# Wavelets, Approximation Theory, and Signal Processing

Güntürk, Fall 2010  
_Scribe: Evan Chou_

References: Mallat, _Wavelet tour of signal processing_

**Table of contents**

<table>
<thead>
<tr>
<th>Week 1</th>
<th>(9/9/2010)</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fourier Transform</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Uncertainty Principles</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Time Frequency Representations</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>(9/16/2010)</td>
<td>6</td>
</tr>
<tr>
<td>Wavelet Transform</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>The Haar System</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>(9/23/2010)</td>
<td>11</td>
</tr>
<tr>
<td>Convergence of wavelet series</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>(9/30/2010)</td>
<td>16</td>
</tr>
<tr>
<td>Multiresolution Analysis (MRA) Framework (Meyer, Mallat)</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>(10/7/2010)</td>
<td>20</td>
</tr>
<tr>
<td>Finding the Mother Wavelet</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>(10/14/2010)</td>
<td>23</td>
</tr>
<tr>
<td>Week 7</td>
<td>(10/21/2010)</td>
<td>27</td>
</tr>
<tr>
<td>Designing Wavelets</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Meyer’s construction of wavelets in $S$</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Week 8</td>
<td>(10/28/2010)</td>
<td>32</td>
</tr>
<tr>
<td>Vanishing Moments</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Spline Wavelets</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Week 9</td>
<td>(11/4/2010)</td>
<td>36</td>
</tr>
<tr>
<td>Compactly Supported Wavelets</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Week 10</td>
<td>(11/11/2010)</td>
<td>40</td>
</tr>
<tr>
<td>Week 11</td>
<td>(11/24/2010)</td>
<td>44</td>
</tr>
<tr>
<td>Vanishing Moments, Again</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Decay of wavelet coefficients from vanishing moments</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Lipschitz regularity from decay of wavelet coefficients</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Week 12</td>
<td>(12/1/2010)</td>
<td>48</td>
</tr>
<tr>
<td>$L^p$ convergence of wavelet series</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Approximation by wavelet series</td>
<td>51</td>
<td></td>
</tr>
</tbody>
</table>
Fourier Transform

To start, we’ll start with the Fourier transform, studying its deficiencies and how to overcome them.

The convention for Fourier transform we’ll use in this course is the one with $2\pi$ in the exponent, for ease of inversion, Plancherel, sampling at integers without $2\pi$, among other reasons:

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}} f(t) e^{-2\pi i \xi t} dt$$

Initially this makes sense for $f \in L^1$, and as an operator $\mathcal{F}: L^1 \to L^\infty$ (and in fact maps to $C_0$). By the standard extension procedure we can define this for $L^2$ (approximate by $L^1 \cap L^2$), and we have that $\mathcal{F}: L^2 \to L^2$ and is an isometry, so that $\|\hat{f}\|_2 = \|f\|_2$. By interpolation (Riesz-Thorin), we also have have that $\mathcal{F}: L^p \to L^{p'}$ for $1 < p < 2$, a result by the name of Hausdorff-Young.

Also, if we consider the Fourier transform as an operator on the Schwarz space $S$, then $\mathcal{F}: S \to S$ and by duality $\mathcal{F}: S' \to S'$, an operator on the space of tempered distributions such as the dirac delta.

To recover $f$ from $\hat{f}$, we have the inversion formula:

$$f(t) = \int \hat{f}(\xi) e^{2\pi i \xi t} d\xi$$

and viewing $\hat{f} = \mathcal{F}f$ as an analysis of $f$, this can be considered the synthesis (or reconstruction) of $f$ from its frequency content $\hat{f}$.

**Deficiency.** This is a very nonlocal transform. To recover $f(t)$ for even a single $t$, we need all frequency data $\hat{f}(\xi)$ for all $\xi \in \mathbb{R}$.

A quick reason for this is the fact that the complex exponential $e^{2\pi i \xi t}$ is globally supported. This is the main deficiency of the Fourier Transform, when local information about a signal is needed. For instance, in music we can discern which frequencies are being played at which time, and if we were to take the Fourier transform of the music, it would not give this information; instead, it would give only a vague sense of which frequencies were present in the entire song. Such local frequency information would be useful in general for signals with transients: seismic signals, medical signals, images and edges, just to name a few.

Here we investigate why the Fourier transform does not provide an efficient way of capturing local frequency information.

**Uncertainty Principles**

These essentially say that a function cannot be simultaneously localized in time and frequency.

0. Scaling properties:

$$\hat{f}(\delta \xi) = \frac{1}{\delta} \hat{f}\left(\frac{\xi}{\delta}\right)$$

Thus dilating $f$ has the effect of spreading out $\hat{f}$. 
1. If \( f \neq 0 \), then at most one of \( \text{supp} f \) and \( \text{supp} \hat{f} \) can be compact. The reason for this is that if \( \text{supp} f \) is compact, then
\[
\hat{f}(z) = \int f(t)e^{2\pi izt}dt
\]
is analytic, and therefore \( \hat{f}(\xi) \) cannot be compactly supported.

2. **Heisenberg Uncertainty Principle.** Suppose we had \( f \in L^2 \) with \( \|f\|_2 = 1 \). In physics, we can think of \( |f(t)|^2 \) as a pdf. In this context we then define the mean
\[
m_f := \int t|f(t)|^2 dt
\]
and the variance
\[
\sigma_f^2 := \int (t - m_f)^2|f(t)|^2 dt
\]
Similarly we define \( m_\hat{f} \) and \( \sigma_\hat{f}^2 \). Then we have

**Theorem 1. (Heisenberg Uncertainty Principle)**

\[
\sigma_f \sigma_\hat{f} \geq \frac{1}{4\pi}
\]

**Proof.** Note a few facts: \( \sigma_f \) and \( \sigma_\hat{f} \) are invariant under translations and modulations, i.e. if we define \( f_{\alpha, \beta}(t) = e^{2\pi i\alpha t}f(t - \beta) \) then \( \sigma_{f_{\alpha, \beta}} = \sigma_f \). The same holds for \( \hat{f} \) since the Fourier transform exchanges translations and modulations. Thus without loss of generality we may assume that \( m_f = m_\hat{f} = 0 \).

For what follows, we assume \( f \) is Schwarz, and the general case can be obtained by approximation.

\[
1 = \int |f|^2 dt = \int f \bar{f} dt = t(\bar{f}f)'|_\infty^-\infty - \int t(\bar{f}f)' dt = -2\text{Re} \int t f \bar{f}' dt
\]
(Cauchy-Schwarz)
\[
\leq 2 \left( \int t^2 f^2 dt \right)^{1/2} \left( \int |\bar{f}'^2| dt \right)^{1/2}
\]
(Plancherel)
\[
= 2\sigma_f \left( \int |2\pi i \xi \hat{f}(\xi)|^2 d\xi \right)^{1/2}
\]
\[
= 4\pi \sigma_f \sigma_\hat{f}
\]

We conclude that
\[
\frac{1}{4\pi} \leq \sigma_f \sigma_\hat{f}
\]

For general \( f \), we can just approximate with a Schwarz function and take limits. \( \square \)

The minimizers of \( \sigma_f \sigma_\hat{f} \) turn out to be Gaussians.
Examples:

- \( f = \delta_0, \hat{f} = 1 \). (Not quite an instance of above, but an extreme case). Here we have perfect localization in time, but no frequency localization.
- \( f = 1_{[-1/2, 1/2]}, \hat{f} = \frac{\sin \pi \xi}{\pi \xi} \), where \( \sigma_f < \infty \) but \( \sigma_{\hat{f}} = \infty \). This is a milder case.
- \( f(t) = e^{-\pi t^2/2} \). Then \( \sigma_f \sigma_{\hat{f}} = \frac{1}{4\pi} \). This is the balanced case.

3. **Amrein-Betheir.** If \( f \in L^1 \), and \( f \neq 0 \), then

\[
|\text{supp } f| \cdot |\text{supp } \hat{f}| = \infty
\]

where \( |A| \) denotes the measure.

4. **Donaho-Stark** If \((\int_{-\tau}^{\tau} |f|^2)^{1/2} \leq \varepsilon_T \|f\|_2 \) and \((\int_{\Omega} |\hat{f}|^2)^{1/2} \leq \varepsilon_{\Omega} \|\hat{f}\|_2 \), then

\[
|T| |\Omega| \geq (1 - \varepsilon_T - \varepsilon_{\Omega})^2
\]

so that even if we wanted to find sets \( T, \Omega \) such that \( f|_T \) and \( \hat{f}|_\Omega \) capture most of the norms of \( f, \hat{f} \), there is a limitation on how small we can make \( T, \Omega \).

5. **Uniform Uncertainty Principle(s).** **Candés, Tao, others.** For \( f \in \mathbb{R}^N \), and \( k, m \) related by

\[
m \geq ck \log N
\]

If \( \#(\text{supp } f) \leq k \), then there exists a set \( \Lambda \subset \{1, \ldots, N\} \) such that \( |\Lambda| \leq m \) and \( f|_\Lambda \) determines \( f \).

This means that if \( f, g \) are two signals with support of size \( k \), and \( \hat{f}|_\Lambda = \hat{g}|_\Lambda \), then \( f = g \).

Also, if \( h \) is a signal of support size \( 2k \), then \( \hat{h} \) cannot vanish on \( \Lambda \), otherwise we can break \( h \) into a difference two signals of size \( k \) with disjoint support, whose Fourier transforms agree on \( \Lambda \). So this is a discrete version which says that a signal of small support must have some frequency content in \( \Lambda \), and equivalently, if a signal vanishes on \( \Lambda \), then the support of the signal cannot be smaller than \( 2k \).

**Time Frequency Representations**

We will investigate different ways of analyzing local frequency content of functions, starting with the short time (windowed) Fourier transform, and later we will look at wavelets (which analyzes functions at different time scales).

Since the Fourier transform does not give us a good view of the frequency content of a signal near particular times, the idea is to preprocess the signal by restricting it to values near a particular time (to a window of time) and then feeding it to the Fourier transform.
Above $\varphi$ is the window, translated to time $\tau$. We define the windowed Fourier transform to be

$$(T_{\varphi}f)(\xi, \tau) := \int f(t) \overline{\varphi(t-\tau)} e^{-2\pi i \xi t} dt = \langle f, \varphi_{\xi, \tau} \rangle$$

which can be interpreted as inner products with a family of functions $\varphi_{\xi, \tau}(t) = e^{2\pi i \xi t} \varphi(t-\tau)$. This is a continuous transform, and later we will be considering whether we can sample at discrete points on a lattice in the time-frequency plane $(\xi, \tau)$.

How do we recover $f$? Since these are inner products, it is tempting to try placing the coefficients back with the functions, and this does work:

**Theorem 2. (Reconstruction from windowed Fourier transform)**

$$\int \int (T_{\varphi}f)(\xi, \tau) \varphi_{\xi, \tau}(t) d\xi dt$$

**Proof.** Again we will assume that $f$ is sufficiently nice to apply Fubini. The rest is just computation:

$$\int \int (T_{\varphi}f)(\xi, \tau) e^{2\pi i \xi t} \varphi(t-\tau) d\xi dt = \int \int \int f(s) \overline{\varphi(s-\tau)} e^{-2\pi i \xi s} ds e^{2\pi i \xi t} \varphi(t-\tau) d\xi d\tau$$

$$= \int \int f(s) e^{2\pi i \xi (t-s)} \left[ \int \varphi(t-\tau) \varphi(s-\tau) d\tau \right] d\xi ds$$

$$= \int e^{2\pi i \xi t} \mathcal{F} \{ f \Phi_t \}(\xi) d\xi$$

$$= f(t) \Phi_t(t)$$

$$= f(t)$$

This gives us a better idea about which frequencies are present near a particular time. Also, in the reconstruction formula for $f(t)$, we note that if $\varphi$ has sufficiently fast decay, we can truncate the integral with a small loss in error, so we do not really need all values $T_{\varphi}f(\xi, \tau)$ for $(\xi, \tau) \in \mathbb{R}^2$ to recover $f(t)$ approximately. Also, we suspect that the information in $T_{\varphi}f$ is fairly redundant, at least at a glance we are using a $\mathbb{R}^2 \rightarrow \mathbb{R}$ function to represent our signal $f: \mathbb{R} \rightarrow \mathbb{R}$. There is an example that makes this redundancy very apparent.

**Example 3.** Let $\varphi = 1_{[0,1]}$. Then it is already obvious that we lose no information by restricting $\tau$ to integers, as $\sum \varphi(t-k) = 1$ form a partition of unity. Furthermore, for each piece $f(t) \varphi(t-k)$ supported on $[k, k+1]$ we have a Fourier series representation, so we can sample $\xi$ at integers also. In fact, $\{ \varphi_{\xi, \tau}, (\xi, \tau) \in \mathbb{Z}^2 \}$ form an orthonormal basis for $L^2(\mathbb{R})$, and thus

$$f = \sum_{(\xi, \tau) \in \mathbb{Z}^2} \langle f, \varphi_{\xi, \tau} \rangle \varphi_{\xi, \tau} = \sum_{(\xi, \tau) \in \mathbb{Z}^2} (T_{\varphi}f)(\xi, \tau) \varphi_{\xi, \tau}$$

As a quick note, we cannot sample any less here or else we will actually lose part of the original signal. The other thing to note is that $1_{[0,1]}$ is not nice, as $f 1_{[0,1]}$ is not even continuous (even on the torus $\mathbb{T}$), and thus the decay of the Fourier series coefficients will be slow, which will not allow us to truncate the series for a finite approximation.
In the 1940’s, Gabor suggested using Gaussian windows, since they balance smoothness and locality well. However, it turns out that sampling the windowed Fourier transform at integers does not work.

**Notation:** For $\xi_0 > 0, \tau_0 > 0$, we write

$$\varphi_\xi \tau := \varphi_{m_n} \xi \tau$$

and

$$G(\varphi, \xi_0, \tau_0) = \{ \varphi_{m_n} \xi \tau, (m, n) \in \mathbb{Z}^2 \}$$

The family $G$ we will call a **Gabor system**. We can then ask the following questions.

**Question 1:** When is a given Gabor system an orthonormal basis?

We already have an example above. A necessary condition is that $\xi_0 \tau_0 = 1$. But this condition is not sufficient, just considering the earlier example, since we cannot sample at $(\frac{1}{2}, 2)$ for instance. In general, consider $\varphi$ with $\text{supp}(\varphi) \subset [0, \tau_0]$

**Question 2:** When is the system complete? (i.e. it has the property that if $\langle f, \varphi_{m_n} \rangle = 0$ for all $m, n$, then $f = 0$).

A necessary condition is that $\xi_0 \tau_0 \leq 1$. But this condition is not sufficient for the same reason.

Later we will also show

**Theorem 4. (Balian-Low)** If $G(\varphi, 1, 1)$ is an orthonormal basis, then either $\sigma_{\varphi} = \infty$ or $\sigma_{\varphi} = \infty$.

In particular, this means that Gaussians can never yield a Gabor system that is an orthonormal basis.

**Week 2**  
**(9/16/2010)**

When is a good time frequency localization possible for $\varphi$? Last time we saw a limitation from the Heisenberg Uncertainty Principle (Theorem 1):

$$\sigma_f \sigma_{\hat{f}} \geq \frac{1}{4\pi}$$

Now we turn to proving the Balian-Low theorem concerning Gabor systems.

**Proof. (of Theorem 4)** Suppose $G(\varphi, 1, 1) = \{ \varphi_{m_n} \}_{m,n \in \mathbb{Z}}$ is an orthonormal basis. Define

$$(Pf)(x) := x f(x)$$

$$(Qf)(x) := (Pf)^\vee(x) = \frac{1}{2\pi i} \frac{df}{dx}$$

(i.e. $(Qf)^\vee = Pf$). We want to show that either $P\varphi \notin L^2$ or $Q\varphi \notin L^2$. Also, without loss of generality we can center the functions at 0, i.e. we may assume $m_f = m_j = 0$. Towards a contradiction, assume that both $P\varphi$ and $Q\varphi$ are in $L^2$. We thus have

$$\langle P\varphi, Q\varphi \rangle = \sum_{m_n} \langle P\varphi, \varphi_{m_n} \rangle \langle \varphi_{m_n}, Q\varphi \rangle$$

First, note that

$$\langle P\varphi, \varphi_{m_n} \rangle = \langle P\varphi, \varphi_{m_n} \rangle - n \langle \varphi, \varphi_{m_n} \rangle$$
since $\varphi = \varphi_{0,0}$ and thus $\langle \varphi_{0,0}, \varphi_{m,n} \rangle = 0$ for $(m, n) \neq (0, 0)$, and otherwise $n = 0$. Continuing, we have that

$$\langle P\varphi, \varphi_{m,n} \rangle = \langle (P - nI)\varphi, \varphi_{m,n} \rangle$$

$$= \int (x-n)\varphi(x)e^{-2\pi imx}\overline{\varphi(x-n)} \, dx$$

$$= \int x\varphi(x+n)e^{-2\pi imx}\overline{\varphi(x)} \, dx$$

$$= \langle \varphi_{-m,-n}, P\varphi \rangle$$

We have a similar identity with $Q\varphi$, which we prove from the identity for $P\varphi$:

$$\langle \varphi_{m,n}, Q\varphi \rangle = \langle (\varphi_{m,n})^\wedge, (Q\varphi)^\wedge \rangle$$

$$= \langle (\hat{\varphi})_{n,m}, P\hat{\varphi} \rangle$$

$$= \langle P\hat{\varphi}, (\hat{\varphi})_{n,-m} \rangle$$

$$= \langle (Q\varphi)^\wedge, (\varphi_{-m,-n})^\wedge \rangle$$

$$= \langle Q\varphi, \varphi_{-m,-n} \rangle$$

This implies that

$$\langle P\varphi, Q\varphi \rangle = \sum_{m,n} \langle P\varphi, \varphi_{m,n} \rangle \langle \varphi_{m,n}, Q\varphi \rangle$$

$$= \sum_{m,n} \langle Q\varphi, \varphi_{m,n} \rangle \langle \varphi_{m,n}, P\varphi \rangle$$

$$= \langle Q\varphi, P\varphi \rangle$$

Note also that

$$\langle Pf, g \rangle = \int xf(x)\overline{g(x)} \, dx = \langle f, Pg \rangle$$

$$\langle Qf, g \rangle = \langle (Qf)^\wedge, \hat{g} \rangle = \langle P\hat{f}, \hat{g} \rangle$$

$$= \langle \hat{f}, Pg \rangle = \langle \hat{f}, (Qg)^\wedge \rangle$$

$$= \langle f, Qg \rangle$$

so that both $P, Q$ are self-adjoint. Thus,

$$\langle \varphi, PQ\varphi \rangle = \langle P\varphi, Q\varphi \rangle = \langle Q\varphi, P\varphi \rangle = \langle \varphi, QP\varphi \rangle$$

and

$$\langle \varphi, (PQ - QP)\varphi \rangle = 0 \quad \text{for all } \varphi$$

But also, we can compute $(PQ - QP)\varphi$:

$$(PQ - QP)\varphi = \frac{x}{2\pi i} \frac{d}{dx} \frac{d\varphi}{dx} - \frac{1}{2\pi i} \frac{d}{dx} (x\varphi)$$

$$= -\frac{1}{2\pi i} \varphi$$

So we have a contradiction since

$$\langle \varphi, (PQ - QP)\varphi \rangle = -\frac{1}{2\pi i} \langle \varphi, \varphi \rangle^2 \neq 0$$
As mentioned earlier, the Balian-Low theorem implies that $G(\varphi, 1, 1)$ is not an orthonormal basis for any Gaussian $\varphi$. It is true that if $\xi_0 \tau_0 < 1$ then $G(\varphi, \xi_0, \tau_0)$ forms a tight frame (Seip, Lyubarski).

It turns out that we can get around this theorem if we measure the variance slightly differently.

**Generalizations:**

Define for any $p > 0$,

$$
\sigma_{f,p} := \inf_{m} \left( \int (t - m)^2 p |f|^2 \right)^{1/2p}
$$

For $p = 1$, the $m$ that minimizes this is exactly the mean $m_f$, and hence $\sigma_{f,1} = \sigma_f$. Note that if $\sigma_{f,p} < \infty$, then $\sigma_{f,q} < \infty$ for $q < p$ by Hölder (treat $|f|^2 dt$ as a probability measure).

We then have the following results, which we will not prove:

**Theorem 5. (Balian)** There exists $\varphi$ such that $\sigma_{\varphi,p} \sigma_{\hat{\varphi},p} < \infty$ for all $p < 1$ and $G(\varphi, 1, 1)$ is an orthonormal basis for $L^2$

Note that the example of the rectangular window $\varphi = 1_{[-1/2, 1/2]}$ does not satisfy this, as $\hat{\varphi}(\xi) = \frac{2\sin \pi \xi}{\pi \xi}$ and in order for $\sigma_{\hat{\varphi},p} < \infty$ we need $p < 1/2$ so that the decay is still $o\left(\frac{1}{\xi}\right)$.

**Theorem 6. (Steges)** For any orthonormal basis $\{\psi_k\}$ of $L^2(\mathbb{R})$, and for any $p > 1$,

$$
\left( \sup_k \sigma_{\psi_k,p} \right) \left( \sup_k \sigma_{\hat{\psi}_k,p} \right) = \infty
$$

**Theorem 7. (Bourgain)** There exists $\{\psi_k\}$ an orthonormal basis of $L^2(\mathbb{R})$ such that

$$
\left( \sup_k \sigma_{\psi_k} \right) \left( \sup_k \sigma_{\hat{\psi}_k} \right) < \infty
$$

**Wavelet Transform**

Let $\psi \in L^2(\mathbb{R})$. We will use the notation

$$
\psi^{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t - b}{a} \right)
$$

for $a > 0, b \in \mathbb{R}$. The $\sqrt{a}$ normalization is chosen so that $\psi^{a,b}$ has the same $L^2$ norm as $\psi$. The continuous wavelet transform is defined as

$$
(W_\psi f)(a, b) = \langle f, \psi^{a,b} \rangle
$$

where $a$ is the scale and $b$ is the position. The reconstruction formula is similar to that of the windowed Fourier transform:

$$
f(t) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} db \int_{0}^{\infty} \frac{da}{a^2} (W_\psi f)(a, b) \psi^{a,b}(t)
$$
where \( c_\psi = \int \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi < \infty \) (assumed to be finite). In particular, if this condition is satisfied, it is necessary that \( \hat{\psi}(0) = \int \psi = 0 \) and a sufficient condition for this to be satisfied is that \( \psi \in L^2 \cap L^1 \), \( \hat{\psi}(0) = \int \psi = 0 \) and some regularity of \( \psi \) at \( \xi = 0 \).

