Approximation in bases

Background reading

These notes are based on the following sources:

1. Chapters 4,7,9 of: Stephane Mallat. *A wavelet tour of signal processing: the sparse way.* Academic press, 2008.
2. Chapter 2 of: Eugenio Hernández and Guido Weiss. *A first course on wavelets.* CRC press, 1996.

Lecture 1

1 Introduction

A ubiquitous problem in engineering or mathematics is the following:

> Given a function $f(t)$ defined for $t \in \mathbb{R}$, for simplicity suppose that $f \in L^2(\mathbb{R})$, how can we transmit/store/analyse this function from finitely many values?

To give some examples:

1. f is a voice signal and we want to transmit it over a telephone line.
2. f is the cross-section of a body whose image we want to reconstruct using finitely many samples.
3. f is an image that we want to put onto a compact disk.

Suppose we have an orthonormal basis $\{g_n : n \in \mathbb{Z}\}$ in $L^2(\mathbb{R})$, then we know that

$$f = \sum_{n\in\mathbb{Z}} c_n g_n, \qquad c_n = \langle f, g_n \rangle.$$

Then, the coefficients $\{c_n\}_{n\in\mathbb{Z}}$ provides a discrete representation of f. In practice, we will choose some finite set $\Lambda \subset \mathbb{Z}$ and process only the coefficients $\{c_n\}_{n\in\Lambda}$. One would hope that

$$f \approx \sum_{n\in\Lambda} c_n g_n.$$

From classical Fourier analysis, we know that $\{\frac{1}{\sqrt{2B\pi}} e^{iB^{-1}k\cdot} : k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2([-B\pi, B\pi])$. So, given any $f \in L^2([-B\pi, B\pi])$,

$$f(x) = \frac{1}{2B\pi} \sum_{k\in\mathbb{Z}} \hat{f}(kB^{-1}) e^{ikB^{-1}x}. \tag{1}$$

Recall that, the Fourier transform of $f \in L^1(\mathbb{R})$ is defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-ix\xi} \mathrm{d}x, \qquad \xi \in \mathbb{R}$$

and this definition can be extended to $L^2(\mathbb{R})$ since $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is dense in $L^2(\mathbb{R})$.

A direct consequence of (1) is the celebrated Shannon-Nyquist-Whittaker sampling theorem:

Theorem 1. *Suppose $\hat{f}$ is piecewise smooth and continuous and $\hat{f}(\xi) = 0$ for all $|\xi| > B\pi$. Then,*

$$f(x) = \sum_{k\in\mathbb{Z}} f\left(\frac{k}{B}\right)\varphi\left(x - \frac{k}{B}\right),$$

where $\varphi(x) = \frac{\sin(\pi Bx)}{\pi Bx}$. We also have that

$$f_N = \sum_{|k|\leqslant N} f\left(\frac{k}{B}\right)\varphi\left(\cdot - \frac{k}{B}\right) \to f \qquad \text{in } L^\infty(\mathbb{R}).$$

Although the Shannon-Nyquist-Whittaker theorem provides a discrete representation of functions and describes how one may approximate f with finitely many values. However, interpolation with φ is rarely used in practice due to its slow decay. Furthermore, Fourier representations have the drawback of requiring many samples or coefficients to represent localized events. More precisely, the support of the functions $e^{ikB^{-1}\cdot}$ over the entire real line, so changing f locally will result in a change in all its coefficients $\hat{f}(kB^{-1})$.

Wavelets:

Definition 1. *We say that a function $\psi \in L^2(\mathbb{R})$ is a wavelet for $L^2(\mathbb{R})$ if*

$$\left\{\psi_{j,k} := 2^{j/2}\psi(2^j \cdot -k)\ ;\ j,k \in \mathbb{Z}\right\}$$

forms an orthonormal basis of $L^2(\mathbb{R})$.

In 1910, Haar constructed the wavelet basis (although it was not known as such), by choosing

$$\psi = \mathbb{1}_{[-1,-1/2)} - \mathbb{1}_{[-1/2,0)},$$

he showed that

$$\left\{\psi_{j,k} := 2^{j/2}\psi(2^j \cdot -k)\ ;\ j,k \in \mathbb{Z}\right\}$$

forms an orthonormal basis of $L^2(\mathbb{R})$. The basis functions are compactly supported, and large coefficients occur at sharp signal transitions (discontinuities) only.

In 1980, Strömberg discovered a piecewise linear wavelet ψ which yields better approximation properties for smooth functions. Unaware of this result, Meyer tried to prove that there does not exist a regular wavelet which generates an orthonormal basis. Instead of proving this, his attempt led to the construction of an entire family of orthonormal wavelet bases which are infinitely continuously differentiable. The work of Meyer led to a scurry of research on wavelets throughout the late 1980's and 1990's. In the following sections, we shall study the systematic approach of constructing orthonormal wavelet bases via multiresolution analysis, which was established by Meyer and Mallat.

2 Multiresolution analysis

Definition 2. *A multiresolution analysis (MRA) consists of a sequence of closed subspaces V_j of $L^2(\mathbb{R})$, with $j \in \mathbb{Z}$, satisfying the following.*

(I) $V_j \subset V_{j+1}$ *for all* $j \in \mathbb{Z}$.

