fixed $\tau$ and $M$ choosing a more regular (i.e., increasing $K$) wavelet basis reduces $\epsilon_\tau$.

### Robust Multiresolution Analysis - $L^p$ to $L^1$

Sometimes, in an approximation the maximum error or $L^\infty$ 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,