**Theorem 8.** Assume \( c_\psi = \int \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi < \infty \). Then for all \( f, g \in L^2(\mathbb{R}) \),

\[
c_\psi \langle f, g \rangle = \int_{-\infty}^{\infty} db \int_{0}^{\infty} \frac{da}{a^2} (W_\psi f)(a, b) (W_\psi g)(a, b)
\]

This implies the reconstruction formula, since if \( \langle f_1, g \rangle = \langle f_2, g \rangle \) for all \( g \), then \( f_1 = f_2 \) (take \( g = f_1 - f_2 \) so that \( \langle f_1 - f_2, g \rangle = \| f_1 - f_2 \|^2 = 0 \)). Note that

\[
\int_{-\infty}^{\infty} db \int_{0}^{\infty} \frac{da}{a^2} (W_\psi f)(a, b) (W_\psi g)(a, b) = \left\langle \int_{-\infty}^{\infty} db \int_{0}^{\infty} \frac{da}{a^2} (W_\psi f)(a, b) \psi^{a,b}(\cdot), g \right\rangle
\]

**Proof.** By properties of Fourier transform,

\[
(\psi^{a,b})^\wedge (\xi) = \sqrt{a} e^{-2\pi ib \xi} \hat{\psi}(a\xi)
\]

Also,

\[
W_\psi f(a, b) = \langle f, \psi^{a,b} \rangle = \left\langle \hat{f}, \hat{\psi}^{a,b} \right\rangle
\]

\[
= \int \hat{f}(\xi) \sqrt{a} e^{-2\pi ib \xi} \hat{\psi}(a\xi) d\xi
\]

**Computation:**

\[
\int db \int \frac{da}{a^2} W_\psi f(a, b) W_\psi g(a, b) = \int db \int \frac{da}{a^2} \left[ \int d\xi \hat{f}(\xi) \hat{\psi}(a\xi) e^{2\pi ib \xi} \right] \left[ \int d\xi' \hat{g}(\xi') \hat{\psi}(a\xi') e^{-2\pi ib \xi'} \right]
\]

\[
= \int \frac{da}{a} \int d\xi \hat{f}(\xi) \hat{\psi}(a\xi) \hat{g}(\xi) \hat{\psi}(a\xi)
\]

\[
= \int d\xi \hat{f}(\xi) \hat{g}(\xi) \int \frac{da}{a} |\hat{\psi}(a\xi)|^2
\]

\[
= \int d\xi \hat{f}(\xi) \hat{g}(\xi) \int \frac{da}{a} |\hat{\psi}(a\xi)|^2
\]

\[
= c_\psi \langle \hat{f}, \hat{g} \rangle = c_\psi \langle f, g \rangle
\]

\[\square\]

We can also talk about the discrete wavelet transform which samples \( a, b \) in the continuous wavelet transform. For sampling dilations and translations it is natural to look at dyadic scales

\[
a \in \{2^{-j}, j \in \mathbb{Z} \}, b \in \{2^{-j}k, j \in \mathbb{Z}, k \in \mathbb{Z} \}
\]

so that for each scale \( a \), the supports of \( \psi^{a,b} \) cover all of \( \mathbb{R} \). We will use the notation

\[
\psi_{j,k}(x) := \psi_{2^{-j},2^{-j}k}(x) = 2^{j/2} \psi(2^j(x - 2^{-j}k)) = 2^{j/2} \psi(2^jx - k)
\]
As we saw earlier, a necessary admissibility condition is that \( \hat{\psi}(0) = 0 \). Recall

\[
\hat{\psi}_{j,k}(\xi) = 2^{-j/2} e^{2\pi i 2^{-j}k \xi} \hat{\psi}(2^{-j}\xi)
\]

\( k \) only affects the phase of \( \hat{\psi}_{j,k} \) whereas \( j \) dilates \( \hat{\psi} \), giving higher frequency localization for smaller \( j \). If we consider the time-frequency locality of the family of functions \( \psi_{j,k} \), we see a different tiling of the time-frequency plane. In the windowed Fourier transform, the sampled functions all have the same time-frequency localization. In the wavelet transform, \( \psi_{j,k} \) has higher frequency localization for negative \( j \) and higher time localization for positive \( j \):

\[
\begin{align*}
&\psi_{j,k}(\xi) \\
j = 1 \\
j = 0 \\
j = -1 \\
j = -2
\end{align*}
\]

The Haar System

(A Haar, 1910) For \( \mathcal{H} = L^2([0, 1]) \), define

\[
H(x) = \psi(x) = \begin{cases} 
1 & x \in [0, 1/2) \\
-1 & x \in [1/2, 1]
\end{cases}
\]

and \( H_{j,k} = \psi_{j,k} \). Then \( \{1_{[0,1]}\} \cup \{H_{j,k}\}_{j=0,k=0,\ldots,2^{-j}-1}^\infty \) forms an orthonormal basis for \( \mathcal{H} \). Orthogonality is obvious. \( 1_{[0,1]} \) and \( H_{j,k} \) are orthogonal since \( H_{j,k} \) have mean zero. Otherwise, for \( \langle H_{j,k}, H_{j',k'} \rangle \), if \( j = j' \) and \( k \neq k' \) then the supports are disjoint, so the inner product is zero. If \( j \neq j' \), then either the supports are disjoint or the wider function is constant on the support of the narrower function, which has mean zero, and thus the inner product is again zero.

We will show completeness next time.
Week 3 (9/23/2010)

Convergence of wavelet series

Some definitions:

- \( I_{j,k} = [k2^{-j}, (k+1)2^{-j}) \). \( \{I_{j,k}, k \in \mathbb{Z}\} \) partitions \( \mathbb{R} \) for all \( j \).

- \( V_j = \{ f \in L^2 : f \) is constant on each \( I_{j,k}, k \in \mathbb{Z}\} \)

  \( f \in V_j \iff f = \sum_k \alpha_k 1_{I_{j,k}} \) with \( \sum |\alpha_k|^2 < \infty \)

  Note that \( V_j \subseteq V_{j+1} \) for all \( j \).

- \( P_j := \) orthogonal projection on \( V_j \).

  **Fact:** \( \min_\alpha \| f - \alpha \|_{L^2(A)} \) is attained at the average value \( \alpha^* = \frac{1}{|A|} \int_A f \).

  Given any other \( \alpha, f - \alpha^* \) and \( \alpha^* - \alpha \) will be orthogonal with respect to the inner product \( \langle f, g \rangle = \frac{1}{|A|} \int f \bar{g} \, dx \), and the rest follows by Pythagorean identity

\[
\| f - \alpha \|^2 = \| f - \alpha^* \|^2 + \| \alpha^* - \alpha \|^2 \geq \| f - \alpha^* \|^2
\]

Thus

\[
P_j f = \sum_k \bar{f}_{j,k} 1_{I_{j,k}}
\]

where \( \bar{f}_{j,k} = \frac{1}{|I_{j,k}|} \int_{I_{j,k}} f = 2^j \int_{I_{j,k}} f \).

- Let \( \varphi = 1_{[0,1)} \), and define

\[
\varphi_{j,k}(x) = 2^j/2 \varphi(2^j x - k) = 2^j/2 1_{I_{j,k}}(x)
\]

Then \( \{\varphi_{j,k} : k \in \mathbb{Z}\} \) is an orthonormal basis for \( V_j \) for all \( j \).

If we set \( c_{j,k}(f) = \langle f, \varphi_{j,k} \rangle = 2^{-j/2} \bar{f}_{j,k} \), then \( P_j f = \sum_k \langle f, \varphi_{j,k} \rangle \varphi_{j,k} \)

- Note that \( P_j \) is well-defined on \( L^1_{loc} \) by the same formula.

Now we define \( W_j \) to be the orthogonal complement of \( V_j \) in \( V_{j+1} \), i.e.

\[
V_{j+1} = V_j \oplus W_j
\]

and let \( Q_j \) be the orthogonal projection on \( W_j \). We call \( W_j \) the “detail space” at scale \( 2^{-j} \). Note that \( Q_j = P_{j+1} - P_j \), and

\[
Q_j f = P_{j+1} f - P_j f = \sum_k \bar{f}_{j+1,k} 1_{I_{j+1,k}} - \sum_k \bar{f}_{j,k} 1_{I_{j,k}}
\]

Note that \( I_{j,k} = I_{j+1,2k} \cup I_{j+1,2k+1} \), so that \( \bar{f}_{j,k} = \frac{1}{2}(\bar{f}_{j+1,2k} + \bar{f}_{j+1,2k+1}) \). Then above we can split the first sum into even and odd terms and compute:

\[
Q_j f = \sum_k \left( \bar{f}_{j+1,2k} 1_{I_{j+1,2k}} + \bar{f}_{j+1,2k+1} 1_{I_{j+1,2k+1}} - \bar{f}_{j,k} (1_{I_{j+1,2k}} + 1_{I_{j+1,2k+1}}) \right)
\]

\[
= \sum_k \left( [\bar{f}_{j+1,2k} - \bar{f}_{j,k}] 1_{I_{j+1,2k}} - [\bar{f}_{j,k} - \bar{f}_{j+1,2k+1}] 1_{I_{j+1,2k+1}} \right)
\]
and since $2 \tilde{f}_{j,k} = \tilde{f}_{j+1,2k} + \tilde{f}_{j+1,2k+1}$, we have that $\tilde{f}_{j+1,2k} - \tilde{f}_{j,k} = \tilde{f}_{j,k} - \tilde{f}_{j+1,2k+1}$, and denoting this quantity $b_{j,k}$, we have that

$$Q_j f = \sum_k b_{j,k} (1_{I_{j+1,2k}} - 1_{I_{j+1,2k+1}}) = \sum_k b_{j,k} 2^{-j/2} \psi_{j,k}$$

where $\psi_{j,k}$ was defined earlier for the Haar system. Note that $\{\psi_{j,k}, k \in \mathbb{Z}\}$ form an orthonormal system for $W_j$. In fact, it is a basis by the above computation, where we can express any $Q_j f \in W_j$ in terms of the elements $\psi_{j,k}$. Define

$$d_{j,k} := \langle f, \psi_{j,k} \rangle = 2^{-j/2} b_{j,k}$$

Now pick any pair $J_0 < J_1$. We have that

$$V_{J_1} = V_{J_0} \oplus W_{J_0} \oplus W_{J_0+1} \oplus \cdots \oplus W_{J_1-1}$$

by repeatedly applying the decomposition $V_{j+1} = V_j \oplus W_j$. $V_{J_0}$ is the function at the coarser scale $2^{-J_0}$, and $W_{J_0}, W_{J_0+1}, \cdots$ are successive refinements (detail spaces). Equivalently,

$$P_{J_1} = P_{J_0} + \sum_{k=J_0}^{J_1-1} Q_k$$

Naturally, we want to know what happens as we take $J_0 \to -\infty$ and $J_1 \to \infty$.

**Theorem 9.** For all $f \in L^2(\mathbb{R})$, $P_j f \to f$ in $L^2$ as $j \to \infty$, i.e. $\|P_j f - f\|_{L^2} \to 0$

**Proof.** The first observation is that it suffices to check this for a dense subset $X$ of $L^2$. This is a standard approximation argument: for $\varepsilon > 0$, we can find $g \in X$ with $\|f - g\|_2 \leq \varepsilon$. Then

$$\|P_j f - f\|_2 \leq \|P_j (f - g)\|_2 + \|P_j g - g\|_2 + \|g - f\|_2 \leq 2\varepsilon + \|P_j g - g\|_2$$

(note $\|P_j (f - g)\|_2 \leq \|f - g\|_2$ since $P_j$ is a projection in $L^2$).

Now if we take the limsup of both sides as $j \to \infty$, we know that $\|P_j g - g\|_2 \to 0$, so

$$\limsup_j \|P_j f - f\|_2 \leq 2\varepsilon$$

as $\varepsilon$ is arbitrary, we have that $\limsup_j \|P_j f - f\|_2 \leq 0$ and thus $\lim_j \|P_j f - f\|_2 = 0$.

So now we find a convenient dense subspace to work with. Here we will use $X = C_c(\mathbb{R})$, the continuous functions with compact support, which is dense in $L^2$. Note that these functions are uniformly continuous, i.e. the modulus of continuity

$$\omega_g(\delta) := \sup_{|x - y| < \delta} |g(x) - g(y)| \to 0 \text{ as } \delta \to 0$$

This means that

$$\sup_k \|g - \tilde{g}_{j,k}\|_{L^\infty(I_{j,k})} \leq \omega_g(2^{-j}) \to 0 \text{ as } j \to \infty$$

This implies that $\|P_j g - g\|_{L^\infty} \to 0$ as $j \to \infty$ (recall $P_j g$ is just $\tilde{g}_{j,k}$ on $I_{j,k}$), and thus by Hölder

$$\|P_j g - g\|_{L^2} \leq |\text{supp}(P_j g - g)|^{1/2} \|P_j g - g\|_2 \to 0$$
noting that we can find $|\text{supp}(P_j g - g)| \leq C$ for large $j$ (in fact $C = |\text{supp}(g)| + 1$ for $j > 0$)

We therefore have the following immediate corollary:

**Corollary 10.** For any $f \in L^2(\mathbb{R})$,

$$f = P_0 f + \sum_{j=0}^{\infty} Q_j f$$

(Take $J_1 \to \infty$ and $J_0 = 0$)

If we restrict ourselves to functions on $[0, 1]$, then this says that

$$f = \tilde{f}_{0,0} + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j - 1} \langle f, \psi_{j,k} \rangle \psi_{j,k}$$

for any $f \in L^2([0, 1])$. Furthermore, if $f$ is continuous, then the convergence is uniform (directly from the proof). Compare this with Fourier series, where uniform convergence is not guaranteed for continuous functions (in fact, not even pointwise convergence...). This Haar basis was the first orthonormal basis for $L^2[0, 1]$ found with this property. Here the basis is

$$\{1_{[0,1]}\} \cup \{\psi_{j,k}, j > 0, k \in \mathbb{Z}\}$$

For $L^2(\mathbb{R})$, we have the orthonormal basis

$$\{\varphi_{0,k}, k \in \mathbb{Z}\} \cup \{\psi_{j,k}, j \geq 0, k \in \mathbb{Z}\}$$

for which we have uniform convergence of the basis expansion of $f \in L^2$ if $f$ is uniformly continuous.

Now let’s consider what happens if we take $J_0 \to -\infty$.

**Theorem 11.** For all $f \in L^2$, $P_j f \to 0$ in $L^2$ as $j \to -\infty$.

**Proof.** Again, it suffices to show the result for a dense subspace $X \subset L^2$, with the exact same reasoning.

For the dense subspace, it will be convenient to work with $X = L_1 \cap L_2$ since as $j \to -\infty$, the support of $P_j f$ grows to $\mathbb{R}$, in which case we will be integrating $f$ on larger and larger sets.

Note $P_j g = \sum_k \langle g, \varphi_{j,k} \rangle \varphi_{j,k}$, and thus by triangle inequality,

$$\|P_j g\|_2 \leq \sum_k |\langle g, \varphi_{j,k} \rangle| = \sum_k \left| \int_{I_{j,k}} g \right| \leq 2^{j/2} \|g\|_1$$

which tends to 0 as $j \to -\infty$.

**Corollary 12.** Taking $J_1 \to \infty$ and $J_0 \to -\infty$, we therefore have that

$$f = \sum_{j=-\infty}^{\infty} Q_j f = \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi_{j,k}$$

where convergence is in $L^2$. 

13
Hence, \( \{ \psi_{j,k}, j, k \in \mathbb{Z} \} \) is an orthonormal basis for \( L^2(\mathbb{R}) \).

**Remark:** Note that \( \int \psi_{j,k} = 0 \) for all \( j, k \), yet \( \int f \) need not be zero, even though we can write \( f \) as a linear combination of \( \psi_{j,k} \). In fact, this will be the case for any \( f \in L^1 \cap L^2 \) with \( \int f \neq 0 \). Note that this is not contradictory since the convergence is not in \( L^1 \).

We can also talk about \( L^1 \) convergence.

**Proposition 13.**

1. If \( f \in L^1 \), then \( P_j f \to f \) in \( L^1 \) as \( j \to \infty \)

2. It is not necessarily true that \( P_j f \to 0 \) in \( L^1 \) as \( j \to -\infty \)

**Proof.** To show (1), it suffices to show that \( \| P_j \|_{1,1} \) are uniformly bounded as \( j \to \infty \). Note that in the proof of Theorem 9, we can use the same dense subspace \( X = C_c \) to show that \( P_j g \to g \) in \( L^1 \) as \( j \to \infty \) (we can transfer from uniform convergence to \( L^1 \) convergence). Now if we take \( g \in X \) with \( \| f - g \|_1 \leq \varepsilon \), and use the triangle,

\[
\| P_j f - f \|_1 \leq \| P_j (f - g) \|_1 + \| P_j g - g \|_1 + \| g - f \|_1 \leq (1 + \| P_j \|_{1,1})\varepsilon + \| P_j g - g \|_1
\]

we would be able to carry out the same argument if we can bound \( \| P_j \|_{1,1} \) uniformly.

Note that if \( f \geq 0 \) then \( P_j f \geq 0 \), so that \( P_j \) is a positive operator (\( P_j \) replaces \( f \) by its average value on intervals, and if \( f \) is never negative, the average value will never be negative). This means that since \( -|h| \leq h \leq |h| \), \(-P_j |h| \leq P_j h \leq P_j |h| \) and thus

\[
\| P_j h \|_1 \leq \int P_j |h| = \sum_{j,k} \int_{I_{j,k}} \left( 2^j \int_{I_{j,k}} |h| \, dx \right) dy = \sum_{j,k} 2^{-j} 2^j \int_{I_{j,k}} |h| \, dx = \int |h| = \| h \|_1
\]

This implies that \( \| P_j \|_{1,1} \leq 1 \) (operator norm), and we can follow through with the density argument as described above.

To show (2), note that \( \int P_j f = \int f \) by the same reasoning as above (note we need to apply Fubini, which is okay since \( f \in L^1 ) \), and therefore

\[
\left| \int f \right| = \left| \int P_j f \right| \leq \int \| P_j f \|_1 = \| P_j f \|_1
\]

Thus if \( \int f \neq 0 \), then \( \| P_j f \|_1 \to 0 \) as \( j \to -\infty \).

\[\square\]

**Remarks**

- Note that this tells us that Haar’s expansion \( f = \langle f, \varphi \rangle \varphi_0 + \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi_{j,k} \) on \([0,1]\) converges in the \( L^1 \) sense as well (which is also not true for the Fourier series).
It might be tempting to think of $P_j f$ as a conditional expectation $E[f \mid F_j]$ where $F_j$ is the $\sigma$-field generated by the intervals $I_{j,k}$; however, we are dealing with Lebesgue measure on $\mathbb{R}$ which is not a probability distribution. Interestingly, if we replace the Lebesgue measure with any probability measure (which would require renormalization of the orthonormal basis), then we can think of conditional expectations, and in fact $P_j f = E[f \mid F_j] \longrightarrow f$ (in $L^p$, $1 \leq p < \infty$, including $\infty$ if $f$ is uniformly continuous) for $j \to \infty$ would follow from martingale convergence theorem (although the elementary proof above works just as well).

Note that

$$\{\emptyset, (-\infty, 0), [0, \infty), \mathbb{R}\} = \mathcal{F}_{-\infty} \subset \cdots \subset \mathcal{F}_{-1} \subset \mathcal{F}_0 \subset \cdots \subset \mathcal{F}_\infty = \mathcal{B}(\mathbb{R})$$, Borel sets on $\mathbb{R}$

For $j \to -\infty$, we would have $E[f \mid \mathcal{F}_j] \longrightarrow E[f \mid \mathcal{F}_{-\infty}]$, which is $\frac{1}{\mu(x < 0)} \int_{x < 0} f$ for $x < 0$ and $\frac{1}{\mu(x > 0)} \int_{x > 0} f$ for $x > 0$. This can proved directly as well. The idea is that given $\varepsilon > 0$, we can find some $j$ (negative), for which $\mu([-2^{-j}, 2^j]) \geq 1 - \varepsilon$. Thus, the two intervals $I_{j,-1}$ and $I_{j,0}$ give almost the entire space, and thus $P_j f$ will be close to $E[f \mid \mathcal{F}_{-\infty}]$.

### Implementing the Haar System

The point of the Haar expansion is that it allows us to work with approximations of a function at different scales. As we saw earlier, we can go from a coarse scale $P_{J_0} f$ to a fine scale $P_{J_1} f$ with the successive refinements $Q_k f$, $J_0 \leq k \leq J_1 - 1$. Let us look at how to transition from coarse to fine scales algorithmically.

Using the notation $d_{j,k}(f) = (f, \psi_{j,k})$ and $c_{j,k}(f) = (f, \varphi_{j,k})$ as before, we note that since $\varphi_{j,k} = 2^{j/2} \mathbf{1}_{I_{j,k}}$ and $\psi_{j,k} = 2^{j/2}(\mathbf{1}_{I_{j+1,2k}} - \mathbf{1}_{I_{j+1,2k+1}})$, we have that

$$\psi_{j,k} = \frac{1}{\sqrt{2}}[\varphi_{j+1,2k} - \varphi_{j+1,2k+1}]$$

$$\varphi_{j,k} = \frac{1}{\sqrt{2}}[\varphi_{j+1,2k} + \varphi_{j+1,2k+1}]$$

Thus we have the formulas

$$c_{j,k} = \frac{1}{\sqrt{2}}[c_{j+1,2k} - c_{j+1,2k+1}]$$

$$d_{j,k} = \frac{1}{\sqrt{2}}[c_{j+1,2k} - c_{j+1,2k+1}]$$

Suppose we have $P_{J_1} f$, the finer approximation of $f$. This is described entirely by the coefficients $c_{J_1,k}$. Using the formulas we obtain $c_{J_1-1,k}$, $d_{J_1-1,k}$ from $c_{J_1,k}$, and so on until we obtain $c_{J_0}, d_{J_0}$.

The boxed elements above give the fine to coarse decomposition, and of course we can invert the formulas to go backwards as well. Let’s examine how the algorithm works in the discrete setting (or, finite interval $[0,1]$):
This is a picture in the case where we are going between \( P_3 f \) and \((P_0 f, Q_k f, 0 \leq k \leq 2)\). We can reconstruct the coefficients \(c_{3,k}\) from \(c_{0,0}, d_{0,0}, d_{1,*}, d_{2,*}\) and vice versa.

Each line in the diagram represents a multiply-add operation to transfer between successive levels. At the top row, we need \(16 = 2^4\) operations, and the second we need \(8 = 2^3\), and at the bottom we need \(4 = 2^2\) operations. In general, to transfer between the coefficients of \(P_J f\) and \((P_0 f, Q_k f, 0 \leq k \leq J - 1)\) we need a total of \(2^J + 1 + 2^J + 2^J + 4 \approx 4N\) operations, which is a linear time algorithm (compare to the FFT).