(II) For all $j \in \mathbb{Z}$, $f \in V_j$ *if and only if* $f(2\cdot) \in V_{j+1}$.

(III) $\lim_{j\to-\infty} V_j = \bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

(IV) $\lim_{j\to+\infty} V_j = \overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$.

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

The function φ in (V) is called a scaling function for the MRA.

Note that condition (II) implies that $\{\varphi_{j,k}\ ;\ k \in \mathbb{Z}\}$ is an orthonormal basis for V_j.

Lecture 2

2.1 On the conditions of an MRA

In the following, let $\mathbb{T} = [-\pi, \pi)$. Recall that $\left\{\frac{1}{\sqrt{2\pi}} \exp(in\cdot)\ ;\ n \in \mathbb{Z}\right\}$ is an orthonormal basis for $L^2(\mathbb{T})$.

Theorem 2. *Recall the conditions of Definition 2. Then, (I), (II) and (V) implies (III).*

Proof. Assume (I), (II) and that $\{\varphi(\cdot - k)\ ;\ k \in \mathbb{Z}\}$ is an orthonormal basis of V_0. To prove (III), we need to show that for all $f \in L^2(\mathbb{R})$, $\|P_j f\|_{L^2} \to 0$ as $j \to -\infty$, where P_j is the orthogonal projection onto V_j.

Suppose first that $f \in L^2(\mathbb{R})$ is bounded with $\|f\|_\infty \leqslant A$ and has compact support in $[-2^J, 2^J]$. Since $\{\varphi_{j,k}\ ;\ k \in \mathbb{Z}\}$ is an orthonormal basis for V_j,

$$\begin{aligned}
\|P_j f\|_{L^2}^2 &= \sum_{k\in\mathbb{Z}} |\langle f, \varphi_{j,k}\rangle|^2 = \sum_{k\in\mathbb{Z}} \left| 2^{j/2} \int_{-2^J}^{2^J} f(t)\overline{\varphi(2^j t - k)} \mathrm{d}t \right|^2 \\
&\leqslant A^2 \sum_{k\in\mathbb{Z}} 2^j \left(\int_{-2^J}^{2^J} \left|\varphi(2^j t - k)\right| \mathrm{d}t \right)^2 \leqslant A^2 \sum_{k\in\mathbb{Z}} 2^{J+1} 2^j \int_{-2^J}^{2^J} \left|\varphi(2^j t - k)\right|^2 \mathrm{d}t \\
&= A^2 \sum_{k\in\mathbb{Z}} 2^{J+1} \int_{-2^{J+j}-k}^{2^{J+j}-k} |\varphi(x)|^2 \,\mathrm{d}x = A^2 2^{J+1} \int |\varphi(x)|^2 \mathbb{1}_{S_j}(x) \mathrm{d}x
\end{aligned}$$

where $S_j = \bigcup_{k\in\mathbb{Z}} [k - 2^{J+j}, k + 2^{J+j}]$, provided that $J + j \leqslant -1$. Since $\mathbb{1}_{S_j}(x) \to 0$ for a.e. $x \in \mathbb{R}$ as $j \to \infty$, by the dominated convergence theorem,

$$\int |\varphi(x)|^2 \mathbb{1}_{S_j}(x) \mathrm{d}x \to 0, \qquad j \to \infty.$$

Therefore, $\|P_j f\|^2 \to 0$ as $j \to \infty$. Since the set of bounded functions with compact support is dense in $L^2(\mathbb{R})$, it follows that for all $f \in L^2(\mathbb{R})$, $\|P_j f\|^2 \to 0$ as $j \to \infty$.

□

Theorem 3. *Let $\{V_j\}_{j\in\mathbb{Z}}$ be a sequence of closed subspaces of $L^2(\mathbb{R})$ satisfying (I), (II) and (V). Assume that $|\hat{\varphi}|$ is continuous at 0. Then*

$$\hat{\varphi}(0) \neq 0 \iff \overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}).$$

Moreover, if either case holds, $|\hat{\varphi}(0)| = 1$.

Proof. Assume that $\hat{\varphi}(0) \neq 0$. Let $W = \overline{\bigcup_{j\in\mathbb{Z}} V_j}$. To prove that $W = L^2(\mathbb{R})$, we simply need to show that $W^\perp = \{0\}$. For $f \in W^\perp$, let $R > 0$ and let $g \in L^2(\mathbb{R})$ be such that $\hat{g} = \hat{f}\mathbb{1}_{[-R,R]}$. Let P_j be the orthogonal projection onto V_j. By definition of f, $P_j f = 0$ for all $j \in \mathbb{N}$. Moreover,

$$\begin{aligned}\|P_j g\|^2 &= \sum_{k\in\mathbb{Z}} |\langle g,\, \varphi_{j,k}\rangle|^2 = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\Big|\int \hat{g}(\xi)\hat{\varphi}_{j,k}(\xi)\mathrm{d}\xi\Big|^2 \\ &= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\Big|2^{-j/2}\int \hat{g}(\xi)\hat{\varphi}(\frac{\xi}{2^j})e^{ik2^{-j}\xi}\mathrm{d}\xi\Big|^2 = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}}\Big|2^{-j/2}\int_{-2^j\pi}^{2^j\pi} \hat{g}(\xi)\hat{\varphi}(\frac{\xi}{2^j})e^{ik2^{-j}\xi}\mathrm{d}\xi\Big|^2\end{aligned}$$

provided that j is sufficiently large so that $2^j\pi \geqslant R$. Note that $\Big\{e_k := \frac{1}{\sqrt{2\pi 2^j}}\exp(ik\cdot/2^j)\ ;\ k\in\mathbb{Z}\Big\}$ is an orthonormal basis for $L^2([-2^j\pi, 2^j\pi])$. So,

$$\|P_j g\|^2 = \sum_{k\in\mathbb{Z}} \left|\langle \hat{g}\hat{\varphi}(2^{-j}\cdot),\, e_k\rangle\right|^2 = \int_{-R}^{R}\left|\hat{g}(\xi)\hat{\varphi}(2^{-j}\xi)\right|^2\mathrm{d}\xi.$$

