Orthonormal wavelet bases

In the last set of lecture notes, we developed the Haar wavelet basis for decomposing continuous-time signals \( x(t) \in L_2(\mathbb{R}) \):

\[
x(t) = \sum_{n=-\infty}^{\infty} s_{0,n} \phi_{0,n}(t) + \sum_{j=0}^{\infty} \sum_{n=-\infty}^{\infty} w_{j,n} \psi_{j,n}(t),
\]

where the (orthonormal) basis functions are scaled and shifted versions of two template functions:

\[
\begin{align*}
\phi_0(t) &= \begin{cases} 
1, & 0 \leq t \leq 1, \\
0, & \text{otherwise} 
\end{cases}, \\
\psi_0(t) &= \begin{cases} 
1, & 0 \leq t < 1/2, \\
-1, & 1/2 \leq t < 1, \\
0, & \text{otherwise} 
\end{cases}.
\end{align*}
\]

\[
\phi_{0,n}(t) = \phi_0(t - n), \quad \psi_{j,n}(t) = 2^{j/2} \psi_0(2^j t - n).
\]

The two template functions were linear combinations of shifts of a contracted version of \( \phi_0(t) \):

\[
\begin{align*}
\phi_0(t) &= \phi_0(2t) + \phi_0(2t - 1), \\
\psi_0(t) &= \phi_0(2t) - \phi_0(2t - 1).
\end{align*}
\]

This gave us the very nice interpretation of the wavelet coefficients \( w_{j,n} \) capturing the differences between piecewise-constant approximations of \( x(t) \) at different dyadic scales,

\[
x(t) = \underbrace{P_{V_0}[x(t)] + P_{W_0}[x(t)]}_{=P_{V_1}[x(t)]} + \underbrace{P_{W_1}[x(t)]}_{=P_{V_2}[x(t)]} + \underbrace{P_{W_2}[x(t)]}_{=P_{V_3}[x(t)]} + \cdots.
\]

It is natural to ask if we can do something similar for other types of approximation spaces \( V_j \), ones that contain things other than just
piecewise-constant functions. Indeed we can, and it leads to a very rich family of orthonormal wavelet bases.

As in the Haar case, everything will follow from properties of a scaling function $\phi_0(t)$. The first thing we must do is carefully write down some properties of $\phi_0(t)$ that lead to consistent multiscale approximations.

**Multiscale approximation: scaling spaces**

For a given $\phi_0(t)$, the first approximation space $\mathcal{V}_0$ is set of signals we can build up from different linear combinations of the integer shifts of $\phi_0(t)$:

$$\mathcal{V}_0 = \overline{\text{Span}}\{\phi_0(t - n)\}_{n \in \mathbb{Z}}.$$  

The first thing we want is for $\{\phi_0(t - n)\}_{n \in \mathbb{Z}}$ to be an orthobasis, so we ask that

$$(P1) \quad \langle \phi_0(t - k), \phi_0(t - n) \rangle = \begin{cases} 1, & k = n, \\ 0, & k \neq n. \end{cases}$$

Now set

$$\phi_{j,n}(t) = 2^{j/2} \phi_0(2^j t - n),$$

so the function $\phi_0(2^j t - n)$ is formed by contracting $\phi_0(t)$ by a factor of $2^j$, then shifting the result on a grid with spacing $2^{-j}$. For a fixed scale $j$, define

$$\mathcal{V}_j = \overline{\text{Span}}\{\phi_{j,n}(t)\}_{n \in \mathbb{Z}}.$$  

$^1$Technically, this is the set of signals we can approximate arbitrarily well from different linear combinations — this is the closure of the span, which we will denote by $\overline{\text{Span}}$.  