We can also represent this transformation as a matrix:

\[
\begin{pmatrix}
\frac{1}{\sqrt{8}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{2}} & 1 \\
1 & -1 & -1 & 1 \\
1 & 1 & 1 & -1 \\
-1 & -1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
c_{30} \\
c_{31} \\
c_{32} \\
c_{33} \\
c_{34} \\
c_{35} \\
c_{36} \\
c_{37}
\end{pmatrix}
= \begin{pmatrix}
c_{00} \\
d_{00} \\
c_{10} \\
d_{10} \\
c_{11} \\
d_{11} \\
c_{20} \\
d_{20} \\
c_{21} \\
d_{21} \\
c_{22} \\
d_{22} \\
c_{23} \\
d_{23}
\end{pmatrix}
\]

The Haar basis functions (wavelets) are not ideal since they are not smooth. In terms of applying the Haar decomposition to functions, we will not be able to obtain good estimates (especially not if we want to apply Fourier transforms). We will need to work towards finding similar systems with smooth wavelets.

**Week 4** (9/30/2010)

We have that the Haar basis is a basis for \(L^2\) as well as \(L^p\), and also approximates \(C[0,1]\) in the \(L^\infty\) norm (though since the Haar basis functions are not continuous themselves, it is not a basis for \((C[0,1], \| \cdot \|_\infty)\)). In general, there is a framework that captures more general bases, and we will use this to find smooth basis functions of any specified degree.

**Multiresolution Analysis (MRA) Framework (Meyer, Mallat)**

**Definition 14.** An MRA is a sequence \((V_j)_{j \in \mathbb{Z}}\) of linear closed subspaces of \(L^2(\mathbb{R})\) with the following properties:

1. \(V_j \subset V_{j+1}\) for all \(j \in \mathbb{Z}\)
2. \( \bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}) \), i.e. \( L^2(\mathbb{R}) \) can be approximated by the subspaces \( V_j \).

3. \( \bigcap_{j \in \mathbb{Z}} V_j = \{0\} \)

4. \( f \in V_j \iff f(2 \cdot) \in V_{j+1} \) for all \( j \in \mathbb{Z} \), i.e. \( V_j \) are scaled copies of one another.

5. There exists \( \varphi \in V_0 \) such that \( \{ \varphi(\cdot-k), k \in \mathbb{Z} \} \) is an orthonormal basis for \( V_0 \).

(4) and (5) together imply that \( \{ 2^{j/2} \varphi(2^j x - k), k \in \mathbb{Z} \} \) is an orthonormal basis for \( V_j \). We will also see that the conditions are redundant: (3) can be derived from the other properties.

(5) can also be relaxed to require that \( \{ \varphi(\cdot-k), k \in \mathbb{Z} \} \) be a Riesz basis rather than an orthonormal basis, i.e. \( \{ \varphi(\cdot-k), k \in \mathbb{Z} \} \) spans \( V_0 \) and

\[
C_1 \| c \|_{l^2} \leq \left\| \sum_k c_k \varphi(\cdot-k) \right\|_{L^2} \leq C_2 \| c \|_{l^2}, \text{ for all } c \in l^2
\]

As in the Haar basis case, we define

- \( P_j \) as the orthogonal projection on \( V_j \)
- \( W_j \) is the orthogonal complement of \( V_j \) in \( V_{j+1} \), i.e. \( V_{j+1} = V_j \oplus W_j \). Note that \( f \in W_j \) if and only if \( f(2 \cdot) \in W_{j+1} \).
- \( Q_j \) is the orthogonal projection on \( W_j \), so \( P_{j+1} = P_j + Q_j \).

(1) and (2) above say that

\[
\lim_{j \to \infty} \| P_j f - f \|_2 = 0
\]

which we proved in Theorem 9

and (1) and (3) say that

\[
\lim_{j \to -\infty} \| P_j f \|_2 = 0, \text{ for all } f \in L^2
\]

proved in Theorem 11.

As in Corollary 12, we also have

\[
f = \sum_{j \in \mathbb{Z}} Q_j f
\]

with convergence in \( L^2 \).

Summarizing,

**Proposition 15.**

1. \( \lim_{j \to \infty} \| P_j f - f \|_2 = 0 \)
2. \( \lim_{j \to -\infty} \| P_j f \|_2 = 0, \text{ for all } f \in L^2 \)
3. \( f = \sum_{j \in \mathbb{Z}} Q_j f \) with convergence in \( L^2 \)

**Proposition 16.** (3) follows from properties (1),(2),(4),(5)
Proof. Let \( f \in \bigcap_{j \in \mathbb{Z}} V_j \). Then \( f \in V_{-j} \) for \( j > 0 \). Property (4) implies that \( f(2^j \cdot) \in V_0 \) for \( j > 0 \). Property (5) implies that \( 2^{j/2} f(2^j x) = \sum_k \alpha_k^{(j)} \varphi(x - k) \), with \( \|\alpha^{(j)}\|_2 = \|f\|_{L^2} \) for all \( j \). Taking the Fourier transform, we have that

\[
2^{-j/2} \hat{f}(2^{-j} \xi) = \left[ \sum_k \alpha_k^{(j)} e^{-2\pi i k \xi} \right] \hat{\varphi}(\xi)
\]

Denote \( m_j(\xi) := \sum_k \alpha_k^{(j)} e^{-2\pi i k \xi} \), a 1-periodic function. Note that \( \|m_j\|_{L^2(\mathbb{T})} = \|\alpha^{(j)}\|_2 = \|f\|_{L^2} \). Replace \( \xi \) by \( 2^j \xi \) above:

\[
\hat{f}(\xi) = 2^{j/2} m_j(2^j \xi) \hat{\varphi}(2^j \xi)
\]

We will show that \( \hat{f} = 0 \) by showing that \( \int_I \hat{f} = 0 \) for any interval \( I \). Let \( a, b > 0 \) first. Then

\[
\int_a^b |\hat{f}(\xi)|^2 d\xi \leq \left( \int_a^b |m_j(2^j \xi)|^2 d\xi \right)^{1/2} \left( \int_a^b |\hat{\varphi}(2^j \xi)|^2 d\xi \right)^{1/2}
\]

\[
\leq \left( 2^{-j} \int_{2a}^{2b} |m_j(\xi)|^2 d\xi \right)^{1/2} \left( \int_{2a}^{2b} |\hat{\varphi}(\xi)|^2 d\xi \right)^{1/2}
\]

\[
\leq (2^{-j} [2^j (b-a)])^{1/2} \left( \int_{2a}^{2b} |\hat{\varphi}(\xi)|^2 d\xi \right)^{1/2}
\]

\[
\leq C \left( \int_{2a}^{2b} |\hat{\varphi}(\xi)|^2 d\xi \right)^{1/2} \to 0
\]

as \( j \to \infty \). This argument works the same if \( a, b < 0 \) as well, and in general we can split the interval into three parts, \([a, \delta], [-\delta, \delta], [\delta, b]\), noting that the integral over \([-\delta, \delta]\) can be made arbitrarily small. This implies that \( \hat{f} = 0 \) and therefore \( f = 0 \).

\[ \square \]

Definition: A wavelet \( \psi \) is a function in \( L^2 \) such that

\[
\{ \psi_{j,k} := 2^{j/2} \psi(2^j x - k), j, k \in \mathbb{Z} \}
\]

is an orthonormal basis for \( L^2(\mathbb{R}) \).

We say that the wavelet is associated to a given MRA \( (V_j)_{j \in \mathbb{Z}} \) if \( \{ \psi( \cdot - k) \}_{k \in \mathbb{Z}} \) is an orthonormal basis for \( W_0 \), in which case \( \psi_{j,k} \) defined above is an orthonormal basis for \( L^2(\mathbb{R}) \).

Remark 17. We can find wavelets \( \psi \) for which \( \{ \psi_{j,k} \}_{j, k \in \mathbb{Z}} \) is a basis, but that \( \psi \) is not associated to any MRA. In this case, if we try to find the associated MRA, which by definition must be

\[
V_j = \bigoplus_{l=-\infty}^{\infty} W_l
\]

property (5) can fail. Nevertheless, the MRA framework is very convenient to work with, and many examples will fail under this framework.

Example 18. Let \( V_j := \{ f \in L^2 : \text{supp}(\hat{f}) \subset [-2^j, 2^j] \}, j \in \mathbb{Z} \). Properties (1),(2),(3),(4) are all trivial.

\[
V_0 = \{ f \in L^2 : \text{supp}(\hat{f}) \subset [-1/2, 1/2] \}
\]
Set $\varphi = \mathcal{F}^{-1}(1_{[-1/2,1/2]})$, then since $\{2^{-2\pi i k \xi} \hat{\varphi}, k \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(-1/2, 1/2)$, we have that $\{\varphi(x-k), k \in \mathbb{Z}\}$ is an orthonormal basis for $V_0$, and $\varphi(x) = \frac{\sin \pi x}{\pi x}$. Also,

$$W_0 = \{f \in L^2: \text{supp}(\hat{f}) \subset [-1/2, 1/2]\}$$

the wavelet is $\psi = \mathcal{F}^{-1}(1_{[-1,-1/2]\cup[1/2,1]})$. Note that $J = [-1, -1/2] \cup [1/2, 1]$ is congruent to $[0, 1]$ mod $\mathbb{Z}$, i.e. $J + k, k \in \mathbb{Z}$ is a partition of $\mathbb{R}$. This means that $\{e^{-2\pi i k \xi} 1_J, k \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(J) = \mathcal{F}(W_0)$ and thus $\{\psi(\cdot-k), k \in \mathbb{Z}\}$ is an orthonormal basis for $W_0$. Computing $\psi$ explicitly:

$$\hat{\psi} = 1_{[-1,1]} - 1_{[-1/2,1/2]}$$

$$\psi = \frac{2 \sin (2\pi x) - \sin \pi x}{2 \pi x}$$

$$= \frac{\sin \pi x}{\pi x} \left[2 \cos \pi x - 1\right]$$

which still has bad localization properties. This is often called the Shannon wavelet in connection to the sampling theorem, which says that if $f \in V_0$ then $f(x) = \sum_k f(k) \varphi(x-k)$.

This is related to the Littlewood Paley decomposition, which replaces $\hat{\varphi}$ with smoother cutoffs, but in the process property (5) no longer holds.

So we have the Haar wavelet which lacks smoothness in time but has good localization in time, and the Shannon wavelet which lacks localization in time but has great smoothness in time. (In frequency the roles reverse). They lie at the two extremes of time frequency localization.

**Goal:** Find intermediate cases. Daubechies found orthonormal bases of compactly supported wavelets in $C^k$ for any specified $k$.

**Further Goal:** Understand / characterize smoothness function spaces (Sobolev, Besov, etc) in terms of the wavelet coefficients. (This is impossible without good localization / smoothness)

First, we pose the following question:

**Question:** For what $g \in L^2$ is $\{g(\cdot-k), k \in \mathbb{Z}\}$ an orthonormal system? Or equivalently, when is $\{e^{-2\pi i k \xi} \hat{g}(\xi), k \in \mathbb{Z}\}$ an orthonormal system. (equivalent since $\mathcal{F}$ is an isometry)

**Theorem 19.** $\{g(\cdot-k), k \in \mathbb{Z}\}$ is an orthonormal system if and only if $\sum |\hat{g}(\xi+l)|^2 = 1$ a.e.

(Already we have the example where $g(x) = \frac{\sin \pi x}{\pi x}$)

**Proof.** Define $G(\xi) = \sum_{l \in \mathbb{Z}} |\hat{g}(\xi+l)|^2$, which is 1-periodic, and also well defined since

$$\int_0^1 G(\xi) d\xi = \int_{-\infty}^\infty |\hat{g}(\xi)|^2 = \|g\|_{L^2}^2 < \infty$$

so that $G \in L^1(\mathbb{T})$ and hence is finite a.e. We can then look at the Fourier series

$$\hat{G}(k) := \int_0^1 G(\xi) e^{-2\pi i k \xi} d\xi = \int_{-\infty}^\infty e^{-2\pi i k \xi} \hat{g}(\xi) \overline{\hat{g}(\xi)} d\xi$$

$$= \int_{-\infty}^\infty g(x-k) \overline{g(x)} dx$$

19
using Plancherel. This implies that

$$G \equiv 1$$ on $T \iff \hat{G}(k) = \langle g, g(\cdot - k) \rangle = \delta_k \iff \langle g(\cdot - k), g(\cdot - k') \rangle = \delta_{k,k'}$$

which gives the result. \(\square\)

**Exercise 1.** Let $\hat{\psi} = 1_J$ where $J = I \cup (-I)$ where $I = [2/7, 1/2] \cup [2, 2 + 2/7]$. Then $|J| = 1$, and check that

- $\{2^j J, j \in \mathbb{Z}\}$ and $\{J + l, l \in \mathbb{Z}\}$ are both partitions of $\mathbb{R}$
- Defining $W_j = \{f \in L^2, \text{supp}(\hat{f}) \subset 2^j J\}$ and $V_j = \bigoplus_{-\infty}^{-1} W_l$. Property (5) fails.

**Short term goal:** Given an MRA, what is the corresponding wavelet $\psi$? We will cover this next time.

### Week 5 (10/7/2010)

**Finding the Mother Wavelet**

Given an MRA, we are looking for the function $\psi$ (mother wavelet) for which $\{\psi(\cdot - k)\}$ is an orthonormal basis for $W_0$, then $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ will be an orthonormal basis for $L^2(\mathbb{R})$.

**Basic Relations**

- $f \in V_0$ if and only if $f = \sum \langle f, \varphi_{0,k} \rangle \varphi_{0,k}$, and denoting $c_k(f) = \langle f, \varphi_{0,k} \rangle$, this is true if and only if

  $$\hat{f}(\xi) = \left[ \sum c_k e^{-2\pi i k \xi} \right] \hat{\varphi}(\xi)$$

  Call $a_f(\xi) = \sum c_k e^{-2\pi i k \xi}$, a one-periodic function, and we note $\|a_f\|_{L^2(\mathbb{T})} = \|c_k(f)\|_2 = \|f\|_{L^2(\mathbb{R})}$

- Since $V_0 \subset V_1$, $f \in V_0$ implies that $f \in V_1$ so that $f(x) = \sum \langle f, \varphi_{1,k} \rangle \varphi_{1,k}$ where $\varphi_{1,k} = \sqrt{2} \varphi(2\cdot - k)$. Then

  $$\hat{f}(\xi) = \left[ \frac{1}{\sqrt{2}} \sum \langle f, \varphi_{1,k} \rangle e^{-\pi i k \xi} \right] \hat{\varphi}(\xi/2)$$

  Call $m_f(\xi/2) = \frac{1}{\sqrt{2}} \sum \langle f, \varphi_{1,k} \rangle e^{-\pi i k \xi}$, then we have $\|m_f\|_{L^2(\mathbb{T})} = \frac{1}{\sqrt{2}} \|f\|_{L^2(\mathbb{R})}$.

  So $\hat{f}(\xi) = m_f(\xi/2) \hat{\varphi}(\xi/2)$

**Notation:** For $f = \varphi$, we will write $m_0 := m_{\varphi}$.

For $f = \varphi$, we have that

$$\varphi(x) = \sum h_k \varphi_{1,k}(x)$$

where $h_k = \langle \varphi, \varphi_{1,k} \rangle$ and $\varphi_{1,k}(x) = \sqrt{2} \varphi(2x - k)$. Then

$$\varphi(x) = \sqrt{2} \sum_k h_k \varphi(2x - k)$$

with $\|h\|_2 = 1$. This is called the refinement equation (2 scale difference equation).
In the case of the Haar basis, \( \varphi = 1_{[0,1]} \) and
\[
1_{[0,1]} = \sqrt{2} \left( \frac{1}{\sqrt{2}} 1_{[0,1/2]} + \frac{1}{\sqrt{2}} 1_{[1/2,1]} \right)
\]
so that \( h_0 = h_1 = \frac{1}{\sqrt{2}} \) and \( \varphi_{1,0} = 1_{[0,1/2]}, \varphi_{1,1} = \frac{1}{\sqrt{2}} 1_{[1/2,1]} \).

Let us examine \( m_0(\xi) = \frac{1}{\sqrt{2}} \sum h_k e^{-2\pi ik \xi} \). We have \( \hat{\varphi}(\xi) = m_0(\xi/2) \hat{\varphi}(\xi/2) \). Recall from last time (Theorem 19) that if \( \{ g(\cdot - k), k \in \mathbb{Z} \} \) form an orthonormal basis for \( L^2 \), then \( \sum_l |\hat{g}(\xi + l)|^2 = 1 \) for a.e. \( \xi \). Applying this to \( g = \varphi \), we have
\[
1 = \sum_{l \in \mathbb{Z}} |\hat{\varphi}(2\xi + l)|^2 = \sum_{l \in \mathbb{Z}} |\hat{\varphi}(2\xi + 2l)|^2 + \sum_{l \in \mathbb{Z}} |\hat{\varphi}(2\xi + 2l + 1)|^2 \quad \text{a.e. } \xi
\]
\[
= \sum_{l \in \mathbb{Z}} |m_0(\xi + l)|^2 |\hat{\varphi}(\xi + l)|^2 + \sum_{l \in \mathbb{Z}} |m_0(\xi + 1/2)|^2 |\hat{\varphi}(\xi + l + 1/2)|^2
\]
\[
1 = |m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2
\]

Above we have used the fact that \( m_0 \) is 1-periodic so that \( m_0(\xi + l) = m_0(\xi) \). Summarizing the facts, we have

**Proposition 20.**

- If \( f \in V_0 \), then \( \hat{f}(\xi) = m_0(\xi/2) \hat{f}(\xi/2) \). In particular, \( \hat{\varphi}(\xi) = m_0(\xi/2) \hat{\varphi}(\xi/2) \)
- \( 1 = |m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2 \)

**Example 21.**

1. For Haar, we have \( \varphi = 1_{[0,1]} = 1_{[-1/2,1/2]}(\cdot - 1/2) \), so \( \hat{\varphi}(\xi) = e^{-\pi i \xi} \frac{\sin \pi \xi}{\pi \xi} \), and
\[
m_0(\xi) = \frac{1}{\sqrt{2}} \left( \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} e^{-2\pi i \xi} \right) = \frac{1 + e^{-2\pi i \xi}}{2} = e^{-\pi i \xi} \cos(\pi \xi)
\]
It is easy to see that both properties above are satisfied using the half angle formula. Also, we note that \( m_0(\xi) \) is a low-pass filter, mainly supported near \( 0 \in \mathbb{T} \), and \( m_0(\xi + 1/2) \) is a high-pass filter, mainly supported near \( 1/2 \in \mathbb{T} \).

2. For Shannon, we have \( \varphi = 1_{[-1/2,1/2]} \) so \( \varphi(x) = \frac{\sin \pi x}{\pi x} \),
\[
1_{[-1/2,1/2]} = 1_{[-1/2,1/2]} 1_{[-1,1]}
\]
so \( m_0(\xi) = 1_{[-1/4,1/4]} \). Again, it is easy to check the two properties and that \( m_0(\xi) \) is low pass and \( m_0(\xi + 1/2) \) is high-pass.

Let us now characterize the detail space \( W_0 \). \( f \in W_0 \) if and only if \( f \in V_1 \cap V_0^\perp \). Now we already know that \( f \in V_1 \) if and only if \( \hat{f}(\xi) = m_f(\xi/2) \hat{\varphi}(\xi/2) \) with \( \|m_f\|_{L^2(\mathbb{T})} = \|f\|_{L^2} \).

\( f \) is perpendicular to \( V_0 \) if and only if \( f \) is perpendicular to \( \varphi_{0,k} \) for all \( k \in \mathbb{Z} \), which is true if and only if \( \hat{f} \) is perpendicular to \( e^{-2\pi i k \xi} \hat{\varphi}(\xi) \). Computationally, (the same computation as in Theorem 19) this means that
\[
0 = \int \hat{f}(\xi) \overline{\hat{\varphi}(\xi)} e^{2\pi i k \xi} d\xi = \int_0^1 \left( \sum_l \hat{f}(\xi + l) \overline{\hat{\varphi}(\xi + l)} \right) e^{2\pi i k \xi} \text{ for all } k
noting that \( \sum_l \hat{f}(\xi + l) \hat{\varphi}(\xi + l) \) is a one-periodic function that is well-defined, as
\[
\left| \sum_l \hat{f}(\xi + l) \hat{\varphi}(\xi + l) \right|_{L^1(\mathbb{T})} \leq \sum_l \left| \hat{f}(\xi + l) \hat{\varphi}(\xi + l) \right|_{L^1(\mathbb{T})} \\
\leq \sum_l \left| \hat{f}(\xi + l) \right|_{L^2(\mathbb{T})} \left| \hat{\varphi}(\xi + l) \right|_{L^2(\mathbb{T})} \\
\leq \left( \sum_l \left| \hat{f}(\xi + l) \right|_{L^2(\mathbb{T})}^2 \right)^{1/2} \left( \sum_l \left| \hat{\varphi}(\xi + l) \right|_{L^2(\mathbb{T})}^2 \right)^{1/2} \\
= \| \hat{f} \|_{L^2} \| \hat{\varphi} \|_{L^2} < \infty
\]
Continuing on, the previous statement is equivalent to
\[
\sum_l \hat{f}(\xi + l) \hat{\varphi}(\xi + l) = 0 \quad \text{a.e. } \xi
\]
Now we split into evens and odds again, and replace \( \xi \) by \( 2\xi \). Now we additionally use the fact that \( f \in V_1 \) so that \( \hat{f}(\xi) = m_\lambda(\xi/2) \hat{\varphi}(\xi/2) \).
\[
0 = \sum_l \hat{f}(2\xi + 2l) \hat{\varphi}(2\xi + 2l) + \sum_l \hat{f}(2\xi + 2l + 1) \hat{\varphi}(2\xi + 2l + 1) \\
= \sum_l m_\lambda(\xi + l) \hat{\varphi}(\xi + l) m_0(\xi + l) \hat{\varphi}(\xi + l) + \sum_l m_\lambda(\xi + l + 1) \hat{\varphi}(\xi + l + 1) m_0(\xi + l + 1) \hat{\varphi}(\xi + l + 1) \\
= m_\lambda(\xi) m_0(\xi) \sum_l |\hat{\varphi}(\xi + l)|^2 + m_\lambda(\xi + 1/2) m_0(\xi + 1/2) \sum_l |\hat{\varphi}(\xi + l + 1/2)|^2 \\
0 = m_\lambda(\xi) m_0(\xi) + m_\lambda(\xi + 1/2) m_0(\xi + 1/2)
\]
We use the following elementary lemma: Let \((z_1, z_2) \in \mathbb{C}^2 \) and \((\nu_1, \nu_2) \in \mathbb{C}^2 \) such that \( z_1 \nu_1 + z_2 \nu_2 = 0 \). Then there exists \( \lambda \in \mathbb{C} \) such that \((z_1, z_2) = \lambda(\nu_2, -\nu_1) \). (Proof: if \( \nu_1 \neq 0 \), set \( \lambda = -z_2/\nu_1 \), otherwise \( \nu_2 \neq 0 \) and we set \( \lambda = z_1/\nu_2 \).)