Now, $\left|\hat{\varphi}(2^{-j}\xi)\right| \to |\hat{\varphi}(0)|$ uniformly on $[-R,R]$ as $j\to\infty$, so

$$\int_{-R}^{R}\left|\hat{g}(\xi)\hat{\varphi}(2^{-j}\xi)\right|^2\mathrm{d}\xi \to |\hat{\varphi}(0)|^2\int_{-R}^{R}|\hat{g}(\xi)|^2\,\mathrm{d}\xi = |\hat{\varphi}(0)|^2\int_{-R}^{R}\left|\hat{f}(\xi)\right|^2\mathrm{d}\xi,$$

as $j\to\infty$. Also,

$$\|P_j g\|^2 = \|P_j g - P_j f\|^2 \leqslant \|g-f\|^2 = \frac{1}{2\pi}\int_{|\xi|\geqslant R}\left|\hat{f}(\xi)\right|^2\mathrm{d}\xi.$$

Therefore,

$$|\hat{\varphi}(0)|^2\int_{-R}^{R}\left|\hat{f}(\xi)\right|^2\mathrm{d}\xi \leqslant \frac{1}{2\pi}\int_{|\xi|\geqslant R}\left|\hat{f}(\xi)\right|^2\mathrm{d}\xi.$$

Since R is arbitrary and $\hat{\varphi}(0)\neq 0$, it follows that $f=0$.

To prove the converse direction, let $W = \overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$. Let f be such that $\hat{f} = \mathbb{1}_{[-1,1]}$. Then,

$$\|f\|_{L^2}^2 = \frac{1}{2\pi}\|\hat{f}\|_{L^2}^2 = \frac{1}{\pi}.$$

Let P_j be the orthogonal projection onto V_j. Then, $\|P_j f - f\|_{L^2} \to 0$ as $j\to\infty$ by assumption. So, $\|P_j f\|_{L^2} \to \|f\|_{L^2}$. In particular,

$$\|P_j f\|_{L^2}^2 = \sum_{k\in\mathbb{Z}}|\langle f,\, \varphi_{j,k}\rangle|^2 \to \frac{1}{\pi}, \qquad j\to\infty.$$

By Plancherel, we have that

$$\langle f,\, \varphi_{j,k}\rangle = \frac{1}{2\pi}\int_{\mathbb{R}}\hat{f}(\xi)\overline{\hat{\varphi}_{j,k}(\xi)}\mathrm{d}\xi.$$

So,

$$\begin{aligned}\|P_j f\|_{L^2}^2 &= \frac{1}{4\pi^2}\sum_{k\in\mathbb{Z}}\left|\int_{\mathbb{R}}\hat{f}(\xi)\overline{\hat{\varphi}_{j,k}(\xi)}\mathrm{d}\xi\right|^2 = \frac{1}{4\pi^2}\sum_{k\in\mathbb{Z}}\left|\int_{-1}^{1}2^{-j/2}\overline{\hat{\varphi}(\frac{\xi}{2^j})e^{-i\xi k/2^j}}\mathrm{d}\xi\right|^2 \\ &= \frac{2^j}{4\pi^2}\sum_{k\in\mathbb{Z}}\left|\int_{-2^{-j}}^{2^{-j}}\overline{\hat{\varphi}(\xi)}e^{i\xi k}\mathrm{d}\xi\right|^2 = \frac{2^j}{2\pi}\sum_{k\in\mathbb{Z}}\left|\langle\frac{1}{\sqrt{2\pi}}e^{i\cdot k},\, \hat{\varphi}\mathbb{1}_{[-2^{-j},2^{-j}]}\rangle\right|^2,\end{aligned}$$

for j large enough such that $[-2^{-j},2^{-j}]\subset[-\pi,\pi]$. Since $\hat{\varphi}\mathbb{1}_{[-2^{-j},2^{-j}]}\in L^2(\mathbb{T})$, we have that

$$\frac{2^j}{2\pi}\int\left|\hat{\varphi}(\mu)\mathbb{1}_{[-2^{-j},2^{-j}]}(\mu)\right|^2\mathrm{d}\mu = \frac{2^j}{2\pi}\int_{-2^{-j}}^{2^{-j}}|\hat{\varphi}(\mu)|^2\,\mathrm{d}\mu \to \frac{1}{\pi}$$

as $j \to \infty$. By continuity of $|\hat{\varphi}|$ at 0, the integral expression on the LHS tends to $\frac{1}{\pi}|\hat{\varphi}(0)|$ and it thus follows that $|\hat{\varphi}(0)| = 1$. $\square$