138
It follows immediately from the definitions that
\[ x(t) \in \mathcal{V}_0 \iff x(t - k) \in \mathcal{V}_0 \quad \text{for all } k \in \mathbb{Z}, \]
or more generally,
\[ x(t) \in \mathcal{V}_j \iff x(t - 2^{-j}k) \in \mathcal{V}_j \quad \text{for all } k \in \mathbb{Z}. \]
This means that if \( \mathcal{V}_j \) contains a signal, than it also contains every shift of that signal by integer multiples of \( 2^{-j} \). It also follows immediately that
\[ x(t) \in \mathcal{V}_0 \iff x(2t) \in \mathcal{V}_1 \]
\[ \iff x(4t) \in \mathcal{V}_2 \]
\[ \vdots \]
\[ \iff x(2^jt) \in \mathcal{V}_j. \]

Following the Haar case, there are two more key properties we ask of this sequence of approximation spaces; we would like these spaces to be nested,

\[
\text{(P2)} \quad \mathcal{V}_j \subset \mathcal{V}_{j+1}, \quad \text{so } x(t) \in \mathcal{V}_j \Rightarrow x(t) \in \mathcal{V}_{j+1},
\]
and we also want these approximation spaces to cover all of \( L_2(\mathbb{R}) \) in their limit:

\[
\text{(P3)} \quad \lim_{j \to \infty} \mathcal{V}_j = L_2(\mathbb{R}), \quad \text{so } \lim_{j \to \infty} P_{\mathcal{V}_j}[x(t)] = x(t) \quad \text{for all } x(t) \in L_2(\mathbb{R}).
\]

Now the question is: What properties does \( \phi_0(t) \) have to have to ensure (P1)–(P3) hold? While the answer is not straightforward, this question was answered completely in the late 1980s/early 1990s. The
conditions on $\phi_0(t)$ are most easily expressed in terms of the inter-scale relationships between the $\{\phi_{j,n}\}_{n \in \mathbb{Z}}$ and $\{\phi_{j+1,n}\}_{n \in \mathbb{Z}}$. This relationship also connects wavelets to digital filterbanks, a fact that allows discrete wavelet transforms to be computed very efficiently.

Given a $\phi_0(t)$, define the sequence of numbers $g[n]$

$$g_0[n] = \langle \phi_0(t), \sqrt{2} \phi_0(2t - n) \rangle. \quad (1)$$

It turns out that whether properties (P1)–(P3) hold depends entirely on properties of this sequence of numbers. Let $G_0(e^{j\omega})$ be the discrete-time Fourier transform of $g_0[n]$. Then we have following major result:

| If $g_0[n]$ obeys the following three properties, then the approximation spaces $\{\mathcal{V}_j\}_{j \geq 0}$ obey properties (P1)–(P3):
| --- |
| (G1) $|G_0(e^{j\omega})|^2 + |G_0(e^{j(\omega+\pi)})|^2 = 2$, for all $-\pi \leq \omega \leq \pi$
| (G2) $G_0(e^{j0}) = \sum_n g_0[n] = \sqrt{2}$,
| (G3) $|G_0(e^{j\omega})| > 0$ for all $|\omega| \leq \frac{\pi}{2}$.

Proof of the above is long and complicated\(^2\) Note that with (P2) established, we know that $\phi_0(t) \in \mathcal{V}_1$. This gives us an additional

\(^2\)There are a few good references here. I will recommend Chapter 7 of A Wavelet Tour of Signal Processing, by S. Mallat, and Daubechies’ book Ten Lectures on Wavelets.
interpretation of the $g_0[n]$; they tell us how to build up $\phi_0(t)$ out of shifts of the contracted version $\phi_0(2t)$:

$$\phi_0(t) = \sum_{n=-\infty}^{\infty} g_0[n] \sqrt{2} \phi_0(2t - n).$$

(2)

Given a particular $\phi_0(t)$, we can of course generate the $g_0[n]$ using (1), and check to see if the properties above hold. But we can also go the other way. If we design a sequence $g_0[n]$ that obeys the three properties above, it specifies a unique scaling function $\phi_0(t)$. To get $\phi_0(t)$ from $g_0[n]$, we take the continuous-time Fourier transform of both sides of (2):

$$\Phi_0(j\Omega) = \sum_{n=-\infty}^{\infty} g_0[n] \sqrt{2} \int_{-\infty}^{\infty} \phi_0(2t - n) e^{-j\Omega t} \, dt$$

$$= \sum_{n=-\infty}^{\infty} g_0[n] \frac{1}{\sqrt{2}} e^{j\Omega n/2} \Phi_0(j\Omega/2)$$