All of this implies that \( f \in W_0 \) if and only if there exists \( \lambda_\lambda(\xi) \) such that
\[
(m_\lambda(\xi), m_\lambda(\xi + 1/2)) = \lambda_\lambda(\xi) (m_0(\xi + 1/2), -m_0(\xi))
\]
i.e. \( m_\lambda(\xi) = \lambda_\lambda(\xi) m_0(\xi + 1/2) \) and \( m_\lambda(\xi + 1/2) = -\lambda_\lambda(\xi) m_0(\xi) \). Combining these shows that
\[
\lambda_\lambda(\xi + 1/2) m_0(\xi) = m_\lambda(\xi + 1/2) = -\lambda_\lambda(\xi) m_0(\xi)
\]
so that \( \lambda_\lambda(\xi + 1/2) = -\lambda_\lambda(\xi) \) whenever \( m_0(\xi) \neq 0 \). Note the second property of \( m_0 \) says that \( |m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2 = 1 \) for a.e. \( \xi \), so that either \( m_0(\xi) \neq 0 \) or \( m_0(\xi + 1/2) \neq 0 \) for all \( \xi \). In the first case, this means that \( \lambda_\lambda(\xi + 1/2) = -\lambda_\lambda(\xi) \), and in the second case, this means that \( \lambda_\lambda(\xi + 1) = \lambda_\lambda(\xi) = -\lambda_\lambda(\xi + 1/2) \), so in fact \( \lambda_\lambda(\xi + 1/2) = -\lambda_\lambda(\xi) \) for all \( \xi \).

Now define \( \nu_\lambda(\xi) := e^{-\pi i \xi} \lambda_\lambda(\xi/2) \). \( \nu_\lambda \) is one-periodic by the anti-symmetry of \( \lambda_\lambda \), so
\[
\nu_\lambda(\xi + 1) = -e^{-\pi i \xi} \lambda_\lambda(\xi/2 + 1/2) = e^{-\pi i \xi} \lambda_\lambda(\xi/2) = \nu_\lambda(\xi)
\]
Thus we have the following characterization for \( W_0 \):

**Proposition 22.** \( f \in W_0 \) if and only if
\[
\hat{f}(\xi) = e^{\pi i \xi} \nu_\lambda(\xi) m_0(\xi/2 + 1/2) \hat{\varphi}(\xi/2)
\]
for some \( \nu_f \in L^2(\mathbb{T}) \) with \( \| \nu_f \|_{L^2(\mathbb{T})} = \| f \|_{L^2(\mathbb{R})} \).

\( \nu_f \) contains information specific to \( f \), and \( m_0(\xi/2 + 1/2) \) acts as a high pass filter.

**Proof.** We already have that \( f \in V_1 \) implies \( \hat{f}(\xi) = m_f(\xi/2) \hat{\varphi}(\xi/2) \). We just showed that

\[
m_f(\xi/2) = \lambda_f(\xi/2) m_0(\xi/2 + 1/2) = e^{\pi i \xi} \nu_f(\xi) m_0(\xi/2 + 1/2)
\]

for some \( \nu_f \) (note we can transfer between \( \lambda_f \) and \( \nu_f \) with the definition) Also,

\[
\| f \|_{L^2(\mathbb{R})}^2 = 2 \| m_f \|_{L^2(\mathbb{T})}^2 = 2 \int_0^{1/2} |m_f(\xi)|^2 d\xi + 2 \int_0^{1/2} |m_f(\xi + 1/2)|^2 d\xi = 2 \int_0^{1/2} |\lambda_f(\xi)|^2 (|m_0(\xi + 1/2)|^2 + |m_f(\xi)|^2) d\xi = 2 \int_0^{1/2} |\lambda_f(\xi)|^2 d\xi = \int_0^1 |\nu_f(\xi)|^2
\]

as desired. \( \Box \)

Now for a function in \( W_0 \) to be the mother wavelet, there is an additional property to be satisfied (from Theorem 19). We will show that

**Theorem 23.** Given an MRA with scaling function \( \varphi \) and the associated low pass filter \( m_0 \), \( \psi \in W_0 \) is a mother wavelet for the given MRA \( \text{i.e. } \psi(\cdot - k) \text{ an orthonormal basis for } W_0 \) if and only if

\[
\hat{\psi}(\xi) = e^{\pi i \xi} m_0(\xi/2 + 1/2) \hat{\varphi}(\xi/2) \gamma(\xi)
\]

where \( |\gamma(\xi)| = 1 \) a.e. \( \xi \) and is 1-periodic.

**Week 6** \hspace{3cm} (10/14/2010)

**Proof.** First assume that the equation for \( \hat{\psi} \) holds. By Proposition 22, with \( \nu_f = \gamma \), we have that \( \psi \in W_0 \) and that \( \| \psi \|_{L^2(\mathbb{R})}^2 = \| \gamma \|_{L^2(\mathbb{T})}^2 = 1 \). We want to check that \( \{ \psi(\cdot - k), k \in \mathbb{Z} \} \) forms an orthonormal basis for \( W_0 \). By Theorem 19, we just need to check that \( \sum |\hat{\psi}(\xi + l)|^2 = 1 \) a.e. We already know that \( \sum |\hat{\varphi}(\xi + l)|^2 = 1 \) a.e. since \( \varphi(\cdot - k), k \in \mathbb{Z} \) is an orthonormal basis for \( V_0 \). Applying the usual computation, we look at \( 2\xi \) and split into even and odds:

\[
\sum_l |\hat{\psi}(2\xi + l)|^2 = \sum_l |m_0(\xi + l/2 + 1/2)|^2 |\hat{\varphi}(\xi + l/2)|^2 = \sum_l |m_0(\xi + l + 1/2)|^2 |\hat{\varphi}(\xi + l)|^2 + \sum_l |m_0(\xi + l + 1)|^2 |\hat{\varphi}(\xi + l + 1/2)|^2 = |m_0(\xi + 1/2)|^2 + |m_0(\xi + l)|^2 = 1
\]
noting the facts from Proposition 20 for the last line, and that \$m_0\$ is 1-periodic. Thus \$\psi(\cdot-k), k \in \mathbb{Z}\$ forms an orthonormal system. Now we show that this orthonormal system spans \$W_0\$. Again from the previous Proposition 22, we know that since \$f \in W_0\$ we can write

\[
\hat{f}(\xi) = e^{\pi i \xi} \nu_f(\xi) m_0(\xi/2 + 1/2) \hat{\phi}(\xi/2) = \frac{\nu_f(\xi)}{\gamma(\xi)} \hat{\psi}(\xi)
\]

so that \$\hat{f}(\xi)\$ is the product of a 1-periodic function and \$\hat{\psi}(\xi)\$. Taking the Fourier series of \$\frac{\nu_f(\xi)}{\pi(\xi)}\$ and inverting the Fourier transform (see remark below) shows that

\[
f(\xi) = \sum_k c_k \psi(\xi - k)
\]

so that \$f \in \text{span}\{\psi(\cdot-k), k \in \mathbb{Z}\}\$. This shows that \$\psi\$ is a wavelet for the given MRA.

Conversely, now assume that \$\psi\$ is a wavelet for the given MRA. Then by Proposition 22 we have some \$\nu_\psi\$ such that \$\hat{\psi}(\xi) = e^{\pi i \xi} \nu_\psi(\xi) m_0(\xi/2 + 1/2) \hat{\phi}(\xi/2)\$, and \$\|\nu_\psi\|_{L^2(\mathbb{T})} = \|\psi\|_{L^2(\mathbb{R})} = 1\$. We also have that

\[
\sum_l |\hat{\psi}(\xi + l)|^2 = 1 \text{ a.e. since } \psi \text{ is an orthonormal system. Then by the same computation as above, we have}
\]

\[
1 = \sum_l |\hat{\psi}(2\xi + l)|^2 = |\nu_\psi(2\xi)|^2 \left( \sum_l |m_0(\xi + l + 1/2)|^2 |\hat{\phi}(\xi + l + 1/2)|^2 + \sum_l |m_0(\xi + l)|^2 |\hat{\phi}(\xi + l + 1/2)|^2 \right)
\]

so that \$|\nu_\psi(\xi)| = 1\$ for a.e. \$\xi\$. \qed

The following is a useful characterization of when a function lies in the span of translates, and we used it in the previous proof:

**Proposition 24.** Let \$\{g(\cdot-k), k \in \mathbb{Z}\}\$ be an orthonormal basis for \$Y \subset L^2\$. We note that \$f \in Y = \text{span}\{g(\cdot-k), k \in \mathbb{Z}\}\$ if and only if \$\hat{f}(\xi) = \lambda_f(\xi) \hat{g}(\xi)\$ where \$\lambda_f\$ is a 1-periodic function.

**Proof.** We showed this in the previous proof by using the Fourier series \$\lambda_f(\xi) = \sum \lambda_f(k) e^{-2\pi ik\xi}\$, so that

\[
\hat{f}(\xi) = \sum \lambda_f(k) e^{-2\pi ik\xi} \hat{g}(\xi) \implies f(x) = \sum \lambda_f(k) g(x-k)
\]

Conversely, if \$f \in Y\$, so that \$f = \sum c_k g(x-k)\$, then taking the Fourier transform shows that \$\hat{f}(\xi) = \left( \sum_k c_k e^{-2\pi ik\xi}\right) \hat{g}(\xi)\$. \qed

From this, we can show that for any \$\gamma\$, 1-periodic with \$|\gamma| = 1\$, if we set \$\hat{g}_\gamma(\xi) := \hat{g}(\xi) \gamma(\xi)\$, then \$\{g_{\gamma}(\cdot-k), k \in \mathbb{Z}\}\$ also forms an orthonormal basis for \$Y\$. This follows from the fact that

\[
\sum |\hat{g}(\xi + l)|^2 = 1 \text{ a.e. } \iff \sum |\hat{g}_\gamma(\xi + l)|^2 = 1 \text{ a.e.}
\]

so that \$g_{\gamma}(\cdot-k), k \in \mathbb{Z}\$ forms an orthonormal system. That the span is the same follows immediately from the remark, so that for any \$f \in Y\$,

\[
\hat{f}(\xi) = \lambda_f(\xi) \hat{g}(\xi) = \frac{\lambda_f(\xi)}{\gamma(\xi)} \hat{g}_\gamma(\xi)
\]

24
just as in the previous proof.

For example, if we take $\gamma(\xi) = e^{2\pi i N \xi}$, this shifts $\psi$ by $N$, which is saying the obvious fact that if $\psi(\cdot - k), k \in \mathbb{Z}$ is an orthonormal basis for $W_0$, then $\psi(\cdot - N - k), k \in \mathbb{Z}$ is an orthonormal basis for $W_0$.

Essentially this observation says that the $\gamma$ of Theorem 23 can be an arbitrary 1-periodic function with $|\gamma| = 1$. The canonical choice of course is just to set $\gamma(\xi) = 1$.

This allows us to compute $\psi$ explicitly in terms of the refinement equation. Recall

$$m_0(\xi) = \frac{1}{\sqrt{2}} \sum_k h_k e^{-2\pi i k \xi}$$

where $h_k = \langle \varphi, \varphi_{1,k} \rangle$. Take $\gamma \equiv 1$, so that

$$\hat{\psi}(\xi) = e^{\pi i \xi} m_0(\xi/2 + 1/2) \hat{\varphi}(\xi/2) = \frac{1}{\sqrt{2}} \sum_k \tilde{h}_k e^{2\pi i k (\xi/2 + 1/2)} e^{\pi i \xi} \hat{\varphi}(\xi/2)$$

$$(k = -l - 1) = \frac{1}{\sqrt{2}} \sum_l (-1)^{-l-1} h_{-l-1} e^{-\pi i \xi} \hat{\varphi}(\xi/2)$$

Inverting the Fourier transform gives

**Proposition 25.**

$$\psi(x) = \sqrt{2} \sum_l \tilde{h}_{-l-1} (-1)^{l-1} \varphi(2x - l)$$

so $\psi$ is built from the coefficients $h_k$, except flipped and with modulated signs.

**Example 26.** Check this with the Haar system. In the case of the Haar basis, we had $\varphi = 1_{[0,1]}$,

$$m_0(\xi) = \frac{1}{\sqrt{2}} \left( \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} e^{-2\pi i \xi} \right) = e^{-\pi i \xi} \cos(\pi \xi)$$

and $h_0 = h_1 = \frac{1}{\sqrt{2}}$. Then

$$\psi(x) = \sqrt{2} \left( \frac{1}{\sqrt{2}} 1_{[0,1]}(2x + 1) - \frac{1}{\sqrt{2}} 1_{[0,1]}(2x + 2) \right) = 1_{[0,1]}(2x + 1) - 1_{[0,1]}(2x + 2) = \begin{cases} 1 & -\frac{1}{2} \leq x \leq 0 \\ -1 & -1 \leq x \leq -\frac{1}{2} \end{cases}$$

Note that it is a shifted and flipped version of what we used before, which of course is equivalent from the earlier remarks.

**Computing the Wavelet Coefficients**

Define $g_k := -\tilde{h}_{-k-1} (-1)^{k-1}$, so that $\varphi(x) = \sqrt{2} \sum h_k \varphi(2x - k)$ and $\psi(x) = \sqrt{2} \sum g_k \varphi(2x - k)$. This means that

$$\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k) = 2^{j+1} \sum_l h_l \varphi(2^{j+1} x - k - l) = \sum_l h_l \varphi_{j+1,l+2k}$$

25
Likewise,

\[
\psi_{j,k} = \sum_l g_l \varphi_{j+1,l+2k}
\]

As before, we will set \(c_{j,k}(f) = \langle f, \varphi_{j,k} \rangle\) and \(d_{j,k}(f) = \langle f, \psi_{j,k} \rangle\). Then taking the inner product with the previous relations,

\[
c_{j,k} = \sum_l \bar{h}_l c_{j+1,l+2k} = \sum_l \bar{h}_{l-2k} c_{j+1,l}
\]

We write \(\sum_l \bar{h}_{-(2k-l)} c_{j+1,l}\) as the convolution \((\bar{h}_{-(\cdot)} * c_{j+1,\cdot})_{2k}\)

\[
c_{j,k} = \sum_l \bar{h}_{l-2k} c_{j+1,l} = (\bar{h}_{-(\cdot)} * c_{j+1,\cdot})_{2k}
\]

\[
d_{j,k} = \sum_l \bar{g}_{l-2k} c_{j+1,l} = (\bar{g}_{-(\cdot)} * c_{j+1,\cdot})_{2k}
\]

As a block diagram,

\[
\begin{array}{c}
(c_{j+1}) \\
\downarrow \ \downarrow
\end{array}
\begin{array}{c}
(h_{-n}) \\
\downarrow 2
\end{array}
\begin{array}{c}
(c_j)
\end{array}
\begin{array}{c}
\downarrow 2
\end{array}
\begin{array}{c}
(d_j)
\end{array}
\begin{array}{c}
(\bar{g}_{-n})
\end{array}
\begin{array}{c}
\downarrow 2
\end{array}
\]

where the first blocks denote convolution and \(\downarrow 2\) denotes the downsampling operator

\[(c_0, c_1, c_2, c_3, \ldots) \rightarrow (c_0, c_2, \ldots)\]

(note we want the \(2k\)-th coefficient of the convolution for the \(k\)-th coefficients for \(c_j, d_j\). As before, given the coefficients for a function \(f\) at the finer level \(V_{j+1}\), we can decompose them to the coefficients at the coarser level \(V_j\) and the coefficients in the detail space \(W_j\). In a matrix form, we have

\[
\begin{pmatrix}
\vdots \\
c_{j,k} \\
d_{j,k} \\
c_{j,k+1} \\
d_{j,k+1} \\
\vdots
\end{pmatrix}
= 
\begin{pmatrix}
\vdots \\
\ldots \\
\bar{h}_0 & \bar{h}_1 & \bar{h}_2 & \ldots \\
\ldots & \bar{g}_0 & \bar{g}_1 & \bar{g}_2 & \ldots \\
\ldots & \bar{h}_0 & \bar{h}_1 & \bar{h}_2 & \ldots \\
\ldots & \bar{g}_0 & \bar{g}_1 & \bar{g}_2 & \ldots \\
\vdots
\end{pmatrix}
\begin{pmatrix}
\vdots \\
c_{j+1,2k} \\
c_{j+1,2k+1} \\
c_{j+1,2k+2} \\
c_{j+1,2k+3} \\
c_{j+1,2k+4} \\
\vdots
\end{pmatrix}
\]

This matrix is an orthogonal matrix.

**Exercise 2.** Write out \(|m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2 = 1\) in terms of \(h_k\) only. Then \(\sum_k h_k \bar{h}_{k+2n} = \delta_n\).

Doing the same for \(m_\psi(\xi) m_0(\xi) + m_\psi(\xi + 1/2) m_0(\xi + 1/2) = 0\) shows that \(\sum h_k \bar{g}_{2k+n} = 0\).

For the reconstruction formula, we note that

\[
P_{j+1,ff} = P_j f + Q_j f = \sum c_{j,t} \varphi_{j,t} + \sum d_{j,t} \psi_{j,t}
\]
so that
\[
c_{j+1,k} = \langle f, \varphi_{j+1,k} \rangle \\
= \langle P_{j+1} f, \varphi_{j+1,k} \rangle \\
= \sum c_{j,l} \langle \varphi_{j,l}, \varphi_{j+1,k} \rangle + \sum d_{j,l} \langle \psi_{j,l}, \varphi_{j+1,k} \rangle \\
= \sum c_{j,l} h_{k-2l} + \sum d_{j,l} g_{k-2l}
\]
noting that
\[
\langle \varphi_{j,l}, \varphi_{j+1,k} \rangle = 2^j 2^{j+1/2} \int \varphi(2^j x - l) \varphi(2^{j+1} x - k) dx = \sqrt{2} \int \varphi(x) \varphi(2x - k + 2l) dx = \langle \varphi_0, \varphi_{1,k-2l} \rangle = h_{k-2l}
\]
and a similar computation holds for \( g_{k-2l} \). We can write the above as
\[
c_{j+1,k} = \left( h \ast \tilde{c}_j,(\cdot) \right) + \left( g \ast \tilde{d}_j,(\cdot) \right)
\]
where \( \tilde{c}_j = (\ldots c_{-1}, 0, c_0, 0, c_1, 0, \ldots) \) and likewise for \( \tilde{d} \) (this operation is called upsampling, which we will denote by \( \uparrow^2 \) in the block diagram:

\[
\begin{array}{c}
\ (c_j) \\
\downarrow \quad \uparrow^2 \\
\ h \\
\ \downarrow \\
\ (c_{j+1})
\end{array}
\]

\[
\begin{array}{c}
\ (d_j) \\
\downarrow \quad \uparrow^2 \\
\ g \\
\ \downarrow \\
\ (d_{j+1})
\end{array}
\]

Note that for practical purposes, instead of working directly with the functions \( f \in L^2, \varphi, \psi \), these equations allow us to simply work entirely in terms of the wavelet coefficients \( c_j, d_j, h_j, g_j \).

For finite sequences \( c_j \), let us compute the time to compute the decomposition \( d_{j-1}, d_{j-2}, \ldots, d_0, c_0 \). As long as \( h, g \) are finite length sequences, we note that the implementation time is still linear! Computing \( c_{j-1}, d_{j-1} \) from \( c_j \) takes a linear amount of computation, say \( CN \) (depending on the support of \( h, g \)). Then, we note that there are half as many coefficients in \( c_{j-1} \), so to compute \( c_{j-2}, d_{j-2} \) from \( c_{j-1} \) takes half the amount of computation \( CN/2 \), and so forth. Adding these up is a geometric series of length at most \( \log_2 N \) (where \( N \) is the number of coefficients in \( (c_j) \)), so the total amount of computation is
\[
CN \left[ 2^0 + \ldots + 2^{-\log_2 N} \right] \leq 2CN.
\]
(Again, compare to FFT(N) which takes \( N \log N \) time). The multiresolution structure is what buys us this speedup.

Week 7 (10/21/2010)

Designing Wavelets

What do we want in a wavelet basis?

1. A basis not just for \( L^2 \) but for other function spaces (Sobolev, Hölder, Besov). This typically requires that \( \psi \) has sufficient smoothness.

2. Local expansions. This requires that \( \psi \) be well localized in time, ideally compact with small support, otherwise fast decay.
3. If we have \( f = \sum d_{j,k} \psi_{j,k} \) , we want sparse decompositions, or decay of the wavelet coefficients for large \( |j| \) (the scale). This is analogous to the case of Fourier basis, where the smoothness of the function corresponds to decay in the Fourier coefficients. For \( k \) (the position) there is no corresponding property, as decay in positional coefficients only describes the decay of the function, and not the smoothness. In any case, we care about wavelets for which the wavelet coefficients decay in \( |j| \) at a rate that corresponds to the smoothness class of \( f \).

This requires vanishing moments for \( \psi \), i.e. \( \int x^k \psi(x) \, dx = 0 \) or \( \hat{\psi}^{(k)}(0) = 0 \) for \( k = 0, 1, \ldots, r \). This is because if functions are locally smooth, then locally they are well-approximated by polynomials. In this case the wavelet coefficients are nothing more than

\[
\int f \psi_{j,k} = \int (f - p_f) \psi_{j,k}
\]

if the degree of \( p_f \leq r \). If \( p_f \) approximates \( f \) well near where \( \psi_{j,k} \) is supported, then the wavelet coefficient is small.

The reason we care about decay is for truncation purposes: we can describe \( f \) with fewer coefficients while minimizing the distortion when we reconstruct \( f \).

4. Fast algorithms (using MRA framework and good time localization)

So far we’ve seen the route  

MRA((V_j), \varphi) \rightarrow \psi ,

characterizing the wavelet \( \psi \) for an MRA. Now we want to design \( \varphi \) so that we construct an MRA corresponding to a wavelet with the properties that we want, i.e. the route we take now is

\[
\varphi \rightarrow \text{MRA}((V_j), \varphi) \rightarrow \psi
\]

Given some \( \varphi \), we can start by setting \( V_j = \text{span}\{\varphi_{j,k} \, , \, k \in \mathbb{Z}\} \). We then ask when this gives rise to an MRA? Here is one ingredient:

**Proposition 27.** Let \( \varphi \in L^2(\mathbb{R}) \) such that \( \{\varphi(\cdot - k), k \in \mathbb{Z}\} \) forms an orthonormal system. If \( \hat{\varphi} \) is continuous at 0 and \( \hat{\varphi}(0) \neq 0 \) then \( \bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}) \).