Proposition 1. *Let $g \in L^2(\mathbb{R})$. Then, $\{g(\cdot - k)\ ;\ k \in \mathbb{Z}\}$ is an orthonormal system if and only if*

$$\sum_{k\in\mathbb{Z}} |\hat{g}(\xi + 2k\pi)|^2 = 1, \qquad a.e.\, \xi \in \mathbb{R}.$$

Proof. Let $\delta_{k,0}$ equal 0 if $k \neq 0$ and equal 1 if $k = 0$. By Plancherel,

$$\int_{\mathbb{R}} g(x)\overline{g(x-k)}\mathrm{d}x = \frac{1}{2\pi}\int \hat{g}(\xi)\overline{\hat{g}(\xi)}e^{ik\xi}\mathrm{d}\xi.$$

So,

$$\begin{aligned}
\int_{\mathbb{R}} g(x)\overline{g(x-k)}\mathrm{d}x &= \frac{1}{2\pi}\int |\hat{g}(\xi)|^2 e^{ik\xi}\mathrm{d}\xi \\
&= \frac{1}{2\pi}\sum_{n\in\mathbb{Z}}\int_{2\pi n}^{2\pi(n+1)} |\hat{g}(\xi)|^2 e^{ik\xi}\mathrm{d}\xi \\
&= \frac{1}{2\pi}\sum_{n\in\mathbb{Z}}\int_{0}^{2\pi} |\hat{g}(\xi+2\pi n)|^2 e^{ik\xi}\mathrm{d}\xi \\
&= \frac{1}{2\pi}\int_{0}^{2\pi} \left(\sum_{n\in\mathbb{Z}} |\hat{g}(\xi+2\pi n)|^2\right) e^{ik\xi}\mathrm{d}\xi.
\end{aligned}$$

Therefore,

$$\delta_{0,k} = \int_{\mathbb{R}} g(x)\overline{g(x-k)}\mathrm{d}x$$

if and only if the 2π-periodic function $\sum_{n\in\mathbb{Z}} |\hat{g}(\xi+2\pi n)|^2$ has its zeroth Fourier coefficient equal to 1 and all other Fourier coefficients equal to 0, i.e.

$$\sum_{n\in\mathbb{Z}} |\hat{g}(\xi+2\pi n)|^2 = 1, \qquad a.e..$$

$\square$

2.2 Construction of wavelets from an MRA

Let W_0 be the orthogonal complement of V_0 in V_1, so that

$$V_1 = V_0 \oplus W_0.$$

If we dilate elements in W_0 by 2^j (for $\psi \in W_0$, consider $\psi(2^j\cdot)$), we get W_j such that

$$V_{j+1} = V_j \oplus W_j, \qquad \forall j \in \mathbb{Z}.$$

Since $V_j \to \{0\}$ as $j \to -\infty$, we have that

$$V_{j+1} = V_j \oplus W_j = V_{j-1} \oplus W_{j-1} \oplus W_j = \oplus_{l=-\infty}^{j} W_l.$$

Also, since $V_j \to L^2(\mathbb{R})$ as $j \to +\infty$,

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j.$$

If we can find $\psi \in W_0$ such that $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0, then $\{\psi_{j,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_j (by (II)). This implies that

$$\{\psi_{j,k}\ ;\ j,k \in \mathbb{Z}\}$$

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

The low pass filter Note that

$$\frac{1}{2}\varphi(\frac{\cdot}{2}) \in V_{-1} \subset V_0.$$

By (V),

$$\frac{1}{2}\varphi(\frac{\cdot}{2}) = \sum_k \alpha_k \varphi(\cdot + k)$$

where

$$\alpha_k = \frac{1}{2}\int \varphi(\frac{x}{2})\overline{\varphi(x+k)}\mathrm{d}x, \qquad \sum_k |\alpha_k|^2 < \infty.$$

By applying the Fourier transform,

$$\hat{\varphi}(2\xi) = \sum_k \alpha_k \hat{\varphi}(\xi) e^{ik\xi} =: m_0(\xi)\varphi(\xi).$$

The 2π-periodic function $m_0(\xi) = \sum_k \alpha_k e^{ik\xi}$ is an element of $L^2(\mathbb{T})$ and is call the *low pass filter* of φ.

Lemma 1. *For a.e.* $\xi \in \mathbb{R}$, $|m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1$.

Proof. By Proposition 1, $\sum_{k\in\mathbb{Z}} |\hat{\varphi}(2\xi + 2k\pi)|^2 = 1$ for a.e. $\xi \in \mathbb{R}$. Since $\hat{\varphi}(2\xi) = \hat{\varphi}(\xi)m_0(\xi)$,

$$\sum_{k\in\mathbb{Z}} |\hat{\varphi}(\xi + k\pi)m_0(\xi + k\pi)|^2 = 1.$$

By splitting the sum over odd and even indices, and using the fact that m_0 is 2π-periodic,

$$\begin{aligned} 1 &= \sum_{k\in\mathbb{Z}} |\hat{\varphi}(\xi + 2k\pi)m_0(\xi)|^2 + \sum_{k\in\mathbb{Z}} |\hat{\varphi}(\xi + (2k+1)\pi)m_0(\xi+\pi)|^2 \\ &= |m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 . \end{aligned}$$

□

Theorem 4. *Let* $\{V_j\}_{j\in\mathbb{Z}}$ *be an MRA with scaling function* φ *and low pass filter* m_0. *Let* ψ *be such that*

$$\hat{\psi}(\xi) = e^{i\xi/2}\overline{m_0(\xi/2+\pi)}\hat{\varphi}(\xi/2).$$

Let $W_0 = \overline{\mathrm{Span}\,\{\psi_{0,k}\ ;\ k \in \mathbb{Z}\}}$. *Then,*

(i) $\psi \in V_1$,

(ii) $\{\psi_{0,k}\ ;\ k \in \mathbb{Z}\}$ *is an orthonormal basis of* W_0.