$$= \frac{1}{\sqrt{2}} G(e^{j\Omega/2}) \Phi_0(j\Omega/2)$$

We can again expand $\Phi_0(j\Omega/2) = \frac{1}{\sqrt{2}} G(e^{j\Omega/4}) \Phi_0(j\Omega/4)$, etc. Condition (G3) above means that the limit exists, and we have

$$\Phi_0(j\Omega) = \left( \prod_{p=1}^{\infty} \frac{G(e^{j2^{-p}\Omega})}{\sqrt{2}} \right) \Phi_0(j0) = \prod_{p=1}^{\infty} \frac{G(e^{j2^{-p}\Omega})}{\sqrt{2}},$$

since $\Phi_0(j0) = 1$ (this follows from integrating both sides of (2) and applying Condition (G2) above). Unfortunately, except in special cases it is hard to compute $\Phi_0(j\Omega)$ past the iterative expression above. This is why wavelets are usually specified in terms of their corresponding sequences $g_0[n]$. 
Multiscale approximation: wavelet spaces

The complementary wavelet spaces and wavelet basis functions can also be generated from the coefficient sequence $g_0[n]$. This is detailed as our second major result:

Suppose $\phi_0(t)$ with corresponding $g_0[n]$ obeys (G1)–(G3). Set

$$g_1[n] = (-1)^{1-n}g_0[1-n],$$

and

$$\psi_0(t) = \sum_{n=-\infty}^{\infty} g_1[n]\sqrt{2}\phi_0(2t-n).$$

Then, along with integer shifts of the scaling function $\phi_{0,n}(t) = \phi_0(t-n)$, the set of all dyadic shifts and contractions of $\psi_0(t)$,

$$\psi_{j,n}(t) = 2^{j/2}\psi_0(2^jt-n), \quad n \in \mathbb{Z}, \quad j = 0, 1, 2, \ldots,$$

form an orthobasis for $L_2(\mathbb{R})$. That is,

$$x(t) = \sum_{n=-\infty}^{\infty} \langle x, \phi_{0,n} \rangle \phi_{0,n}(t) + \sum_{j=0}^{\infty} \sum_{n=-\infty}^{\infty} \langle x, \psi_{j,n} \rangle \psi_{j,n}(t)$$

for all $x(t) \in L_2(\mathbb{R})$.

As with the Haar case, the wavelet coefficients at scale $j$ represent the difference between the approximation of a signal in $\mathcal{V}_j$ and the approximation in $\mathcal{V}_{j+1}$. That is, if we set

$$\mathcal{W}_j = \overline{\text{Span} \{\psi_{j,n}(t)\}_{n \in \mathbb{Z}}}$$

then
1. For fixed $j$, $\langle \psi_{j,n}, \psi_{j,\ell} \rangle = 0$ for $n \neq \ell$. That is, the $\{\psi_{j,n}(t)\}_{n \in \mathbb{Z}}$ are orthobasis for $\mathcal{W}_j$.

2. $\mathcal{W}_j \perp \mathcal{V}_{j'}$ for all $j' \leq j$. Notice that since $\mathcal{W}_j \subset \mathcal{V}_{j+1}$, it follows that the sequence of spaces $\mathcal{V}_0$, $\mathcal{W}_0$, $\mathcal{W}_1$, ... are all mutually orthogonal.

3. $\mathcal{V}_{j+1} = \mathcal{V}_j \oplus \mathcal{W}_j$. That is, every $v(t) \in \mathcal{V}_{j+1}$ can be written as

$$v(t) = P_{\mathcal{V}_j}[v(t)] + P_{\mathcal{W}_j}[v(t)].$$

As the previous property states, these two components are orthogonal to one another.

In summary, this means we can break $L_2(\mathbb{R})$ into orthogonal parts,

$$L_2(\mathbb{R}) = \mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \cdots$$

and we have an orthobases for each of these.
Vanishing moments and support size

In addition to forming an orthobasis with a certain multiscale form, there are other desirable properties that wavelet systems often have.