**Proof.** We want to show that if \( f \perp V_j \) for all \( j \in \mathbb{Z} \) then \( f \equiv 0 \). Let \( f \) be such a function, and let \( \epsilon > 0 \). There exists \( g \in L^2 \) such that supp \( \hat{g} \) is compact, say supp \( \hat{g} \subset [-2^{j-1}, 2^{j-1}] \) such that \( \|f - g\|_2 \leq \epsilon \). This implies that \( \|P_j g\|_2 = \|P_j(f - g)\|_2 \leq \epsilon \). We will be interested in the finer scales \( j \geq J \). Consider the function

\[
h_j(\xi) = \hat{g}(\xi) \frac{\varphi(2^{-j}\xi)}{\|\varphi(2^{-j}\xi)\|_2} \text{ on } [-2^{j-1}, 2^{j-1}]
\]

and write the Fourier series expansion with respect to \( 2^{-j/2} e^{2\pi i k \xi 2^{-j}}, k \in \mathbb{Z} \):

\[
\|h_j\|_{L^2(\mathbb{R})}^2 = \int_{-2^{j-1}}^{2^{j-1}} |\hat{g}(\xi)|^2 |\varphi(2^{-j}\xi)|^2 d\xi
\]

(support of \( \hat{g} \) is \( [-2^{j-1}, 2^{j-1}] \))

\[
= \sum_{k \in \mathbb{Z}} \left| \int \hat{g}(\xi) \varphi(2^{-j}\xi) 2^{-j} e^{-2\pi i k \xi 2^{-j}} d\xi \right|^2
\]

\[
= \sum_{k \in \mathbb{Z}} |\langle \hat{g}, \varphi_{j,k} \rangle|^2 = \sum_{k \in \mathbb{Z}} |\langle g, \varphi_{j,k} \rangle|^2 \quad \text{Parseval}
\]

\[
= \|P_j g\|_{L^2(\mathbb{R})}^2
\]

28
This means that \( \|h_j\|_2 = \|P_j g\|_2 \leq \varepsilon \). We know that as \( j \to \infty \), \( \|h_j\|_2 \to \|\hat{g}\|_2 \cdot |\hat{\phi}(0)|^2 \) since \( \hat{\phi}(\xi 2^{-j}) \to \hat{\phi}(0) \) uniformly on \( \xi \in [-2^{j-1}, 2^{j-1}] \), using the fact that \( \hat{\phi} \) is continuous at 0. Thus,

\[
\|\hat{g}\|_2 = \frac{\varepsilon}{|\hat{\phi}(0)|^2} \quad \text{for all } \varepsilon > 0
\]

Now we use that \( \hat{\phi}(0) \neq 0 \), and transfer to \( f \):

\[
\|f\|_2 \leq \|f - g\|_2 + \|g\|_2 \leq \varepsilon + \frac{\varepsilon}{|\hat{\phi}(0)|^2} \leq C\varepsilon
\]

and since \( \varepsilon \) is arbitrary, this implies \( \|f\|_2 = 0 \) and \( f \equiv 0 \).

\[\square\]

Note that if \((V_j, \phi)\) also satisfies the other properties of an MRA, then earlier we showed that \( \|P_j g\|_2 \to \|g\|_2 \) (needs nesting of \( V_j \), Proposition 15). Since above we showed that \( \|P_j g\|_2 \to \|g\|_2 |\hat{\phi}(0)|^2 \), this implies that \( |\hat{\phi}(0)| = 1 \). In other words,

**Proposition 28.** Let \((V_j, \phi)\) be an MRA. If \( \hat{\phi} \) is continuous at 0, and \( \hat{\phi}(0) \neq 0 \), then \( |\hat{\phi}(0)| = 1 \)

Thus, if we want to construct an MRA in this manner, we would find \( \phi \) satisfying \( \hat{\phi} \) continuous at 0 with \( |\hat{\phi}(0)| = 1 \).

If we assume something even stronger, that \( \phi \in L^1 \) so that \( \hat{\phi} \) is continuous everywhere, then

**Corollary 29.** Let \((V_j, \phi)\) be an MRA. If \( \phi \in L^1 \cap L^2 \) and \( \hat{\phi}(0) \neq 0 \), then

1. \( \hat{\phi}(k) = 0 \) for all \( k \in \mathbb{Z} \setminus \{0\} \)
2. \( \sum_k \phi(x + k) = 1 \) a.e.

**Proof.** Since \( \phi \) is in \( L^1 \), \( \hat{\phi} \) is continuous, and since the translates of \( \phi \) form an orthonormal system (MRA), \( \sum |\hat{\phi}(\xi + l)|^2 = 1 \) a.e., and by continuity, this in fact holds everywhere. But since \( \hat{\phi}(0) = 1 \) (assumption and the previous proposition), it must be the case that \( \hat{\phi}(k) = 0 \) for \( k \in \mathbb{Z} \setminus \{0\} \), and this shows (1).

Now consider \( \Phi(x) = \sum_k \phi(x + k) \), which is 1-periodic. Taking the Fourier series,

\[
\hat{\Phi}(l) = \hat{\phi}(l) = \delta_l \quad \text{F.S.} \quad \text{F.T.}
\]

so that \( \Phi \equiv 1 \) a.e., which shows (2) \[\square\]

**Another relation on \( \phi, \psi \).**

Earlier we had derived the following relations:

\[
\hat{\phi}(\xi) = m_0(\xi/2) \hat{\phi}(\xi/2) \quad \hat{\psi}(\xi) = e^{i\pi \xi} m_0(\xi/2 + 1/2) \hat{\phi}(\xi/2) \gamma(\xi) \quad 1 = |m_0(\xi) + m_0(\xi + 1/2)|^2
\]

where \( m_0, \gamma \) is 1-periodic, in \( L^2(\mathbb{T}) \) and \( |\gamma| = 1 \) (Proposition 20, Theorem 23)
These imply that
\[ |\hat{\phi}(\xi)|^2 + |\hat{\psi}(\xi)|^2 = |\hat{\phi}(\xi/2)|^2 \] for all \( \xi \)

As an aside, we note that plugging in \( \xi = 0 \), then if we choose \( \phi \) so that \( \hat{\phi}(0) = 0 \), i.e. \( \int \psi = 0 \) (note this is not a necessary condition for \( (V_j, \phi) \) to be an MRA)

Now rewrite the above:
\[
|\hat{\phi}(\xi)|^2 = |\hat{\phi}(2\xi)|^2 + |\hat{\psi}(2\xi)|^2 \\
= |\hat{\phi}(4\xi)|^2 + |\hat{\psi}(4\xi)|^2 + |\hat{\psi}(2\xi)|^2 \\
= |\hat{\phi}(2N\xi)|^2 + \sum_{j=1}^{N} |\hat{\psi}(2^j\xi)|^2
\]

Note that since \( \sum_{l \in \mathbb{Z}} |\hat{\phi}(\xi + l)|^2 = 1 \), \( |\hat{\phi}(\xi)|^2 \leq 1 \) a.e. This implies that both \( \sum_{j=1}^{N} |\hat{\psi}(2^j\xi)|^2 \) and \( \hat{\phi}(2N\xi) \) converge a.e. as \( N \to \infty \). Since \( \hat{\phi}(2N\cdot) \to 0 \) in \( L^2 \), it must be that \( \hat{\phi}(2N\cdot) \to 0 \) a.e. Finally, this means that
\[
|\hat{\phi}(\xi)|^2 = \sum_{j=1}^{\infty} |\hat{\psi}(2^j\xi)|^2 \text{ a.e.}
\]

Note that this does not hold everywhere since plugging in \( \xi = 0 \) gives 0 on the right but 1 on the left, and plugging in other integer \( \xi \) gives 0 on the left but not necessarily on the right.

In any case, this allows us to recover \( |\hat{\phi}| \) from \( |\hat{\psi}| \) and vice versa. (Can look at what this looks like graphically, and can compare to the Haar system).

**Summary**

We have the following ingredients. We will choose \( \phi \in L^2 \) such that

1. \( \sum_{l \in \mathbb{Z}} |\hat{\phi}(\xi + l)|^2 = 1 \), ensuring that \( \{\phi(\cdot - k), k \in \mathbb{Z}\} \) is an orthonormal system.

2. \( \hat{\phi} \) continuous at 0, and \( |\hat{\phi}(0)| = 1 \). (satisfies spanning criteria \( \bigcup_j V_j = L^2 \))

3. \( \hat{\phi}(\xi) = m_0(\xi/2) \hat{\phi}(\xi/2) \) for some 1-periodic function \( m_0 \in L^2(\mathbb{T}) \) (a necessary condition for \( \phi \in V_0 \) in an MRA). Note that the same iteration procedure (as above for \( |\hat{\phi}|, |\hat{\psi}| \)) allows us to recover \( \hat{\phi} \) from \( m_0 \)

Then it turns out that all the properties of an MRA will be satisfied by \( (V_j, \phi) \).

We need to check the conditions of Definition 14. As mentioned above, the second condition follows from (2) and the fifth condition follows from (1).

By construction, we have that \( \{\phi_{j,k}, k \in \mathbb{Z}\} \) is an orthonormal basis for \( V_j \). Then from Proposition 24, we have

\[
f \in V_0 \iff \hat{f}(\xi) = (1\text{-periodic fn}) \hat{\phi}(\xi) \iff \frac{1}{2} \hat{f}(\xi/2) = (2\text{-periodic fn}) \hat{\phi}(\xi/2) \iff f(2\cdot) \in V_i
\]

which is the fourth condition. The final condition is the nesting condition \( V_j \subset V_{j+1} \). Now we use (3), which gives the refinement equation

\[
\phi(x) = \sum c_k \phi(2x - k)
\]
with \( c_k = 2m_0(-k) \) (plug in and take Fourier transforms). This gives the nesting, since for \( f \in V_0 \),

\[
f = \sum_j a_j(f) \varphi_{0,j} = \sum_j a_j(f) \sum_k c_k \varphi(2x - 2j - k) = \sum_j a_j(f) c_k \varphi_{1,k+2j} \in V_1
\]

Note that we showed earlier that the third condition is implied by the others.

**Meyer’s construction of wavelets in \( S \)**

We will construct a wavelet \( \psi \) with \( \hat{\psi} \) compactly supported and \( \psi \in C^\infty \). In particular \( \psi \) will not be compactly supported, which will not be as useful. In fact, there is no wavelet that is compactly supported and \( C^\infty \) (can only obtain up to fixed order \( C^k \)). The idea is as follows: Let \( \phi \in C^\infty \) satisfying:

- a) \( \text{supp} \ (\hat{\phi}) \subset [-2/3, 2/3] \)
- b) \( \phi(\xi) = 1 \) on \([-1/3, 1/3]\)
- c) \( \phi \) real valued and even (so that \( \varphi \) is also real valued and even)
- d) \( |\phi(\xi)|^2 + |\phi(\xi - 1)|^2 = 1 \) for all \( \xi \in [0, 1] \).

We will refer to the ingredients (1,2,3) in the previous discussion.

(a,d) imply (1) since the support condition on \( \hat{\phi} \) make it so that the infinite sum \( \sum_{l \in \mathbb{Z}} |\hat{\phi}(\xi + l)|^2 \) consists of two terms for every \( \xi \), which will be consecutive, and by (d) the sum will be 1.

(b) will imply (2) by construction. (a-d) will imply (3) with

\[
m_0(\xi) = \begin{cases} \phi(2\xi) & |\xi| \leq 1/3 \\ 0 & 1/3 \leq |\xi| \leq 1/2 \end{cases}
\]

(and periodized of course)

Note

\[
\hat{\phi}(2\xi) = \begin{cases} m_0(\xi) & |\xi| < 1/3 \\ 0 & |\xi| \geq 1/3 \end{cases}
\]

and

\[
m_0(\xi) \hat{\phi}(\xi) = \begin{cases} m_0(\xi) & |\xi| \leq 1/3 \\ 0 & 1/3 \leq |\xi| \leq 2/3 \\ m_0(\xi) = 0 \text{ on } [1/2, 2/3] \end{cases}
\]

How do we satisfy (d)? Since we are looking for even \( \hat{\phi} \), we want \( |\hat{\phi}(\xi)|^2 + |\hat{\phi}(\xi - 1)|^2 = 1 \). We want to use \( |\cos(\xi)|^2 + |\cos(\frac{\pi}{2} - \xi)|^2 = |\cos(\xi)|^2 + |\sin(\xi)|^2 = 1 \), but \( \cos(\xi) \) is not compactly supported. However, now the problem reduces to finding \( \eta \in C^\infty \) for which \( \eta(\xi) + \eta(1 - \xi) = \frac{\pi}{2} \) for all \( \xi \) with \( \eta(\xi) = \frac{\pi}{2} \) for \( |\xi| \geq 2/3 \) and \( \eta(\xi) = 0 \) for \( |\xi| \leq 1/3 \).

The construction is as follows. First we find a smooth function which is 1 for \( x < 0 \) and 0 for \( x > 1 \) and strictly between 0 and 1 otherwise. We simply convolve a smooth bump with support \([-\varepsilon, \varepsilon] \), \( \varepsilon < 1/2 \), with the function \( H(x) \) where \( H \) is 1 for \( x < 1/2 \) and 0 for \( x > 1/2 \). Call this function \( f \).

Now we take \( \eta(x) = \frac{f(x)}{f(x) + f(1-x)} \), which is smooth since \( f \) is smooth and \( f(x) + f(1-x) > 0 \). Then by construction we have \( \eta(x) + \eta(1-x) = \frac{f(x) + f(1-x)}{f(x) + f(1-x)} = 1 \), and \( \eta(x) = 1 \) for \( x < 0 \) (since \( f(x) = 1 \) and \( f(1-x) = 0 \) when \( x < 0 \) ) and \( \eta(x) = 0 \) for \( x > 1 \).
Then just shift and scale \( \eta \) to fit the requirements above.

* Alternatively, take a symmetric smooth bump \( \alpha(t) \) with support \([1/3, 2/3]\) centered at \(1/2\) with integral \( \frac{\pi}{2} \) and consider

\[
\eta(\xi) = \int_{-\infty}^{\xi} \alpha(t) dt
\]

Note that \( \eta(\xi) + \eta(1-\xi) = \int_{-\infty}^{\xi} + \int_{1-\xi}^{\xi} = \int_{-\infty}^{\xi} + \int_{\xi}^{\infty} = \frac{\pi}{2} \) (using symmetry of \( \alpha \)).

In summary, \( \hat{\psi} \) is \( C^\infty \) and compactly supported, and thus in \( \mathcal{S} \), so that \( \phi \) is in \( \mathcal{S} \). If \( \psi \) is chosen according to our convention, then \( \psi \in \mathcal{S} \) as well. Moreover, \( \psi \) vanishes in a neighborhood of 0, which we can see from

\[
\hat{\psi}(\xi)^2 = \hat{\phi}(\xi/2)^2 - \hat{\phi}(\xi)^2
\]

noting \( \hat{\phi}(\xi) \) is 1 in a neighborhood of zero, and thus \( \hat{\psi}(\xi) \) is 0 in a neighborhood of 0. This means that \( \psi^{(k)}(0) = 0 \) for all \( k = 0, 1, 2, \ldots \) so all moments vanish: \( \int x^k \psi(x) dx = 0 \) for all \( k \).

As mentioned previously, the Meyer wavelet is not usable in practice since it is not compactly supported.

**Week 8**  
(10/28/2010)

**Vanishing Moments**

First, we revisit specifics about vanishing moments. If \( f \) is smooth and \( \psi \) has vanishing moments, then \( \langle f, \psi_{j,k} \rangle \) decays fast in \( j \to \infty \), depending on the smoothness of \( f \) and the number of vanishing moments.

Here is a basic case:

**Theorem 30.** Let \( f \) be \( \text{Lip}_\alpha \cap L^2 \), i.e. \( |f(x) - f(y)| \leq |f|_{\text{Lip}_\alpha} |x - y|^\alpha \), where \( 0 < \alpha \leq 1 \), and let \( \psi \) be such that \( \int \psi(x) dx = 0 \). If \( c_\psi := \int |x|^\alpha |\psi(x)| dx < \infty \), then \( |\langle f, \psi_{j,k} \rangle| \leq C(f, \psi) 2^{-j(\alpha+1/2)} \).

**Remark:** The result should not be confused with exponential decay in frequency, since \( \psi_{j,k} \) in the frequency domain is concentrated around \( \xi \sim 2^{-j} \) so really the decay in frequency is like \( |\xi|^{-\alpha} \). Note that the condition that \( c_\psi < \infty \) means that \( \psi \) should have some order of decay at as \( |x| \to \infty \).

**Proof.** Note

\[
\langle f, \psi_{j,k} \rangle = 2^{j/2} \int f(x) \psi(2^j x - k) dx = 2^{j/2} \int \left[ f(x) - \frac{f(k2^{-j})}{|k2^{-j}|} \right] \psi(2^j x - k) dx
\]

noting that \( \psi \) integrates to 0 and \( \psi_{j,k} \) centers around \( k2^{-j} \), so we extract \( f(k2^{-j}) \) from \( f(x) \). Then taking absolute values inside, and applying Lipschitz constant:

\[
|\langle f, \psi_{j,k} \rangle| \leq 2^{j/2} |f|_{\text{Lip}_\alpha} \int |x - k2^{-j}|^{\alpha} |\psi(2^j x - k)| dx
= 2^{j(1/2 - \alpha)} |f|_{\text{Lip}_\alpha} \int |2^j x - k|^{\alpha} |\psi(2^j x - k)| 2^j dx
= 2^{j(1/2 - \alpha)} |f|_{\text{Lip}_\alpha} c_\psi
\]

\( \square \)
Thus, having the 0-th vanishing moment for \( \psi \) gives results for \( \text{Lip}_\alpha \). For higher vanishing moments, we can subtract a polynomial instead of the constant, and for this we need more smoothness of \( f \) to get (local) polynomial approximation. \( t \) vanishing moments can handle \( f \in \text{Lip}_{t+\alpha} \). The proof is essentially the same.

Note that for \( f \) in a range of \( \text{Lip}_\alpha \) spaces and say \( \psi \) has all vanishing moments, to obtain the optimal decay in \( j \) we note that we need to look at the how \( |f|_{\text{Lip}_\alpha} \) and \( c_\psi \) depend on \( \alpha \) and optimize over \( \alpha \) (there is a tradeoff between the decay rate of \( 2^{-j\alpha} \) and the growth of \( |f|_{\text{Lip}_\alpha}, c_\psi(\alpha) \)).

### Spline Wavelets

At this point, Meyer wavelets can be used to characterize spaces, but implementation-wise we will need other wavelets. In particular we want wavelets with corresponding high pass and low pass having finite support. As an intermediate step to the Daubechies compactly supported wavelets, we will look at splines, piecewise polynomials on uniformly spaced knot sequences (Knots are the points at which the function changes. Uniform spacing will give us shift invariance).

First, some basic background on splines. Let \( d \geq 0 \) be an integer (degree of spline), and fix \( a > 0 \) (spacing). Define

\[
S^d(a\mathbb{Z}) := \left\{ f: \mathbb{R} \to \mathbb{C}, f|_{[ak, a(k+1)]} \text{ is a polynomial of degree } \leq d, f \in C^{d-1} \right\}
\]

e.g. \( S^0(a\mathbb{Z}) \) are piecewise constants and \( S^1(a\mathbb{Z}) \) are piecewise linear functions which are still continuous.

A basis for \( S^d(\mathbb{Z}) \) is given by \( B_d(\cdot - k), k \in \mathbb{Z} \) where \( B_0 = 1_{[0,1]} \) and \( B_{n+1} = B_n * 1_{[0,1]} \).

(Note that as \( n \to \infty, \sqrt{n} B_n(\sqrt{n} \cdot) \) tends to a Gaussian, by the central limit theorem)

We can check the following properties (inductively):

- \( B_d \in S^d(\mathbb{Z}) \). Assuming \( B_n \) is piecewise polynomial on \([k, k+1]\), we note that

\[
B_{n+1}(x) = \int_{x-1}^{x} B_n(y) dy
\]

and for \( x \in [k, k+1] \), we have

\[
B_{n+1}(x) = \int_{x-1}^{k} B_n(y) dy + \int_{k}^{x} B_n(y) dy
\]

which is polynomial since the antiderivative of polynomials is again polynomial.

Also, we show that \( B_{n+1} \in C^n \) assuming \( B_n \in C^{n-1} \). It suffices to check near integer points \( k \in \mathbb{Z} \), since polynomials are smooth. Using properties of convolution, we have that \( B_{n+1}^{(n)} = \frac{d}{dx} \left[ B_n^{(n-1)} \ast 1_{[0,1]} \right] \), then

\[
B_{n+1}^{(n)}(x) = \frac{d}{dx} \left[ \int_{x-1}^{x} B_n^{(n-1)}(y) dy \right] = B_n^{(n-1)}(x) - B_n^{(n-1)}(x-1)
\]

which shows also that \( B_{n+1}^{(n)} \) is continuous (since \( B_n^{(n-1)}(x) \) is continuous by our induction hypothesis).

- \( B_d \) is symmetric about \( \frac{d+1}{2} \)
We wish to show that $B_{n+1}(x) = B_{n+1}(d+2-x)$ given that $B_n(x) = B_n(d+1-x)$:

\[
B_{n+1}(x) = \int_{x-1}^{x} B_n(y) \, dy = \int_{x-1}^{d+2-x} B_n(d+1-y) \, dy = \int_{d+1-x}^{d+2-x} B_n(y) \, dy = B_{n+1}(d+2-x)
\]

- supp$(B_d) = [0, d+1]$

This follows from convolution properties, supp$(f * g) = $supp$(f) +$supp$(g)$ (Minkowski sum).

- $\sum B_d(x - k) \equiv 1$ for all $x$, i.e. $B_d(\cdot - k)$ form a partition of unity.

Let $G(x) = \sum_k B_d(x - k)$, then $\hat{G}(l) = \mathcal{B}_d(l) = [\mathcal{B}_0(l)]^{d+1}$ ($\hat{G}$ is Fourier series and $\mathcal{B}_d$ is Fourier transform). Noting that $\mathcal{B}_0(\xi) = e^{-i\pi \sin(\pi \xi)}$, this implies that $\hat{G}(l) = \delta_l$ so that $G \equiv 1$.

- $\{B_d(\cdot - k), k \in \mathbb{Z}\}$ form a basis for $S^d(\mathbb{Z})$.

The previous two properties implies that the functions are linearly independent. Suppose we have a finite linear combination of $\sum a_k B_d(x - k) = 0$, note that the functions at the end (or beginning) have support where the other functions are not supported, and thus the coefficient must be 0. By repeating this process, we conclude that all the coefficients are 0, by peeling off the functions that are furthest away from the origin.

To show that these functions span the spline space is more difficult (There’s a reference in the wikipedia article for “B-spline”)

**Spline MRA** $(d$ is fixed $)$

Define $V_j := S^d(2^{-j} \mathbb{Z}) \cap L^2(\mathbb{R})$.

Note that we immediately have $V_j \subset V_{j+1}$ (knot sequence of $V_j$ is a subsequence of the knot sequence of $V_{j+1}$).

Also, $\bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R})$. We showed this for $B_0$ (Haar system), but in general we can approximate $L^2$ functions by splines of a fixed degree (but whose knot sequences are finer and finer). This can be achieved inductively, as if we can find a $d-1$ spline $g \in S^{d-1}(2^{-j} \mathbb{Z})$ for which $\|f - g\|_2 \leq \varepsilon$, if we convolve with $2^k 1_{[0, 2^{-k}]}$ for suitable $k$ we can obtain a $d$ spline $h \in S^d(2^{-k} \mathbb{Z})$ for which $\|g - h\|_2 \leq \varepsilon$.