(iii) $V_0 \perp W_0$.

(iv) $V_1 = V_0 + W_0$.

Proof.

(i) We first show that $f \in V_0$ if and only if there exists a 2π-periodic function $g \in L^2(\mathbb{T})$ such that $\hat{f}(\xi) = g(\xi)\hat{\varphi}(\xi)$:

Let $f \in V_0$. Then, $f = \sum_k b_k \varphi_{0,k}$ for some $b \in \ell^2(\mathbb{Z})$. By applying the Fourier transform

$$\hat{f}(\xi) = \sum_{k\in\mathbb{Z}} b_k e^{-ik\xi}\hat{\varphi}(\xi),$$

where $g(\xi) := \sum_{k\in\mathbb{Z}} b_k e^{-ik\xi}$ is a 2π-periodic function in $L^2(\mathbb{T})$. Conversely, if $\hat{f}(\xi) = g(\xi)\hat{\varphi}(\xi)$ for some 2π-periodic function which is in $L^2(\mathbb{T})$, then

$$\int |\hat{\varphi}(\xi)g(\xi)|^2 \, \mathrm{d}\xi = \int_0^{2\pi} \sum_k |\hat{\varphi}(\xi + 2k\pi)|^2 \, g(\xi)\mathrm{d}\xi = \|g\|^2_{L^2(\mathbb{T})} < \infty.$$

Therefore, $\hat{\varphi}(\xi)g(\xi) \in L^2(\mathbb{R})$ and by applying the inverse Fourier transform to $\hat{f}$, we have that $f \in V_0$.

So, $\psi \in V_1$ if and only if $\psi(\cdot/2) \in V_0$, and by the derived condition, this holds if and only if $2\hat{\psi}(2\xi) = g(\xi)\hat{\varphi}(\xi)$ for some $g \in L^2(\mathbb{T})$, 2π-periodic. This is certainly true by definition of ψ.

Lecture 3

From now on, let us write $\hat{\psi} = g(\cdot/2)\hat{\varphi}(\cdot/2)$ where $g(\xi) = e^{i\xi}\overline{m_0(\xi+\pi)}$.

(ii) By the same argument as in Proposition 1, for $\hat{\psi} = g(\cdot/2)\hat{\varphi}(\cdot/2)$, then one can show that $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal system if and only if

$$|\hat{g}(\xi/2)|^2 + |\hat{g}(\xi/2+\pi)|^2 = 1 \tag{2}$$

for a.e. $\xi \in \mathbb{R}$. By Lemma 1, this is certainly true for our choice of g. Hence, $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0.

(iii) Let us establish that $V_0 \perp W_0$. This is true if and only if $\{\varphi_{0,k}\}_k$ and $\{\psi_{0,k}\}_k$ are orthogonal families, that is:

$$\langle \psi, \, \varphi_{0,n} \rangle = 0, \qquad \forall n \in \mathbb{Z}.$$

Now,

$$0 = \langle \psi, \, \varphi_{0,n} \rangle = \frac{1}{2\pi} \int \hat{\psi}(\xi)\overline{\hat{\varphi}(\xi)} e^{in\xi}\mathrm{d}\xi = \frac{1}{2\pi}\int_0^{2\pi} \sum_{k\in\mathbb{Z}} \hat{\psi}(\xi+2\pi k)\overline{\hat{\varphi}(\xi+2\pi k)} e^{in\xi}\mathrm{d}\xi$$

for all $n \in \mathbb{Z}$ if and only if

$$\sum_{k\in\mathbb{Z}} \hat{\psi}(\xi+2\pi k)\overline{\hat{\varphi}(\xi+2\pi k)} = 0 \tag{3}$$

for a.e. $\xi \in \mathbb{R}$. Recall that $\hat{\psi}(2\xi) = g(\xi)\hat{\varphi}(\xi)$ and $\hat{\varphi}(2\xi) = m_0(\xi)\hat{\varphi}(\xi)$. Therefore, (3) is equivalent to

$$\sum_{k\in\mathbb{Z}} g(\xi/2+\pi k)\overline{m_0(\xi/2+\pi k)} \, |\hat{\varphi}(\xi/2+\pi k)|^2 = 0.$$

By splitting this sum of the odd and even integers and applying Proposition 1, we get that $V_0 \perp W_0$ if and only if

$$g(\xi)\overline{m_0(\xi)} + g(\xi+\pi)\overline{m_0(\xi+\pi)} = 0 \tag{4}$$

which is true by our choice of g.