Vanishing moments. We say that $\psi_0(t)$ has $p$ vanishing moments if

$$\int_{-\infty}^{\infty} t^q \psi_0(t) \, dt = 0, \quad \text{for } q = 0, 1, \ldots, p - 1.$$

This means that $\psi_0(t)$ is orthogonal to all polynomials of degree $p - 1$ or smaller. Since shifting a polynomial just gives you another polynomial of the same order, $\psi_0(t - n)$ is also orthogonal to these polynomials. This means that polynomials that have degree at most $p - 1$ are completely contained in the scaling space $V_0$ — all of the wavelet coefficients of a polynomial are zero.

Compact support. The support of $\psi_0(t)$ is the size of the interval on which it is non-zero. If $\psi_0(t)$ is supported on $[0, L]$, then $\psi_{0,n}(t) = \psi_0(t - n)$ is supported on $[n, n + L]$, and

$$w_{0,n} = \langle x, \psi_{0,n} \rangle = \int_{n}^{n+L} x(t) \psi_{0,n}(t) \, dt.$$

This means that $w_{0,n}$ only depends on what $x(t)$ is doing on $[n, n + L]$ — the wavelet coefficients are recording local information about the behavior of $x(t)$.

These two properties make wavelets very good for representing signals which are smooth except at a few singularities. The following exercise will try to make this point.
Exercise.

1. Suppose that $\psi_0(t)$ is supported on $[0, L]$. What is the support of $\psi_{j,n}(t) = 2^{j/2}\psi_0(2^j t - n)$?

2. Suppose that $x(t)$ is piecewise polynomial as follows

\[ x(t) = \begin{cases} 
0, & t < -1 \\
\text{pth order polynomial}, & -1 \leq t < 1, \\
\text{different pth order polynomial}, & 1 \leq t < 2, \\
0, & t \geq 2.
\end{cases} \]

Suppose that $\psi_0(t)$ is supported on $[0, L]$ and that it has at least $p + 1$ vanishing moments. At most how many wavelet coefficients at scale $j \geq 0$ are non-zero?
Daubechies Wavelets

In the late 1980s, Ingrid Daubechies presented a systematic framework for designing wavelets with vanishing moments and compact support. For any integer $p$, there is a method for solving for the $g_0[n]$ that corresponds to a wavelet with $p$ vanishing moments and has support size $2p - 1$.

Here are the filter coefficients for $p = 2, \ldots, 10$. ($p = 1$ gives you Haar wavelets.):

<table>
<thead>
<tr>
<th>$p$</th>
<th>$h_0[n]$</th>
<th>$g_0[n]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.4829629131499 0.8565163057382 0.2241458908422 -0.125969522551</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.3326705529500 0.8068915909311 0.4398775021182 -0.1359101020100 -0.0854412738821 0.035226291882</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.2350778133092 0.7148465795533 0.5080880707959 0.0279857694147 -0.187034811719 0.0068413818363 0.032883011667 0.010059401785</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.1601029797977 0.6053829259790 0.7243085284383 0.1584281939001 -0.2425098707056 -0.0322448965898 0.077571490840 0.0060241490213</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0811907519999 0.0033357252285</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.0115407435350 0.4966298909388 0.7511350080221 0.3152503517099 -0.2262664996955 -0.1297668675676 0.0975016055876 0.027523885550</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.0098739289340 0.0001582038917 0.00055582201 0.0047722575111 -0.001077301085</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0079820549485 0.3965393194821 0.7291320908462 0.4697882287945 0.000003209592949 0.0224036184994 0.0713092192656 0.0086112091531</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0004929569358 0.0165754416531 0.0122550988556 0.0004295779732 -0.0001806140704</td>
<td></td>
</tr>
</tbody>
</table>

From Mallat, *A Wavelet Tour of Signal Processing*

146
Here are pictures of some of the scaling functions ($N = 2p$ in the captions below):

![Scaling Functions Diagram]

**Figure 6.1.** Daubechies Scaling Functions, $N = 4, 6, 8, \ldots, 40$

From Burrus et al, *Introduction to Wavelets*...
Here are pictures of some of the wavelet functions ($N = 2^p$ in the captions below):

Figure 6.2. Daubechies Wavelets, $N = 4, 6, 8, \ldots, 40$

From Burrus et al, *Introduction to Wavelets* ...