$f \in V_j \iff f(2 \cdot) \in V_{j+1}$ is also immediate from our choice of $V_j$.

What’s left is to find the scaling function $\varphi$ for which $\{\varphi(\cdot - k), k \in \mathbb{Z}\}$ is an orthonormal basis for $V_0$, and then $(V_j, \varphi)$ will satisfy all the properties of an MRA (Definition 14)

We have $\{B_d(\cdot - k), k \in \mathbb{Z}\}$ which is not orthogonal, but is still a basis. It turns out that it is a Riesz sequence as well (prove later)

**Definition:** Let $\mathcal{H}$ be a Hilbert space. $(f_n) \in \mathcal{H}$ is said to be a Riesz sequence if there exists $C_1, C_2 > 0$ such that

$$C_1 \|x\|_2 \leq \left\| \sum_i x_i f_i \right\|_{\mathcal{H}} \leq C_2 \|x\|_2 \quad \text{for all } x \in l_2$$
The left inequality implies linear independence. We have the following characterization of Riesz sequences:

**Proposition 31.** Let \( \varphi \in L^2(\mathbb{R}) \). Then the following are equivalent:

1. \( \{ \varphi(\cdot - k), k \in \mathbb{Z} \} \) is a Riesz sequence in \( L^2 \) with constants \( C_1, C_2 \)

2. \( C_1 \leq \left( \sum \left| \hat{\varphi}(\xi + l) \right|^2 \right)^{1/2} \leq C_2 \) for a.e. \( \xi \)

**Proof.** Computation:

\[
\left\| \sum x_k \varphi(\cdot - k) \right\|_{L^2}^2 = \left\| \hat{\varphi}(\xi) \sum x_k e^{-2\pi ik\xi} \right\|_{L^2(\xi)}^2 \\
= \int_{\mathbb{R}} |\hat{\varphi}(\xi)|^2 |\hat{x}(\xi)|^2 d\xi \\
= \int_0^1 \left( \sum_l |\hat{\varphi}(\xi + l)|^2 \right) |\hat{x}(\xi)|^2 d\xi
\]

Now showing \( 2 \implies 1 \) is easy. Given \( 2 \), we have that

\[
C_1^2 \int_0^1 |\hat{x}(\xi)|^2 d\xi \leq \int_0^1 \left( \sum_l |\hat{\varphi}(\xi + l)|^2 \right) |\hat{x}(\xi)|^2 d\xi \leq C_2^2 \int_0^1 |\hat{x}(\xi)|^2 d\xi
\]

and from Parseval, \( \int_0^1 |\hat{x}(\xi)|^2 d\xi = \|x\|^2 \), and thus using the computation we have

\[
C_1^2 \|x\|^2 \leq \sum x_k \varphi(\cdot - k) \|_{L^2}^2 \leq C_2^2 \|x\|^2
\]

which shows \( 1 \).

To show \( 1 \implies 2 \), one way is through the contrapositive. Suppose \( 2 \) does not hold. Then there is a set \( A \) of positive measure and \( \varepsilon > 0 \) for which \( \left( \sum \left| \hat{\varphi}(\xi + l) \right|^2 \right)^{1/2} \) is larger than \( C_2 + \varepsilon \), say (or smaller than \( C_1 - \varepsilon \), but the same argument will work). If we use \( \hat{x}(\xi) = 1_A \), then we have that

\[
\int_0^1 \left( \sum_l |\hat{\varphi}(\xi + l)|^2 \right) |\hat{x}(\xi)|^2 d\xi > (C_2 + \varepsilon)^2 \int_0^1 |\hat{x}(\xi)|^2 d\xi = (C_2 + \varepsilon)^2 \|x\|^2
\]

so that \( 1 \) is not satisfied.

The other way is similar, assuming \( 1 \) we let \( I \) be any interval and set \( \hat{x} = 1_I \). Then we have that

\[
C_1^2 \|x\|^2 \leq \int_I \sum |\hat{\varphi}(\xi + l)|^2 d\xi \leq C_2^2 \|x\|^2
\]

since \( \|x\|^2 = \|\hat{x}\|^2_{L^2(\mathbb{T})} = |I| \), we have that

\[
C_1^2 \leq \frac{1}{|I|} \int_I \sum |\hat{\varphi}(\xi + l)|^2 \leq C_2^2
\]

Take \( |I| \to 0 \) centered around any \( \xi \), and by Lebesgue differentiation we have that for almost every \( \xi \)

\[
C_1^2 \leq \sum |\hat{\varphi}(\xi + l)|^2 \leq C_2^2
\]
which proves (2) \[ \square \]

Let’s show that \( \{ B_d(\xi - k), k \in \mathbb{Z} \} \) is a Riesz sequence. Note that
\[
\sum_l |B_d(\xi + l)|^2 = \sum_l \left| \frac{\sin(\pi \xi)}{\pi (\xi + l)} \right|^{2(d+1)} \geq \left| \frac{\sin(\pi \xi)}{\pi \xi} \right|^{2(d+1)} \geq A > 0
\]

for \( \xi \in [-1/2, 1/2] \), and since the function is 1-periodic, the lower bound holds for all \( \xi \). Note that \( \frac{\sin x}{x} \to 1 \) as \( x \to 0 \). For the upper bound, we simply note that the sum converges absolutely, and in fact is bounded by \( 2 + \sum_{l \neq 0} \frac{1}{|\pi (l - 1/2)|^{2(d+1)}} = : B < \infty \). Thus we have a Riesz sequence.

Recall that an orthonormal sequence has a similar characterization, except with \( C_1 = C_2 = 1 \). This gives an easy way to convert any Riesz sequence into an orthonormal sequence (normalization). We set \( \varphi_d \) so that
\[
\varphi_d(\xi) := \frac{B_d(\xi)}{\left( \sum |B_d(\xi + l)|^2 \right)^{1/2}}
\]

Then we note that \( \sum_l |\varphi_d(\xi + l)|^2 = \sum_l |B_d(\xi + l)|^2 / \sum_k |B_d(\xi + k)|^2 = 1 \) a.e., and thus \( \{ \varphi_d(\xi - k), k \in \mathbb{Z} \} \) is an orthonormal system, and \( \varphi_d \) is the desired scaling function for our spline MRA.

If we set \( G(\xi) := \frac{1}{\left( \sum |B_d(\xi + l)|^2 \right)^{1/2}} = \sum \alpha_k e^{-2\pi ik\xi} \) (a 1-periodic function with Fourier series coefficients \( \alpha_k \)), then \( \tilde{\varphi_d} = B_d G \), and thus \( \varphi_d(x) = \sum \alpha_k B_d(x - k) \).

It is known that \( \alpha_k \) is an infinite sequence, that there is no smooth spline wavelet with compact support. The corresponding wavelet \( \psi_d \) will have infinite support for \( d > 0 \) (\( d = 0 \) is the Haar system, which is compactly supported). However \( \varphi_d \) and \( \psi_d \) have exponential decay, which can be shown by showing that \( \tilde{\varphi_d} \) and \( \tilde{\psi_d} \) have analytic extensions to some horizontal strip \( 0 < \text{Im } z < b \) (the length of strip corresponds to the rate of exponential decay). We will not prove this here.

**Week 9**

**Compactely Supported Wavelets**

We will approach the construction of compactly supported wavelets through \( m_0 \). Recall that \( \varphi \) satisfies
\[
\varphi(x) = \sqrt{2} \sum_k h_k \varphi(2x - k)
\]

and
\[
\hat{\varphi}(\xi) = m_0(\xi/2) \hat{\varphi}(\xi/2)
\]

where \( m_0(\xi) = \frac{1}{\sqrt{2}} \sum_k h_k e^{-2\pi ik\xi} \). Also,
\[
|m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2 = 1
\]

If \( \hat{\varphi} \) is continuous and \( \hat{\varphi}(0) \neq 0 \), then \( \hat{\varphi}(0) = m_0(0) \hat{\varphi}(0) \) and \( m_0(0) = 1 \). This implies \( m_0(1/2) = 0 \). \( m_0 \) is continuous at all \( \xi \) for which \( \hat{\varphi}(\xi) \neq 0 \).

If \( \varphi \) is compact, then \( h_k \) is necessarily finite, so that \( m_0 \) is a trigonometric polynomial.
We can iterate the formula above:

\[ \hat{\phi}(\xi) = m_0(\xi/2) m_0(\xi/4) \hat{\phi}(\xi/4) = \ldots = \prod_{j=1}^{\infty} m_0(2^{-j}\xi) \hat{\phi}(0) \]

so long as we have convergence in the infinite product. Without loss of generality \( \hat{\phi}(0) = 1 \) (in general we know \(|\hat{\phi}(0)| = 1 \) if \( \hat{\phi}(0) \neq 0 \) and \( \hat{\phi} \) is continuous at 0, Proposition 28).

**Theorem 32.** Let \( m(\xi) = \sum_{k=K}^{L} a_k e^{-2\pi i k \xi} \) be a trigonometric polynomial such that

1. \(|m(\xi)|^2 + |m(\xi + 1/2)|^2 = 1 \)
2. \( m(0) = 1 \)
3. \( m(\xi) \neq 0 \) for \(|\xi| \leq 1/4 \).

Then \( \theta(\xi) = \prod_{j=1}^{\infty} m(2^{-j}\xi) \) defines a continuous function in \( L^2 \). If \( \varphi := \theta' \), then \( \varphi [K, L] \) and \( \varphi \) defines an MRA, and the corresponding wavelet \( \psi \) is also compactly supported (support contained in \([K, L]\)).

**Proof.**

(\( \theta \) is continuous) For all \( \xi \), \( m(2^{-j}\xi) \to m(0) = 1 \), so that the infinite product makes sense. For products, we note \( \prod a_i \) converges if \( \sum |1 - a_i| \) converges (verify with logarithms/exponentiation). Note

\[ |m(2^{-j}\xi) - 1| = |m(2^{-j}(\xi)) - m(0)| \leq \|m'\| \infty 2^{-j} |\xi| \]

Since \( m \) is a trigonometric polynomial, \( \|m'\| \infty \leq C < \infty \) (and can be computed with Bernstein’s inequality). Since \( \sum_j 2^{-j} < \infty \), get convergence for all \( \xi \), and in fact the convergence is uniform for all bounded sets. This means that the limiting function \( \theta \) is continuous.

(\( \theta \in L^2 \)) Since \( |m(\xi)| \leq 1 \) from (1),

\[ |\theta(\xi)| \leq \prod_{j=1}^{N} m(2^{-j}\xi) \]

for any \( N \), where the product is \( 2^N \)-periodic, so that the inequality will only be useful up to \( 2^N \). We will estimate

\[ \int_{-2^{N-1}}^{2^{N-1}} |\theta(\xi)|^2 d\xi \leq \int_{-2^{N-1}}^{2^{N-1}} \left| \prod_{j=1}^{N} m(2^{-j}\xi) \right|^2 d\xi \]

It turns out that the right hand side is 1 for any \( N \). We will show this inductively:

First, for \( N = 1 \), we have that

\[ \int_{-1}^{1} |m(\xi/2)|^2 d\xi = \int_{0}^{2} |m(\xi/2)|^2 d\xi = \int_{0}^{1} |m(\xi/2)|^2 d\xi + \int_{0}^{1} |m(\xi/2 + 1/2)|^2 d\xi = 1 \]
Assuming that it is true for $N - 1$, we have

$$
\int_{-2^{N-1}}^{2^{N-1}} \left| \prod_{j=1}^{N} m(2^{-j}\xi) \right|^2 d\xi = \int_{0}^{2^N} \left| m(2^{-N}\xi) \right|^2 \left| \prod_{j=1}^{N-1} m(2^{-j}\xi) \right|^2 d\xi
$$

$$
= \int_{0}^{2^{N-1}} \left| m(2^{-N}\xi) \right|^2 \left| \prod_{j=1}^{N-1} m(2^{-j}\xi) \right|^2 d\xi
$$

$$
+ \int_{0}^{2^{N-1}} \left| m(2^{-N}\xi + 1/2) \right|^2 \left| \prod_{j=1}^{N-1} m(2^{-j}\xi) \right|^2 d\xi
$$

$$
= \int_{0}^{2^{N-1}} \left| \prod_{j=1}^{N-1} m(2^{-j}\xi) \right|^2 d\xi = 1
$$

noting that $\left| \prod_{j=1}^{N-1} m(2^{-j}\xi) \right|^2$ is a $2^{N-1}$-periodic function. This implies that

$$
\int_{-\infty}^{\infty} |\theta(\xi)|^2 d\xi \leq 1
$$

taking the limit as $N \to \infty$. Thus $\theta \in L^2$.

(\(\varphi = \theta^\vee \) gives rise to an MRA)

We know that $\hat{\varphi} = \theta$ is continuous, and $\hat{\varphi}(0) = 1$. If we show that \{ $\varphi ( \cdot - k), k \in \mathbb{Z}$ \} forms an orthonormal system, we are done (from Proposition 27). The translates form an orthonormal system if and only if

$$
\sum_{l} |\hat{\varphi}(\xi + l)|^2 = 1
$$

which occurs if and only if the Fourier series coefficients

$$
\int_{0}^{1} \sum_{l} |\hat{\varphi}(\xi + l)|^2 e^{-2\pi ik\xi} d\xi = \delta_k
$$

or

$$
\int_{-\infty}^{\infty} |\theta(\xi)|^2 e^{-2\pi ik\xi} d\xi = \delta_k
$$

The claim is that

$$
\int_{-2^{N-1}}^{2^{N-1}} |\theta_N(\xi)|^2 e^{-2\pi ik\xi} d\xi = \delta_k
$$

for all $N$ and $k \in \mathbb{Z}$. The proof is identical to the inductive proof earlier that

$$
\int_{-2^{N-1}}^{2^{N-1}} |\theta_N(\xi)|^2 d\xi = 1.
$$

Note that $1_{[-2^{N-1}, 2^{N-1}]} \theta_N e^{-2\pi ik\xi} \to \theta(\xi) e^{-2\pi ik\xi}$ pointwise for all $k$, so that if we can find a square-integrable function $\gamma$ for which $|1_{[-2^{N-1}, 2^{N-1}]} \theta_N| \leq |\gamma|$ for all $\xi$, then by dominated convergence theorem

$$
\int_{-\infty}^{\infty} |\theta(\xi)|^2 e^{-2\pi ik\xi} d\xi = \lim_{N \to \infty} \int_{-2^{N-1}}^{2^{N-1}} |\theta_N(\xi)|^2 e^{-2\pi ik\xi} d\xi = 1
$$

Note that

$$
\theta_N(\xi) = \prod_{j=1}^{N} m(2^{-j}\xi) = \frac{\prod_{j=1}^{\infty} m(2^{-j}\xi)}{\prod_{k=1}^{\infty} m(2^{-j2^{-N}\xi})} = \frac{\theta(\xi)}{\theta(2^{-N}\xi)}
$$

Thus we just need to show that $\theta(2^{-N}\xi)$ has a lower bound in $[-2^{N-1}, 2^{N-1}]$. Let $\omega = 2^{-N}\xi$, then we wish to show that $|\theta(\omega)| \geq C > 0$ for $|\omega| \leq \frac{1}{2}$. We will use (3) here.

$$
\theta(\omega) = \prod_{j=1}^{\infty} m(2^{-j}\omega) = m(\omega/2) m(\omega/4) \ldots
$$
We are interested in $|\omega| \leq \frac{1}{2}$. We know that $m(\xi) \neq 0$ for $|\xi| \leq 1/4$, and by continuity on a compact set there exists $c_0$ for which $|m(\xi)| \geq c_0 > 0$ for $|\xi| \leq 1/4$. This means all terms have the bound $|m(\omega/2^j)| \geq c_0$ if $|\omega| \leq 1/2$. We will use this bound for the first few terms, and rely on convergence for the rest. Due to uniform convergence on bounded sets of the infinite product, we note that there is an integer $M$ such that

$$\prod_{j=M}^{\infty} |m(2^{-j} \omega)| \geq \frac{1}{2} \text{ for all } |\omega| \leq 1/2$$

This implies that

$$\left| \prod_{j=1}^{\infty} m(2^{-j} \omega) \right| \geq \frac{1}{2} e^{-1} \text{ for all } |\omega| \leq 1/2$$

and thus $|\theta_N(2^{-j} \omega)| \leq \frac{2}{c_0} |\theta(\xi)|$ for all $|\xi| \leq 2^{N-1}$, and thus dominated convergence applies.

($\varphi$ is supported in $[K, L]$)

**Claim:** Without loss of generality we may assume that $K = 0$. If we write

$$m(\xi) = e^{-2\pi i K} \sum_{k=-K}^{L-K} a_k e^{-2\pi i k \xi} = e^{-2\pi i K} \hat{m}(\xi)$$

we have

$$\theta(\xi) = \prod_{j=1}^{\infty} e^{-2\pi i K \xi} \prod_{j=1}^{\infty} \hat{m}(2^{-j} \xi) = e^{-2\pi i K} \prod_{j=1}^{\infty} \hat{m}(2^{-j} \xi)$$

now let $\hat{\theta} = \prod_{j=1}^{\infty} \hat{m}(2^{-j} \xi)$. We have that $\theta^\vee = \hat{\theta} \cdot (\cdot - K)$. So supp $\hat{\theta} \subset [K, L]$ if and only if supp $\theta^\vee \subset [0, L - K]$. Thus if we prove the result for $K = 0$ we can prove the result for general $K$ by shifting.

$\theta_N(\xi) = \prod_{j=1}^{N} m(2^{-j} \xi)$. Let $\mu$ be the measure $\mu = \sum_{k=0}^{L} a_k \delta_k$ so that $\hat{\mu} = m(\xi)$. Note supp $\mu = \{0, 1, \ldots, L \}$. Define $\mu_j = \sum_{k=2^{-j}}^{L} a_k \delta_k$, so that $\hat{\mu_j} = m(2^{-j} \xi)$. Note supp $\hat{\mu_j} = \{0, 2^{-j}, \ldots, L 2^{-j} \}$. We have that

$$\theta_N^\vee = \mu_1 \ast \ldots \ast \mu_N$$

Note that the support of a convolution is the Minkowski sum of the supports, so

$$\text{supp}(\theta_N^\vee) = \text{supp}(\mu_1) + \text{supp}(\mu_2) + \ldots + \text{supp}(\mu_N)$$

with $\text{supp}(\mu_1) \subseteq [0, L/2]$, $\text{supp}(\mu_2) \subseteq [0, L/4]$ and $\text{supp}(\mu_k) \subseteq [0, L/2^k]$. This means that $\text{supp}(\theta_N^\vee) \subseteq [0, L/2 + L/4 + \ldots] = [0, L]$. Since $\theta_N \rightarrow \theta$, this means that $\theta_N^\vee \rightarrow \theta^\vee = \varphi$ weakly. Then $\text{supp}(\theta_N^\vee) \subseteq [0, L]$ for all $N$ implies that $\text{supp}(\varphi) \subseteq [0, L]$.

($\psi$ is also compactly supported)

We can reconstruct a wavelet from $\varphi$ with the formula Proposition 25

$$\psi(x) = \sqrt{2} \sum_{l=-L-1}^{-K-1} \tilde{\kappa}_{L-1} \varphi(2x - l)$$
Since \(|\text{supp}(\varphi)| \leq L - K\), we have that \(|\text{supp}(\varphi(2x - l))| \leq \frac{L - K}{2}\) and thus \(|\text{supp}(\psi)| \leq L - K\): the left-most function in the sum has support \([a, a + \frac{L - K}{2}]\) and the right-most function in the sum has support \([a + \frac{L - K}{2}, a + L - K]\), so the total support length is \(L - K\).

\[\square\]

**Wish List**

We seek \(m\) with the given properties:

- \(|m(\xi)|^2 + |m(\xi + 1/2)|^2 = 1\)
- \(m(0) = 1\)
- \(m(\xi) \neq 0\) for \(|\xi| \leq 1/4\)

We can verify these properties for the Haar system already. \(m(\xi) = \frac{1}{\sqrt{2}} \sum h_k e^{-2\pi ik\xi}, h_0 = h_1 = 1,\) so \(m(\xi) = 1 + e^{-2\pi i\xi} = e^{-\pi i\xi}\cos(\pi\xi)\). Note that \(m(\xi) \neq 0\) for \(|\xi| \leq 1/2\). The support of \(\varphi\) is in \([0, 1]\) \((K = 0, L = 1)\).

Where will the smoothness of \(\varphi\) come from? We will need to show decay estimates for \(\theta\). If we have the decay estimate \(|\theta(\xi)| \lesssim |\xi|^{-r-1-\varepsilon}\), then \(\varphi \in C^r\), for instance (can use integration by parts to prove).

**Week 10**

(11/11/2010)

This time we show that for all \(r > 0\), there exists a wavelet satisfying our wish list above in \(C^r\). In particular, we can arrange \(|\text{supp} \psi_r| \leq Cr\) for some constant \(C\).

Originally the construction was due to Daubechies. We will be doing a simpler construction.

**Remark:** Let \(V_0 = \text{span}\{\varphi(\cdot - k), k \in \mathbb{Z}\} = \text{span}\{\tilde{\varphi}(\cdot - k), k \in \mathbb{Z}\}\), i.e. \(V_0\) is generated by translates of either \(\varphi\) and \(\tilde{\varphi}\), and suppose that \(\{\varphi(\cdot - k), k \in \mathbb{Z}\}\) and \(\{\tilde{\varphi}(\cdot - k), k \in \mathbb{Z}\}\) are orthonormal bases for \(\mathbb{C}\). We have already characterized orthonormal generators previously. We know that

\[
\sum |\tilde{\varphi}(\xi + l)|^2 = 1 = \sum |\tilde{\varphi}^\wedge(\xi + l)|^2
\]

(Theorem 19) and if we write \(\tilde{\varphi} = \sum a_k \varphi(\cdot - k)\), then \(\tilde{\varphi}^\wedge(\xi) = \hat{a}(\xi) \hat{\varphi}(\xi)\) with \(|\hat{a}(\xi)| = 1\) where \(\hat{a}\) is one-periodic (Proposition 24). If \(\varphi, \tilde{\varphi}\) are both compactly supported, we will see that \(\varphi\) must be an integer translate of \(\tilde{\varphi}\).

**Proof.** Note that if \(\varphi, \tilde{\varphi}\) are compactly supported, then \(a_k\) is a finite sequence, and thus \(\hat{a}\) is a trigonometric polynomial \(\hat{a}(\xi) = \sum_{k=K}^L a_k e^{-2\pi ik\xi}\), with \(a_K \neq 0\) and \(a_L \neq 0\). Let’s write \(\hat{a}\) in the form

\[
\hat{a}(\xi) = e^{-2\pi ikK} \sum_{0}^{N} a_{k+K} e^{-2\pi ik\xi}
\]

and call \(b_k = a_{k+K}\), so that \(b_0 \neq 0\) and \(b_N \neq 0\). Then

\[
|\hat{a}(\xi)|^2 = 1 = \sum_{k,l} b_k b_l e^{-2\pi i(k-l)\xi} = \sum_{n=-N}^{N} \left( \sum_{k-l=n} b_k b_l \right) e^{-2\pi i n \xi}
\]
This means that \( \sum_{k-l=n} b_k b_l^* = \delta_n \).