(iv) To show that $V_1 = V_0 + W_0$, it is equivalent to showing that for all $a \in \ell^2(\mathbb{Z})$, there exists $b, c \in \ell^2(\mathbb{Z})$ such that

$$2\sum_{n\in\mathbb{Z}} a_n \varphi(2t-n) = \sum_{n\in\mathbb{Z}} b_n \varphi(t-n) + \sum_{n\in\mathbb{Z}} c_n \psi(t-n).$$

By applying the Fourier transform, this equation becomes

$$\begin{aligned}\underbrace{\sum_{n\in\mathbb{Z}} a_n e^{-in\xi/2}}_{A(\xi/2)} \hat{\varphi}(\xi/2) &= \underbrace{\sum_{n\in\mathbb{Z}} b_n e^{-in\xi}}_{B(\xi)} \hat{\varphi}(\xi) + \underbrace{\sum_{n\in\mathbb{Z}} c_n e^{-in\xi}}_{C(\xi)} \hat{\psi}(\xi) \\ &= B(\xi)m_0(\xi/2)\hat{\varphi}(\xi/2) + C(\xi)g(\xi/2)\hat{\varphi}(\xi/2).\end{aligned}$$

It is therefore enough to show that for all 2π-periodic functions $A \in L^2(\mathbb{T})$, there exists 2π-periodic functions $B, C \in L^2(\mathbb{T})$ such that

$$A(\xi/2) = B(\xi)m_0(\xi/2) + C(\xi)g(\xi/2).$$

One can verify, as a result of Lemma 1, (2) and (4), that

$$\begin{aligned} B(\xi) &= A(\xi/2)\overline{m_0(\xi/2)} + A(\xi/2+\pi)\overline{m_0(\xi/2+\pi)}, \\ C(\xi) &= A(\xi/2)\overline{g(\xi/2)} + A(\xi/2+\pi)\overline{g(\xi/2+\pi)}, \end{aligned}$$

are appropriate choices.

□

Remark 1. Actually, $\{\psi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0, the orthogonal complement of V_0 in V_1, if and only if there exists some 2π-periodic function ν with $|\nu(x)| = 1$ a.e. such that $\hat{\psi} = e^{i\xi/2}\nu(\xi)\overline{m_0(\xi/2+\pi)}\hat{\varphi}(\xi/2)$.

We have therefore shown that given any MRA and scaling function, we can always construct an orthonormal wavelet by

$$\boxed{\hat{\psi}(\xi) = e^{i\xi/2}\overline{m_0(\xi/2+\pi)}\hat{\varphi}(\xi/2).}$$

Recall also that

$$\hat{\varphi}(2\xi) = \hat{\varphi}(\xi)m_0(\xi), \qquad m_0(\xi) = \sum_{k\in\mathbb{Z}} \alpha_k e^{ik\xi}.$$

So,

$$\hat{\psi}(2\xi) = e^{i\xi}\hat{\varphi}(\xi)\sum_{k\in\mathbb{Z}} \overline{\alpha}_k e^{-ik\xi}(-1)^k \iff \hat{\psi}(\xi) = \hat{\varphi}(\xi/2)\sum_{k\in\mathbb{Z}} \overline{\alpha}_k e^{-i(k-1)\xi/2}(-1)^k$$

and by taking the Fourier transform,

$$\boxed{\psi(x) = 2\sum_{k\in\mathbb{Z}}(-1)^k\overline{\alpha}_k\varphi(2x-(k-1)).}$$

Remark 2. Not all wavelets are associated with an MRA, however, non-MRA wavelets are rare. In particular, if ψ is an orthonormal wavelet such that any of the following conditions hold, then it must be an MRA wavelet.

- ψ is compactly supported.
- $\left|\hat{\psi}\right|$ is continuous and $\left|\hat{\psi}(x)\right| = \mathcal{O}(|x|^{-1/2-\alpha})$ for some $\alpha > 0$.
- ψ is bandlimited and $\left|\hat{\psi}\right|$ is continuous.

3 Choosing a wavelet

Most applications of wavelet bases exploit their ability to efficiently approximate particular classes of functions with few nonzero wavelet coefficients. Therefore, we would like a wavelet such that most of the coefficients $\langle f,\, \psi_{j,n}\rangle \approx 0$. There are generally 3 desirable qualities:

- decay
- vanishing moments
- smoothness

Definition 3. *We say that ψ has p vanishing moments if $\int t^k\psi(t)\mathrm{d}t = 0$ for all $k = 0, \dots, p-1$.*

Note that if ψ has p vanishing moments, then $\langle f, \psi \rangle = 0$ whenever f is a polynomial of degree at most $p-1$. In general, if f has very few discontinuities and is smooth between the discontinuities, then one may want to choose a wavelet with many vanishing moments. On the other hand, as the density of the singularities increase, one may wish to find a wavelet with smaller support at the cost of reducing the number of vanishing moments.

Although the size of the support and the number of vanishing moments are not directly linked, one can show that the size of the support of an orthogonal wavelet with p vanishing moments necessarily have support of size at least $2p-1$.

We will now prove that smoothness in a wavelet in fact implies the vanishing moments property.
Proposition 2. *Suppose that ψ is an orthonormal wavelet. For $l \in \mathbb{N}$, assume that*

- $\psi \in C^l$,
- $\psi^{(s)}$ *is bounded on* $\mathbb{R}$ *for* $s = 0, \ldots, l$,
- $|\psi(x)| \leqslant C/(1+|x|)^\alpha$ *for some* $\alpha > l+1$.

Then, ψ has $l+1$ vanishing moments.

Proof. For a contradiction, let us assume that we can let s be the smallest integer in $\{0, \ldots, l\}$ such that $\int x^s \psi(x) \neq 0$.

Since ψ is not a polynomial, $\psi^{(s)} \neq 0$. Let $a = k2^{-J}$ with $J \geqslant 0$ be such that $\psi^{(s)}(a) \neq 0$. Using Taylor's formula,

$$\psi(x) = \sum_{r=0}^{s} a_r (x-a)^r + R(x)$$

where $a_r = \frac{\psi^{(r)}(a)}{r!}$, and the remainder R satisfies $|R(x)| \leqslant C' |x-a|^s$ for all $x \in \mathbb{R}$ and

$$\forall \varepsilon > 0, \exists \delta > 0 \quad |R(x)| \leqslant \varepsilon |x-a|^s, \qquad \text{whenever } |x-a| \leqslant \delta.$$

From our assumption of $\psi^{(s)}(a) \neq 0$, $a_s \neq 0$. For $j > J$, let $k_j = 2^j a = 2^{j-J} k \in \mathbb{Z}$. Then, since ψ is an orthogonal wavelet,

$$\int \psi(x) \overline{\psi(2^j x - k_j)} \mathrm{d}x = 0.$$

So, letting $u = x - a$,

$$\int \left(\sum_{r=0}^{s} a_r u^r + R(u+a) \right) \overline{\psi(2^j u)} \mathrm{d}u = 0.$$

For $r < s$, by our choice of s, $\int u^r \overline{\psi(2^j u)} \mathrm{d}u = 0$, so

$$-\int a_s u^s \overline{\psi(2^j u)} \mathrm{d}u = \int R(u+a) \overline{\psi(2^j u)} \mathrm{d}u.$$

Putting $x = 2^j u$,

$$-2^{-j-js} \int a_s x^s \overline{\psi(x)} \mathrm{d}x = \int R(u+a) \overline{\psi(2^j u)} \mathrm{d}u \implies \int x^s \overline{\psi(x)} \mathrm{d}x = \frac{-2^{j+js}}{a_s} \int R(u+a) \overline{\psi(2^j u)} \mathrm{d}u$$

To conclude this proof, we simply need to show that the RHS tends to 0 as $j \to \infty$.

By recalling the properties of R, So

$$
\begin{aligned}
&2^{j+js}\int R(u+a)\overline{\psi(2^j u)}\mathrm{d}u = 2^{j+js}\varepsilon\int_{-\delta}^{\delta}|u|^s\,\overline{\psi(2^j u)}\mathrm{d}u + C' 2^{j+js}\int_{|u|>\delta}|u|^s\,\overline{\psi(2^j u)}\mathrm{d}u\\
&\leqslant C2^{j+js}\varepsilon\int_{-\delta}^{\delta}\frac{|u|^s}{(1+|2^j u|)^\alpha}\mathrm{d}u + CC'2^{j+js}\int_{|u|>\delta}\frac{|u|^s}{(1+|2^j u|)^\alpha}\mathrm{d}u\\
&\leqslant \underbrace{C\varepsilon\int_{-2^j\delta}^{2^j\delta}\frac{|t|^s}{(1+|t|)^\alpha}\mathrm{d}t}_{\lesssim\varepsilon} + \underbrace{CC'\int_{|u|>2^j\delta}\frac{|t|^s}{(1+|t|)^\alpha}\mathrm{d}t}_{\to 0 \text{ as } j\to\infty}.
\end{aligned}
$$

□

3.1 Examples of wavelets

The Haar wavelet Let V_j be the set of functions in L^2 which are constant on $[n2^{-j},(n+1)2^{-j})$ for $n\in\mathbb{Z}$. Then $\{V_j\}_{j\in\mathbb{Z}}$ is an MRA with scaling function $\varphi = \mathbb{1}_{[-1,0)}$. The corresponding low pass filter is $m_0 = \frac{1}{2}(1+e^{i\xi})$. Since

$$\hat{\varphi}(\xi) = \frac{1-e^{i\xi}}{-i\xi} = e^{i\xi/2}\frac{\sin(\xi/2)}{(\xi/2)},$$

the wavelet ψ satisfies

$$\hat{\psi}(\xi) = e^{i\xi/2}\frac{(1-e^{-i\xi/2})(1-e^{i\xi/2})}{-i\xi} = ie^{i\xi/2}\frac{\sin^2(\xi/4)}{\xi/4}.$$

This is the Fourier transform of $\psi = \mathbb{1}_{[-1,-1/2)} - \mathbb{1}_{[-1/2,0)}$.