Assume towards a contradiction that \( N > 0 \). Then if \( n = N \), we have that \( l = 0 \) and \( k = N \) and \( b_N b_0 = 0 \), but this contradicts the fact that neither \( b_0 \) nor \( b_N \) are zero. Thus \( N = 0 \) and \( \hat{a}(\xi) = e^{-2\pi ikK} \), and thus \( \hat{\varphi} = \varphi(\cdot - K) \)

\[ \square \]

Now we design \( m \) with the desired properties. We’ll initially design not \( m(\xi) \), but \( |m(\xi)|^2 \), and we will use the following Lemma to go back:

**Lemma 33. (Riesz)** If \( g(\xi) \) is a trigonometric polynomial, \( g(\xi) = \sum_{-K}^{K} \gamma_k e^{-2\pi ik\xi} \) is a non-negative (real) trigonometric polynomial with real coefficients \( \gamma_k \in \mathbb{R} \). Then there exists \( m(\xi) = \sum_{0}^{K} a_k e^{-2\pi ik\xi} \) such that \( a_k \in \mathbb{R} \) and \( |m(\xi)|^2 = g(\xi) \). (Actually we do not need \( \gamma_k \) to be real)

We will prove this lemma later.

A family of trigonometric polynomials \( \{g_k\}_{k \geq 0} \) that will work:

\[ g_k(\xi) := 1 - \frac{\int_0^{2\pi\xi} (\sin t)^{2k+1} dt}{\int_0^{2\pi\xi} (\sin t)^{2k+1} dt}, \quad k = 0, 1, \ldots \]

Call \( c_k := \int_0^{\pi} (\sin t)^{2k+1} dt \). Let’s see that \( \{g_k\} \) are admissible for our purposes:

**Properties:**

- \( g_k(0) = 1, g_k(1/2) = 0 \)

\[ g_k(\xi) + g_k(\xi + 1/2) = 2 - c_k^{-1} \left[ \int_0^{2\pi\xi} + \int_0^{2\pi(\xi + 1/2)} \right] (\sin t)^{2k+1} dt \]

\[ = 2 - c_k^{-1} \left[ c_k + \int_0^{2\pi\xi} + \int_\pi^{2\pi\xi+\pi} \right] (\sin t)^{2k+1} dt \]

\[ = 2 - c_k^{-1} c_k \]

\[ = 1 \]

noting that

\[ \int_\pi^{2\pi\xi+\pi} (\sin t)^{2k+1} dt = \int_0^{2\pi\xi} (\sin (t + \pi))^{2k+1} dt = -\int_0^{2\pi\xi} (\sin t)^{2k+1} dt \]

- \( 0 \leq \int_0^{2\pi\xi} (\sin t)^{2k+1} dt \leq c_k \) so that \( 0 \leq g_k \leq 1 \), and \( g_k \neq 0 \) for \( |\xi| < \frac{1}{2} \).

- If we graph \( g_k \), we will see that as \( k \) grows larger, the function becomes flatter at 0, \( \pi \), and this will contribute to decay in \( \xi \) of the corresponding infinite product.

**Proposition 34.** Applying Riesz to obtain trigonometric polynomials \( m_k \) such that \( |m_k|^2 = g_k \),

\[ \left| \prod_{j=1}^{\infty} m_k(2^{-j}\xi) \right|^2 = \prod_{j=1}^{\infty} g_k(2^{-j}\xi) \lesssim |\xi|^{-\beta k} \]

for some \( \beta > 0 \)
First, we prove the following:

**Claim:** \( g_k(\xi) = \left[ \frac{1 + \cos(2\pi \xi)}{2} \right]^{k+1} \gamma_k(\xi) = [\cos(\pi \xi)]^{2k+2} \gamma_k \) where \( \gamma_k \) is another trigonometric polynomial.

**Proof.** Let \( p_k(x) = c_k^{-1} \int_{-1}^{x} (1-u^2)^k \, du \) which is a degree \( 2k+1 \) polynomial. Set \( u = \cos t \) in the definition of \( g_k \), and we have

\[
g_k(\xi) = 1 + c_k^{-1} \int_{-1}^{\cos(2\pi \xi)} (1-u^2)^k \, du \\
= c_k^{-1} \left[ - \int_{-1}^{\xi} + \int_{\xi}^{\cos(2\pi \xi)} \right] (1-u^2)^k \, du \\
= c_k^{-1} \int_{\xi}^{\cos(2\pi \xi)} (1-u^2)^k \, du \\
= p_k(\cos(2\pi \xi))
\]

Note \( p_k(-1) = 0 \) and \( p_k'(x) = c_k^{-1}(1-x^2)^k \). The derivative has a root of order \( k \) at \(-1\), and thus \( p_k \) has a root of order \( k+1 \), and \( p_k(x) = (x+1)^{k+1} q_k(x) \) for some polynomial \( q_k \). We conclude that

\[
g_k(\xi) = (1 + \cos(2\pi \xi))^{k+1} q_k(\cos(2\pi \xi))
\]

and we take \( \gamma_k(\xi) = 2^{-k-1} q_k(\cos(2\pi \xi)) \). \( \square \)

**Remark:**

\[
\prod_{j=1}^{\infty} \cos(2^{-j} \xi) = \prod_{j=1}^{\infty} \sin(2^{-j+2}\pi \xi) = \lim_{m \to \infty} \frac{\sin(\pi \xi)}{2^m \sin(2^{-m}\pi \xi)} = \frac{\sin(\pi \xi)}{\pi}.
\]

Roughly speaking, each \( \cos \pi \xi \) factor buys us \( \frac{1}{|\xi|} \) decay for the infinite product \( \prod g_k(2^{-j} \xi) \).

**Lemma 35.** Let \( M_k(\xi) := \frac{g_k(\xi)}{[\cos(\pi \xi)]^k} = [\cos(\pi \xi)]^{k+2} \gamma_k(\xi) \). Then there exits \( \alpha < 1 \) such that \( \|M_k\|_\infty \lesssim 2^{\alpha k} \)

**Proof.** Since \( g_k(\xi) = p_k(\cos 2\pi \xi) \),

\[
\|M_k\|_\infty = \sup_{-1 \leq x \leq 1} p_k(x) \left( 1 + \frac{x}{2} \right)^{-k/2} \quad \text{(take } x = \cos 2\pi \xi \text{)}
\]

Then

\[
(1+x)^{-k/2} p_k(x) = c_k^{-1} \int_{-1}^{x} \left( 1-u^2 \right)^k \, du \\
= c_k^{-1} \int_{-1}^{x} \left[ \frac{1+u}{1+x} \right]^{k/2} (1+u)^{k/2}(1-u)^k \, du \\
\leq c_k^{-1} \int_{-1}^{x} \left[ (1+u)(1-u)^2 \right]^{k/2} \, du \\
\leq 2c_k^{-1} \left( \sup_{|u|\leq 1} (1+u)(1-u)^2 \right)^{k/2} \\
\leq 2c_k^{-1} 2^{k/2}
\]

42
where we used that \((1 + u)(1 - u^2) = (1 - u^2)(1 - u) \leq 2\) for \(|u| \leq 1\). Now we just need a lower bound for \(c_k\).

\[
c_k = \int_0^\pi (\sin t)^{2k + 1} dt
= \int_{-\pi/2}^{\pi/2} (\cos t)^{2k + 1} dt
\geq \int_{|t| \leq \frac{1}{\sqrt{2k + 1}}} \left(1 - \frac{t^2}{2}\right)^{2k + 1} dt
\geq \left[1 - \frac{1}{2(2k + 1)}\right]^{2k + 1} \frac{1}{\sqrt{k}}
\geq \frac{C_0}{\sqrt{k}}
\]

and this implies that

\[
\|M_k\|_\infty \leq 2 \sqrt{\frac{k}{C_0}} \frac{2^{k/2}}{2}
\]

Note that \(\cos t \geq 1 - \frac{t^2}{2}\) holds when \(|t| \leq \frac{\sqrt{6}}{6!} \geq \ldots\), or when \(|t|^2 \leq 30\) (alternating series estimates), and in particular on the set \(|t| \leq \frac{1}{\sqrt{2k + 1}}\).

Thus we have shown the result for \(\alpha = 1/2\).

\[\square\]

Now we prove the estimate.

**Proof. (of Proposition 34)** \(g_k(\xi) = M_k(\xi) |\cos \pi \xi|^k\). Then

\[
\prod_{j=1}^\infty g_k(2^{-j}\xi) = \left(\frac{\sin \pi \xi}{\pi \xi}\right)^k \prod_{j=1}^\infty M_k(2^{-j}\xi)
\]

We note that the infinite product converges uniformly on bounded sets, since

\[
|M_k(2^{-j}\xi) - 1| = |M_k(2^{-j}\xi) - M(0)| \leq C_k 2^{-j}\xi
\]

Note \(M_k(0) = \frac{g_k(0)}{\cos(0)^k} = 1\). The sup-norm upper bound we derived earlier is only good for finitely many terms, so we will obtain the estimate by breaking the product into two parts:

\[
\prod_{j=1}^{\infty} M_k(2^{-j}\xi) = \prod_{j=1}^j M_k(2^{-j}\xi) \prod_{j+1}^\infty M_k(2^{-j}\xi)
\]

We know that for \(|\xi| \leq 1\), \(\prod_{j=1}^{\infty} M_k(2^{-j}\xi) \leq C'_k\), since we have uniform convergence of the infinite product, and thus the limiting function is continuous, and thus bounded on the compact set \(|\xi| \leq 1\). This implies that

\[
\prod_{j=1}^{\infty} M_k(2^{-j}\xi) \leq C'_k \text{ for } |\xi| \leq 2^j
\]

43
Now for each \( \xi \in \mathbb{R} \), there is a unique \( J \) such that \( 2^{J - 1} \leq |\xi| \leq 2^J \). Then

\[
\prod_{j=1}^{\infty} M_k(2^{-j} \xi) \leq \|M_k\|_{\infty} C'_k \leq 2^{\alpha Jk} C'_k \leq |2\xi|^\alpha C'_k
\]

This implies that

\[
\prod_{j=1}^{\infty} g_k(2^{-j} \xi) \leq C'_k |2|\xi|^{\alpha k} \frac{1}{|\pi \xi|^k} \sim \tilde{C}_k |\xi|^{-\beta k}
\]

with \( \alpha < 1 \), and \( -\beta = \alpha - 1 \), so \( \beta > 0 \). (using the previous lemma we get \( \beta = 1/2 \)) □

From this, in the context of Theorem 32, we note that \( |\theta(\xi)| \lesssim_k |\xi|^{-\beta k/2} \) so that \( \varphi \in C^{\beta k/2 - 2} \), so we end up with a compactly supported wavelet with arbitrarily high smoothness. The one missing piece is the the proof of the Riesz lemma:

**Proof. (of Lemma 33)** From http://people.virginia.edu/~jlr5m/Papers/FejerRiesz.pdf

Letting \( w(z) = \sum_{-K}^K \gamma_k z^k \) so that \( w(e^{-2\pi i \xi}) = g(\xi) \), the goal is to find a polynomial \( p(z) \) such that \( p(z) g(z) = w(z) \) for \( |z| = 1 \). We note that \( \gamma_{-k} = \bar{\gamma}_k \) since \( g \) is real, and thus by inspection we have that \( w(1/z) = \overline{w(z)} \) for all \( z \neq 0 \). In particular, the zeros of \( w \) that are not on the unit circle occur in pairs \( (\alpha_j, 1/\bar{\alpha}_j) \) having equal multiplicity. Also, the zeros of \( w \) on the unit circle must have even multiplicity, because the symmetry properties shows that the map \( z \mapsto w(z) \) is locally 2 (or some other even number) to 1 near such zeros.

Now consider \( q(z) = z^K w(z) \), which is now a polynomial of degree \( 2K \) with \( q(0) = \gamma_{-K} \). Without loss of generality we may assume \( \gamma_{-K} \neq 0 \) by working with the conjugate of \( g \) if necessary. Noting the above, we can factor \( q \) as

\[
q(z) = \prod_{j=1}^{K} (z - \alpha_j)(z - 1/\bar{\alpha}_j) = C \prod_{j=1}^{K} (z - \alpha_j)z(z^{-1} - \bar{\alpha}_j)
\]

so that

\[
w(z) = C \prod_{j=1}^{K} (z - \alpha_j)(z^{-1} - \bar{\alpha}_j)
\]

Then we just take \( p(\xi) = \sqrt{C} \prod_{j=1}^{K} (e^{-2\pi i \xi} - \alpha_j) \). (If in addition \( \gamma_k \) are all real, we note the additional relation \( w(1/z) = w(z) \) which then forces the zero-pairs \( (\alpha, 1/\alpha) \) to be real) □

**Week 11 (11/24/2010)**

**Vanishing Moments, Again**

From the formula of Theorem 23 we note that a 0 of order \( L \) at \( m(1/2) \) gives a 0 of order \( L \) at \( \hat{\psi}(0) \), so we have \( L - 1 \) vanishing moments for \( \psi \). Here we consider a set of conditions which is sufficient for having vanishing moments (note that it does not assume that \( \psi \) is a wavelet)
Theorem 36. Suppose \( \psi \) satisfies the following properties:

- \( \{ \psi_{j,k} \} \) is an orthonormal system for \( j \)
- \( \psi \in C^L(\mathbb{R}) \) and \( \psi^{(l)} \in L^\infty \) for \( l = 0, \ldots, L \)
- \( |\psi(x)| \lesssim \frac{1}{(1+|x|)^A} \) for \( A > L + 1 \)

Then \( \int x^l \psi(x) \, dx = 0 \) for \( l = 0, \ldots, L \).

**Proof.** The idea is that we know \( \langle \psi, \psi_{j,k} \rangle = 0 \) for all \( j, k \neq (0,0) \), and locally \( \psi \) is well-approximated by a polynomial.

Let \( l_0 \) be the smallest \( l \) such that \( \int x^l \psi(x) \, dx \neq 0 \). If \( l_0 > L \), there is nothing to prove, so we assume \( l_0 \leq L \) towards a contradiction. \( \psi^{(l_0)} \) cannot vanish identically since \( \psi \) is not a polynomial (boundedness condition). Thus, there exists some \( a \) such that \( \psi^{(l_0)}(a) \neq 0 \), and without loss of generality we may assume \( a = k 2^{-J} \) for \( J \geq 0 \) and \( J, K \in \mathbb{Z} \). Consider the Taylor formula for \( \psi \) at \( a \):

\[
\psi(x) = \sum_{l=0}^{l_0} \frac{\psi^{(l)}(a)}{l!} (x-a)^l + R(x)
\]

where \( R(x) = o(|x-a|^{l_0}) \) as \( x \to a \). \( \frac{R(x)}{|x-a|^{l_0}} \) is a bounded function, noting the form

\[
R(x) = \int_a^x [\psi^{(l_0)}(t) - \psi^{(l_0)}(a)] \frac{(t-a)^{l_0-1}}{(l_0-1)!} \, dt
\]

and the assumption that \( \psi^{(l_0)} \in L^\infty \).

Let \( c_l = \frac{\psi^{(l)}(a)}{l!} \) be the Taylor polynomial coefficients. We know \( c_{l_0} \neq 0 \).

Consider \( \langle \psi, \psi_{j,k} \rangle = 0 \), i.e. \( \int \psi(x) \overline{\psi(2^j x - k)} \, dx = 0 \). Let \( k_j = 2^j a \) for \( j \geq J \) so that \( k_j = K 2^j - J \in \mathbb{Z} \). Then we have that

\[
\int \psi(x) \overline{\psi(2^j (x-a))} \, dx = 0
\]

for all \( j \geq J \). Putting in the Taylor formula, we have

\[
\int \left[ \sum_{l=0}^{l_0} c_l (x-a)^l + R(x) \right] \overline{\psi(2^j (x-a))} \, dx = 0
\]

Since \( \int x^l \psi(x) \, dx = 0 \) for all \( l < l_0 \), we’re left with

\[
\int \left[ c_{l_0} (x-a)^{l_0} + R(x) \right] \overline{\psi(2^j (x-a))} \, dx = 0
\]

We can rearrange the above as

\[
\int x^{l_0} \psi(x) \, dx = 2^{j(l_0+1)} \int x^{l_0} \psi(2^j x) \, dx = -\frac{2^{j(l_0+1)}}{c_{l_0}} \int R(x) \overline{\psi(2^j (x-a))} \, dx
\]

and we will show that the RHS tends to 0 as \( j \to \infty \).
Using the assumptions, we have

\[
|\text{RHS}| \leq \frac{2^{j(l_0+1)}}{c_{l_0}} \int \left| R(x) \right| \frac{|x-a|^{l_0}}{(1+2^j|x-a|)^d} \, dx
\]

\[
= \frac{2^{j(l_0+1)}}{c_{l_0}} \int \left| R(a+2^{-j}u) \right| \frac{|2^{-j}u|^{l_0}}{(1+|u|)^d} \, 2^{-j} \, du
\]

\[
= \frac{1}{c_{l_0}} \int \left| R(a+2^{-j}u) \right| \frac{|u|^{l_0}}{(1+|u|)^d} \, du
\]

\[
(x = a + 2^{-j}u)
\]

As \( j \to \infty \), pointwise the integrand goes to 0 \( (R(x) = o(|x-a|^{l_0}) \text{ near } a) \). The integrand is also bounded by

\[
\left\| \frac{R(\cdot)}{|x-a|^{l_0}} \right\|_{L^1} \left\| \frac{|u|^{l_0}}{(1+|u|)^d} \right\|_{L^1} \in L^1
\]

and so dominated convergence holds, and we have that \( \int x^l \psi(x) \, dx = 0 \), which contradicts the definition of \( l_0 \).

\[\square\]

**Decay of wavelet coefficients from vanishing moments**

**Theorem 37.** Let \( \psi \) be such that \( \int x^l \psi(x) \, dx = 0 \) for \( l = 0, \ldots, L \). and \( \int (1 + |x|)^{L+1} \psi(x) \, dx < \infty \). Then if \( f \in C^L \) and \( f^{(L)} \in \text{Lip}_\alpha \) for \( 0 < \alpha < 1 \), then

\[|\langle f, \psi_{j,k} \rangle| \leq 2^{-j(L+\alpha+1/2)}\]

**Remarks:** Above, the \( 1/2 \) is an artifact from \( l_2 \) normalization. We proved this for \( L = 0 \) previously (Theorem 30). Also, this is an if and only if statement, which will only partly show later.

**Proof.** By Taylor’s formula, for each \( a \), there exists a polynomial \( p_{a,L} \) of degree \( L \) such that

\[|f(x) - p_{a,L}(x)| \leq C |x-a|^{L+\alpha}\]

This is from the integral form of Taylor remainder:

\[
\left| \int_a^x \left[ f^{(L)}(t) - f^{(L)}(a) \right] \frac{(t-a)^{L-1}}{(L-1)!} \, dt \right| \leq |x-a|^{\alpha} \frac{|x-a|^L}{L!}
\]

Now

\[
\langle f, \psi_{j,k} \rangle = \int f(x) 2^{j/2} \psi(2^{-j}x-k) \, dx
\]

(vanishing moments)

\[
(a = 2^{-j}k) \leq C |x-2^{-j}k|^{L+\alpha} 2^{j/2} |\psi(2^{-j}x-k)| \, dx
\]

\[
(u = 2^{j}x-k) \leq 2^{-j(L+\alpha+1/2)} \int C |u|^{L+\alpha} |\psi(u)| \, du
\]

where \( C' \) is a constant that does not depend on \( j \) (just \( f, \psi \)).

\[\square\]
Lipschitz regularity from decay of wavelet coefficients

0 < α < 1. Let φ, ψ ∈ C^1 be compactly supported, ψ is the wavelet corresponding to the MRA generated by φ.

**Theorem 38.** For f ∈ L^2, suppose |⟨f, ψ_j,k⟩| ≤ 2^{-j(α+1/2)} for all j, k ∈ ℤ. Then f ∈ Lip_α.

**Proof.** Decompose f = P_0f + ∑_{j=0}^{∞} Q_j f,

\[ P_0 f = \sum_k \langle f, \varphi_{0,k} \rangle \varphi_{0,k} \]
\[ Q_j f = \sum_k \langle f, \psi_{j,k} \rangle \psi_{j,k} \]

where the equations above are valid in L^2 as well as pointwise (compactly supported wavelets, so only finitely many terms in each sum). Thus P_0f, Q_jf ∈ C^1.

\[ |Q_j f(x)| \leq \left( \sup_k |\langle f, \psi_j,k \rangle| \right) 2^{j/2} \sum_k |\psi(2^j x - k)| \]

Note \( \sum_k |\psi(2^j x - k)| \) has finitely many terms, and the number of terms does not depend on j. Thus

\[ \|Q_j f\|_\infty \lesssim 2^{-j\alpha} \]

In particular, \( \sum_k Q_j f \) is uniformly convergent, and since each Q_jf ∈ C^1, the limit \( \sum_k Q_j f \) is continuous. Now showing that \( |f(x) - f(y)| \lesssim |x - y|^\alpha \) boils down to showing that

\[ \left| \sum_j Q_j f(x) - Q_j f(y) \right| \lesssim |x - y|^\alpha \]

noting that P_jf ∈ C^1 already (an thus in Lip_α). In fact we will show that

\[ \sum_j |Q_j f(x) - Q_j f(y)| \lesssim |x - y|^\alpha \]
\[ Q_j f(x) - Q_j f(y) = \sum_k \langle f, \psi_{j,k} \rangle (\psi_{j,k}(x) - \psi_{j,k}(y)) \]
\[ |Q_j f(x) - Q_j f(y)| \lesssim 2^{-j\alpha} \sum_k |\psi(2^j x - k) - \psi(2^j y - k)| \]
(mean value theorem, t between x and y)
\[ \leq 2^{j(1-\alpha)}|x - y| \sum_k |\psi'(2^j t - k)| \]
\[ \lesssim 2^{j(1-\alpha)}|x - y| \]

noting again that \( \sum_k |\psi'(2^j t - k)| \) has finitely many nonzero terms. This estimate is only useful for small j.

For large j, we can use the simpler estimate \( \|Q_j f\|_\infty \lesssim 2^{-j\alpha} \) so that

\[ |Q_j f(x) - Q_j f(y)| \lesssim 2^{-j\alpha} \]

47
Then we can split the sum into two parts and use the two estimates:

\[
\sum_{j=0}^{\infty} |Q_j f(x) - Q_j f(y)| \lesssim \sum_{2^j \leq \frac{1}{|x-y|}} |Q_j f(x) - Q_j f(y)| + \sum_{2^j > \frac{1}{|x-y|}} |Q_j f(x) - Q_j f(y)|
\]
\[
\lesssim |x-y|(|x-y|^{-1})^{1-\alpha} + (|x-y|^{-1})^{-\alpha}
\]
\[
\lesssim |x-y|^{\alpha}
\]

Here we are using that the sum of exponentially controlled increasing/decreasing sequence is controlled by the largest term.

\[\square\]

**Week 12**

\[12/1/2010\]

**L\(p\) convergence of wavelet series**

Here we address conditions for when \(P_j f \to f\) in \(L\(p\)\). We already analyzed this for \(p = 2\) and touched on \(p = 1\) case in Week 3.

**Definition 39.** We say that \(\varphi\) is **majorized** by a function \(\mu\) if \(|\varphi(x)| \leq \mu(|x|)\) where \(\mu: \mathbb{R}^+ \to \mathbb{R}^+\) is monotone decreasing and \(\mu \in L^1\)

**Proposition 40.** Let \(\varphi\) be majorized by \(\mu\) and suppose \(\{\varphi(\cdot - k), k \in \mathbb{Z}\}\) is an orthonormal system. Then there exists constants \(C_1, C_2 > 0\) (depending on \(\mu\)) such that

\[
C_1 \|a\|_{l^p} \leq \left\| \sum_k a_k \varphi(\cdot - k) \right\|_{L^p} \leq C_2 \|a\|_{l^p}
\]

for \(1 \leq p \leq \infty\).

**Proof.** Let \(a \in l^p\).

For \(p = \infty\), we have the following upper bound:

\[
\left\| \sum a_k \varphi(\cdot - k) \right\|_{L^\infty} \leq \|a\|_\infty \sup_{x \in [0, 1]} \left( \sum_k |\varphi(x - k)| \right)_{1\text{-periodic}}
\]
\[
\leq \|a\|_\infty \left( \mu(0) + \sum_k \mu(k) \right)
\]

where we have used that \(\varphi\) is majorized by \(\mu\) and that \(\mu\) is monotone in the second line. We can then take \(C_2 = 2 \sum_k \mu(k)\) which is finite since \(\mu \in L^1\).

For the lower bound, let \(f = \sum_k a_k \varphi(\cdot - k)\). Then \(a_k = \langle f, \varphi(\cdot - k) \rangle\), so that

\[
|a_k| \leq \|f\|_\infty \|\varphi\|_1
\]

Note \(\|\varphi\|_1 \leq 2\|\mu\|_{L^1(\mathbb{R}^+)} \leq 2 \sum_{k=0}^{\infty} \mu(k) = C_2(\mu)\) (monotonicity of \(\mu\), so

\[
|a_k| \leq C_2(\mu) \|f\|_\infty
\]

Taking supremum over \(k\) gives the lower bound.
Proof. (of (1)). Let \( f \) be a continuous function. The goal is to show that \( \mathbb{P}f \rightarrow f \) in \( L^p \) as \( j \rightarrow \infty \) for all \( f \in L^p \) if \( 1 \leq p < \infty \) and for bounded, uniformly continuous \( f \) if \( p = \infty \).

**Proposition 41.** As before, assume \( \varphi \) is majorized by \( \mu \in L^1(\mathbb{R}^+) \). Then

1. \( \|P_0\|_{p \rightarrow p} \leq C^2_\mu \) for all \( 1 \leq p \leq \infty \)
2. \( \|P_j\|_{p \rightarrow p} = \|P_0\|_{p \rightarrow p} \) for all \( j \)

**Proof.** (of (1)). Let \( f \in L^p \cap L^2 \). Then since \( P_0f = \sum_k \langle f, \psi_{0,k} \rangle \varphi_{0,k} \), denoting \( a_k = \langle f, \psi_{0,k} \rangle \) we have

\[
\|P_0f\|_{L^p} \leq \|a\|_{L^p} \leq C^2_\mu \|f\|_{L^p}
\]

(of (2)).

\[
P_jf(x) = \sum_{j,k} 2^j \int f(y) \overline{\varphi(2^jy-k)} dy \varphi(2^jx-k)
\]

Note \( 2^j \int f(y) \overline{\varphi(2^jy-k)} dy = \int f(2^{-j}u) \varphi(u-k) du \) (\( u = 2^jy \)), so

\[
[P_jf(2^j \cdot)](2^{-j}x) = \sum_{j,k} \left( \int f(y) \overline{\varphi(u-k)} du \right) \varphi(x-k) = [P_0f](x)
\]
This means that with change of variables,

$$\int |(P_j f(2^j \cdot))(x)|^p \, dx = \int |P_0 f(2^j x)|^p \, dx$$

$$= 2^{-j} \|P_0 f\|_p^p$$

$$\leq \|P_0\|_{p \to p} 2^{-j} \|f\|_p^p$$

$$= \|P_0\|_{p \to p} \|f(2^j \cdot)\|_p^p$$

so that \(\|P_j\|_{p \to p} = \|P_0\|_{p \to p}\)

\(\square\)

Up to now, we have not used the full properties of MRA; we’ve used orthogonality and the definition of \(P_j\).

**Theorem 42.** Suppose \(\phi\) is majorized by \(\mu\) and gives rise to an MRA. Then \(P_j f \to f\) in \(L^p\) if \(1 \leq p < \infty\) and if \(p = \infty\), it holds if \(f\) is uniformly continuous and bounded.

**Proof.** First we write \(P_j f\) as a convolution operator:

$$P_j f(x) = \sum_{k} \int f(y) \overline{\phi_{j,k}(y)} \, dy \phi_{j,k}(x) = \int \left[ \sum_{k} \phi_{j,k}(x) \overline{\phi_{j,k}(y)} \right] f(y) \, dy = \int K_j(x, y) f(y) \, dy$$

where \(K_j(x, y) = 2^j K_0(2^j x, 2^j y), K_0(x, y) = \sum_k \phi(x-k) \overline{\phi(y-k)}\).

Next we prove the following **Claim:** \(|K_0(x, y)| \lesssim C \mu \left( \frac{|x-y|}{2} \right)\)

$$|K_0(x, y)| \leq \sum_k |\phi(x-k)||\phi(y-k)|$$

$$\leq \sum_k \mu(|x-k|)\mu(|y-k|)$$

(Monotonicity) \(\leq \frac{1}{2} \mu \left( \frac{|x-y|}{2} \right) \sum_k [\mu(|x-k|) + \mu(|y-k|)]$$

\[\leq 2C \mu \left( \frac{|x-y|}{2} \right)\]

Continuing on, noting that

$$\int K_j(x, y) \, dy = \sum_k |\phi(x-k)| = 1$$

(property of MRA, \(P_0(1_R) = 1_R\)), we have

\[(P_j f - f)(x) = \int K_j(x, y) f(y) \, dy - f(x)\]

$$= \int K_j(x, y) [f(y) - f(x)] \, dy$$

$$|P_j f - f|(x) \lesssim \int 2^j \mu \left( \frac{|x-y|}{2} \right) |f(y) - f(x)| \, dy$$

$$= \int 2^j \mu (2^j |u|)|f(\cdot - u) - f(\cdot)| \, dy$$
Then Minkowski gives
\[ \|P_j f - f\|_p \lesssim \int 2^j \mu(2^{j-1}|u|) \|f(\cdot - u) - f(\cdot)\|_p \, du \]
\[ = \mu(\frac{|t|}{2}) \|f(\cdot - 2it) - f\|_{L^p} \, dt \]
Now \( \|f(\cdot - 2it) - f\|_{L^p} \to 0 \) as \( j \to \infty \) (continuity of translation operator in \( L^p \), \( p < \infty \) and for uniformly continuous functions when \( p = \infty \)). Then dominated convergence applies since integrand is bounded by \( 2\|f\|_{L^p} \mu(\cdot) \) which is integrable, and we have the result.

\[ \square \]

**Remark:** If we consider \( f = 1_{[0, \sqrt{2}]} \) with the Haar system \((\varphi, \psi)\), then \( P_j f \) always has some interval containing \( \sqrt{2} \) where \( P_j f \) is constant and \( f \) takes values 0 and 1 so that \( \|P_j f - f\|_\infty \geq 1/2 \) for all \( j \). Thus we cannot expect the result to be true for \( p = \infty \) in general.

The above result allows us to write \( f = P_0 f + \sum_{j=0}^\infty Q_j f \) with convergence in \( L^p \). For the full expansion, we need \( P_j f \to 0 \) as \( j \to -\infty \). The proof of this result is similar to the proof of the result in the Haar system.

**Theorem 43.** \( P_j f \to 0 \) as \( j \to -\infty \) if \( 1 < p < \infty \). The result is not true if \( p = 1, \infty \).

**Proof.** Exactly the same proof as Theorem 11 (Week 3)

For \( 1 < p < \infty \), we can write
\[ f = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k} = \sum_j Q_j f \]
with convergence in \( L^p \).

For \( p = 1, \infty \), we can just do \( f = \sum_k \langle f, \varphi_{0,k} \rangle \varphi_{0,k} + \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi_{j,k} = P_0 f + \sum_{j\geq 0} Q_j f \).

**Approximation by wavelet series**

We showed earlier that \( f \in \text{Lip}_\alpha \iff |\langle f, \psi_{j,k} \rangle| \lesssim_{f, \psi} 2^{-j(\alpha + 1/2)} \) for all \( j \in \mathbb{Z} \) (Theorem 37), and it can be seen from the proof that \( |\langle f, \psi_{j,k} \rangle| \leq \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x| - |y|} \lesssim \sup_{j,k} 2^{j(\alpha + 1/2)} |\langle f, \psi_{j,k} \rangle| \).

We established the relationship between smoothness and decay of wavelet coefficients. Now we establish another connection to the approximation error, the rate of approximation for functions in a smoothness space using truncated wavelet series.

For \( f \in \text{Lip}_\alpha \cap L^\infty \),
\[ \|f - P_j f\|_\infty = \left\| \sum_{j=0}^{\infty} Q_j f \right\|_\infty \leq \sum_{j=0}^{\infty} \|Q_j f\|_\infty \lesssim 2^{-j\alpha} \]

We bound \( \|Q_j f\|_\infty \) as follows: \( Q_j f = \sum 2^{j/2} \langle f, \psi_{j,k} \rangle \psi(2^j x - k) \), so that
\[ \|Q_j f\|_\infty \leq C_\mu 2^{j/2} \sup_k |\langle f, \psi_{j,k} \rangle| \lesssim 2^{-j(\alpha + 1/2)} \]
and thus \( \|Q_j f\|_\infty \lesssim 2^{-j\alpha} \).

This can be reversed as well to show that if \( \|f - P_j f\|_\infty \lesssim 2^{-j\alpha} \), then \( f \in \text{Lip}_\alpha \).
In general, given a subspace $X$ of $L^p$, we may ask what is the rate of decay of $\| f - P_j f \|_p$ for $f \in X$?

**Notation:** Single bars $|f|$... will be used to denote seminorms, which will generally be components of norms, denoted in double bars $\| f \|$...

**Examples:** For the Sobolev space $X = H^s(\mathbb{R})$, $s > 0$, where $|f|_{H^s} := \int |\xi|^{2s} |\hat{f}(\xi)|^2 d\xi$ and the norm is defined as

$$\| f \|_{H^s} = \left\| \int (1 + |\xi|^{2s}) |\hat{f}(\xi)|^2 d\xi \right\|$$

(equivalent definitions up to constants)

We will characterize this space (wavelet coefficient decay) with the Shannon wavelet (will work for arbitrary wavelets with more work). Recall Shannon wavelet had $\varphi = \mathcal{F}^{-1}\{1_{[1/2,1/2]}\}$ and $\psi = \mathcal{F}^{-1}\{1_{[1/2,1]}\}$. Note $\hat{\psi}_{j,k}(\xi) = 2^{-j/2} \hat{\psi}(\xi/2^j) 2^{-2\pi i \xi/2^j}$ so $\{\hat{\psi}_{j,k}, k \in \mathbb{Z}\}$ is an orthonormal basis for $L^2(\pm [2^{j-1}, 2^j])$. Then splitting into dyadic frequencies,

$$|f|_{H^s}^2 = \sum_{j \in \mathbb{Z}} \int_{\pm [2^{j-1}, 2^j]} |\xi|^{2s} |\hat{f}(\xi)|^2 d\xi$$

$$\leq \sum_{j \in \mathbb{Z}} 2^{2js} \int_{[2^{j-1}, 2^j]} |\hat{f}(\xi)|^2 d\xi$$

$$= \sum_{j \in \mathbb{Z}} 2^{2js} \sum_k |\langle f, \psi_{j,k} \rangle|^2$$

so that

$$\| f \|_{H^s} \approx \sum_{j,k} (1 + 2^{js}) |\langle f, \psi_{j,k} \rangle|^2$$

Note for the seminorm we can rewrite this as

$$|f|_{H^s} \approx \left\| \left(2^{js} \|(d_{j,k})_k\|_2\right)\right\|_2$$

($d_{j,k} = \langle f, \psi_{j,k} \rangle$) i.e. the seminorm is equivalent to taking the 2 norm of the wavelet coefficients in the translation index, and then the 2 norm of the resulting sequence weighted with the dyadic scale

To relate to approximation error, we can also show that $f \in H^s \iff (2^{js}\|f - P_j f\|_{L^2})_{j \in \mathbb{Z}} \in l^2$ (Exercise, easy using exact formulas).

**Week 13**

(12/9/2010)

Recall that

$$f \in \text{Lip}_\alpha \iff \| f - P_j f \|_2 \lesssim 2^{-js\alpha}$$

$$\iff |\langle f, \psi_{j,k} \rangle| \lesssim 2^{-j(\alpha + 1/2)}$$

Note that before we were using $L^2$-normalized $\psi_{j,k}$. Consider instead $\psi_{j,k}^\infty = \psi(2^j x - k) = 2^{-j/2}\psi_{j,k}$. Writing $f = \sum d_{j,k}^\infty \psi_{j,k}$ using the new normalization means $d_{j,k}^\infty = d_{j,k}2^{j/2}$ so that

$$|f|_{\text{Lip}_\alpha} \approx \left\| \left(2^{js} \|(d_{j,k}^\infty)_k\|_\infty\right)\right\|_\infty$$
Compare this to what we have for $H^s$, we see the exact same expression, with $s$ instead of $\alpha$ and using a different norm. We can generalize this to define another space $B^\alpha_{p,q}$ where $\alpha$ is the degree of smoothness, $p$ is the primary space where smoothness is measured and $q$ is a secondary measure. The ultimate characterization for the norm would be something of the form

$$\|f\|_{B^\alpha_{p,q}} := \|f\|_{L^p} + |f|_{B^\alpha_{p,q}}$$

$$|f|_{B^\alpha_{p,q}} = \left\| \left( 2^{j\alpha} \left( \left\| 2^{(p)} (\lambda^{(p)})_{j,k} \right\|_{L^p} \right)_{j} \right) \right\|_q$$

where $f = \sum d_{j,k}^{(p)}\psi_{j,k}$, the $p$-normalized basis.

Back to $H^\alpha$: How does $\|f - P_j f\|_{L^2}$ behave?

$$\|f - P_j f\|_{L^2}^2 = \sum_{j=J}^{\infty} \|Q_j f\|_2^2$$

$L^2 = V_j \bigoplus_{j=J+1}^{\infty} W_j$

On one hand,

$$\sum_j 2^{2j\alpha} \|f - P_j f\|_{L^2}^2 = \sum_{j=-\infty}^{\infty} 2^{2j\alpha} \sum_{l=j}^{\infty} \|Q_l f\|_2^2$$

$$= \sum_{l=-\infty}^{\infty} \|Q_l f\|_2^2 \sum_{j=-\infty}^{l} 2^{2j\alpha}$$

$$\lesssim \sum_{j=-\infty}^{\infty} 2^{2l\alpha} \frac{\|Q_l f\|_2^2}{\sum_k |\langle f, \psi_{j,k} \rangle|^2}$$

$$\lesssim \|f\|_{H^\alpha}^2$$

We also have

$$\sum_j 2^{2j\alpha} \|f - P_j f\|_{L^2}^2 = \sum_{j=-\infty}^{\infty} 2^{2j\alpha} \sum_{l=j}^{\infty} \|Q_l f\|_2^2$$

$$\geq \sum_{j=-\infty}^{\infty} 2^{2j\alpha} \|Q_j f\|_2^2$$

$$\lesssim \|f\|_{H^\alpha}^2$$

So that

$$\|f\|_{H^\alpha}^2 \approx \left\| \left( 2^{2j\alpha} \|f - P_j f\|_{L^2} \right)_{j} \right\|_{l^2}$$

**Three Interacting Components:**

- Classical Smoothness Spaces
- Approximation Spaces (abstract, class of functions that can be approximated at a certain rate)
- Interpolation Spaces, i.e. $X = L^2$, $Y = W^{1,2}$, then get $Z = (X, Y)_{\alpha,q}$. $(L^p, W^{1,p})_{\alpha,q} = B^\alpha_{p,q}$, $0 < q < \infty, 0 < \alpha < 1$. This interpolation is achieved via the definition of a $K$-functional that combines the two norms with some weights...
Linear Approximation v Nonlinear Approximation

Let $\mathcal{H}$ be a Hilbert space, separable, $\{e_n\}_{n=1}^{\infty}$ be an orthonormal basis.

Let $E_n := \text{span}\{e_1, \ldots, e_n\}$ (e.g. trigonometric polynomials, etc). For $f \in \mathcal{H}$, we are interested in

$$\inf_{g \in E_n} \|f - g\|_\mathcal{H} = \text{dist}(f, E_n) =: E_n(f) \quad \text{(linear approximation)}$$

Let $\Sigma_n := \left\{ \sum_{k \in \Lambda} c_k e_k, |\Lambda| \leq n \right\}$. This space is closed under scalar multiplication but not addition. At best $\Sigma_n + \Sigma_n \subset \Sigma_{2n}$, and so it is not a linear space (an $n$-dimensional manifold).

For $f \in \mathcal{H}$, we are interested in

$$\text{dist}(f, \Sigma_n) = \inf_{g \in \Sigma_n} \|f - g\|_\mathcal{H} =: \sigma_n(f)$$

Exact formulas:

$\begin{align*}
f &= \sum_{k} c_k e_k, \\
E_n(f)^2 &= \sum_{k=1}^{n} |c_k|^2 \quad \text{(norm of projection)}
\end{align*}$

Letting $|c_1^*| > |c_2^*| > \ldots$ be a decreasing rearrangement. Then $\sigma_n(f)^2 = \sum_{k=1}^{n} |c_k^*|^2$.

In general, suppose we are given $X_1 \subset X_2 \subset \ldots \subset X_n \subset \ldots \subset \mathcal{H} = X$. Define

$$\mathcal{A}^\alpha(X, \{X_n\}) : = \left\{ f \in X, \inf_{g \in X_n} \|f - g\|_X \leq n^{-\alpha} \right\}$$

This is a linear space. We want to know the description of $\mathcal{A}^\alpha(\{E_n\})$ and $\mathcal{A}^\alpha(\{\Sigma_n\})$.

Define $|f|_{\mathcal{A}^\alpha(X_n)} := \sup n^\alpha \text{dist}(f, X_n)$

For linear approximation,

$$E_n(f) \lesssim n^{-\alpha} \iff E_{2^j}(f) \lesssim 2^{-j\alpha} \quad \text{(monotonicity)}$$

$$(\text{uses the fact that the sum of a geometric series is bounded by the largest term})$$

Thus $|f|_{\mathcal{A}^\alpha(E_n)} \approx \sup_j 2^{j\alpha} \left( \sum_{2^j < k < 2^{j+1}} |c_k(f)|^2 \right)^{1/2}$

For nonlinear approximation,

$$\sigma_n(f) \lesssim n^{-\alpha} \iff \left( \sum_{k=n+1}^{\infty} |c_k^*|^2 \right)^{1/2} \lesssim n^{-\alpha}$$

$$(\text{Claim}) \iff |c_k^*| \lesssim k^{-\alpha - 1/2} \text{ for all } k$$

The $\iff$ part of the claim is immediate. For $\Rightarrow$,

$$n |c_{2n}^2| \lesssim \sum_{n=1}^{2n} |c_k^2| \lesssim n^{-2\alpha} \implies |c_{2n}^*| \lesssim n^{-\alpha - 1/2}$$

Thus $|f|_{\mathcal{A}^\alpha(\Sigma_n)} \approx \sup_n n^{\alpha + 1/2} |c_n^*|$
Besov Space

The original definition of $B_{p,q}^\alpha$. For simplicity, assume $0 < \alpha < 1$.

Define $L_p$ modulus of continuity

$$\omega(f,t)_{L^p} := \sup_{|h|<t} \| f(\cdot + h) - f(\cdot) \|_{L^p}$$

We are interested in a bound like $\omega(f,t)_{L^p} \lesssim t^\alpha$. Define the seminorm

$$|f|_{B_{p,q}^\alpha} := \begin{cases} \sup_t \frac{\omega(f,t)_{L^p}}{t^{\alpha}} & q = \infty \\ \left( \int_0^1 [t^{-\alpha} \omega(f,t)]^q \frac{dt}{t} \right)^{\frac{1}{q}} & 0 < q < \infty \end{cases}$$

(The $dt/t$ term captures boundedness in the second case)

For higher $\alpha$, $r < \alpha < r + 1$, need to look at higher differences $\Delta_h f$ where $\Delta_h f = f(\cdot + h) - f(\cdot)$.

This captures fine behavior of $\omega(f,t)$.

Some cases: $B_{p,\infty}^\alpha = \text{Lip}(\alpha,L^p)$, $B_{2,2}^\alpha = H^\alpha$, $B_{\infty,\infty}^\alpha = \text{Lip}(\alpha)$

Properties:

- For $\alpha_1 > \alpha_2$, $B_{p,q_1}^{\alpha_1} \subset B_{p,q_2}^{\alpha_2}$ for all $q_1, q_2$
- For $p_1 > p_2$, $B_{p_1,q_1}^{\alpha} \subset B_{p_2,q_2}^{\alpha}$ for all $q_1, q_2$

($q$ is a secondary index)

Ex. $H^{1/2+\varepsilon} = B_{2,2}^{1/2+\varepsilon} \hookrightarrow L^\infty$.

Using Wavelet Bases: $X_1 \subset X_2 \subset \ldots$ with $X_j := V_j$. Then $A_q^\alpha(L^p, \{V_j\}) = B_{p,q}^\alpha$, and for the nonlinear case, the n-term wavelet series $\Sigma_n = \left\{ \sum_{j,k \in A} d_{j,k} \psi_{j,k} \right\} \leq n$, $A_q^\alpha(L^2, \{\Sigma_n\}) = B_{q,q}^\alpha$ where $1/q = 1/2 + \alpha$.

Comparing to the case where $p = 2$, the linear approximation space gives $B_{2,q}^\alpha$ and the nonlinear gives $B_{q,q}^\alpha$. 

55