Lecture 4

The Shannon wavelet Let V_j be the set of functions in $L^2(\mathbb{R})$ whose Fourier transform have support contained in $[-2^j\pi, 2^j\pi]$. Then,

- $\varphi(x) = \sin(\pi x)/(\pi x)$ is so that $\{\varphi_{0,n}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of V_0. One can verify that this is a scaling function for MRA $\{V_j\}_{j\in\mathbb{Z}}$.
- Recall that the low pass filter m_0 satisfies

  $$\hat{\varphi}(2\xi) = \mathbb{1}_{[-\pi,\pi]}(2\xi) = \hat{\varphi}(\xi)m_0(\xi) = \mathbb{1}_{[-\pi,\pi]}(\xi)m_0(\xi).$$

  So, $m_0 = \mathbb{1}_{[-\pi/2,\pi/2]}$.
- $\hat{\psi}(\xi) = \mathbb{1}_{[-2\pi,-\pi]\cup[\pi,2\pi]}(\xi)\exp(i\xi/2)$, and so,

  $$\psi(x) = -2\frac{\sin(2\pi x)+\cos(\pi x)}{\pi(2x+1)}.$$

Since $\hat{\psi}$ has compact support, ψ is C^∞, but observe that it decays slowly in time. In particular, $|\psi(x)|$ decays like $|x|^{-1}$ since $\hat{\psi}$ is discontinuous as $\pm\pi$ and $\pm 2\pi$.

Meyer's wavelet The scaling function of Meyer's wavelet is defined in the Fourier domain by

$$\hat{\varphi}(\xi) = \begin{cases} 1 & |\xi| \leqslant 2\pi/3 \\ \cos[\frac{\pi}{2}\nu(\frac{3}{2\pi}|\xi| - 1)] & \frac{2\pi}{3} \leqslant |\xi| \leqslant \frac{4\pi}{3} \\ 0 & \text{otherwise,} \end{cases}$$

where $\nu \in C^k$ or C^∞ is a monotone function which satisfies $\nu(x) = 0$ for all $x \leqslant 0$, $\nu(x) = 1$ for $x \geqslant 1$ and $\nu(x) + \nu(1-x) = 1$ for all $x \in \mathbb{R}$. One can show that the associated wavelet is such that

$$\hat{\psi}(\xi) = \begin{cases} 0 & |\xi| \leqslant 2\pi/3 \\ e^{i\xi/2}\sin[\frac{\pi}{2}\nu(\frac{3}{2\pi}|\xi| - 1)] & \frac{2\pi}{3} \leqslant |\xi| \leqslant \frac{4\pi}{3} \\ e^{i\xi/2}\cos[\frac{\pi}{2}\nu(\frac{3}{4\pi}|\xi| - 1)] & \frac{4\pi}{3} \leqslant |\xi| \leqslant \frac{8\pi}{3} \\ 0 & \text{otherwise.} \end{cases}$$

- Since $\hat{\varphi}$ and $\hat{\psi}$ have compact support, φ and ψ are C^∞.
- The smoother transition in $\hat{\varphi}$ (compare with that of the Shannon scaling function) results in faster decay in time. If $\nu \in C^\infty$, then for all N, there exists A_N such that

  $$|\psi(x)| \leqslant \frac{A_N}{(1+|x|)^N}, \qquad |\varphi(x)| \leqslant \frac{A_N}{(1+|x|)^N}.$$

  We remark however that although there is fast asymptotic decay, the constant A_N grows with N and in practice, the numerical decay of ψ may be slow.
- Note that $\hat{\psi}(0) = 0$ and $\frac{\mathrm{d}^n}{\mathrm{d}\xi^n}\hat{\psi}(0) = 0$ for all $n \in \mathbb{N}$. Since given $g(t) = (-it)^n f(t)$, $\hat{g}(\xi) = \frac{\mathrm{d}^n}{\mathrm{d}\xi^n}\hat{f}(\xi)$, it follows that

  $$\hat{g}(0) = (-i)^n \int t^n f(t)\mathrm{d}t = \hat{f}^{(n)}(0).$$

  Going back to ψ, we see that

  $$\int t^n \psi(t)\mathrm{d}t = 0, \qquad \forall n \in \mathbb{N}.$$

One can in fact relax the requirement of an orthonormal basis (V) to the requirement of a Riesz basis. This is sometimes useful in the construction of wavelets.

Definition 4. *Let $\mathcal{H}$ be a Hilbert space. A sequence $\{g_k\}_{k\in\mathbb{Z}} \subset \mathcal{H}$ is called a Riesz sequence if there exists $B \geqslant A > 0$ such that for all $c \in \ell^2(\mathbb{Z})$,*

$$A\sum_k |c_k|^2 \leqslant \|\sum_{k\in\mathbb{Z}} c_k g_k\|_{\mathcal{H}}^2 \leqslant B\sum_k |c_k|^2.$$

Moreover, if $\overline{\mathrm{Span}\{g_k\}_{k\in\mathbb{Z}}} = \mathcal{H}$, then it is called a Riesz basis.

We first require a result about Riesz bases (for a proof, see [Ch 2, Hernandez & Weiss].):

Lemma 2. *Let $g \in L^2(\mathbb{R})$ and let $\sigma_g(\xi) = \sqrt{\sum_{k\in\mathbb{Z}} |\hat{g}(\xi + 2k\pi)|^2}$. Then, $\{g(\cdot - k) \,;\, k \in \mathbb{Z}\}$ is a Riesz basis, i.e. there exists $B \geqslant A > 0$ such that*

$$A\sum_k |c_k|^2 \leqslant \|\sum_{k\in\mathbb{Z}} c_k g(\cdot - k)\|_{L^2}^2 \leqslant B\sum_k |c_k|^2$$

if and only if

$$\sigma_g(\xi) \in [\sqrt{A}, \sqrt{B}], \qquad a.e.\, \xi \in \mathbb{R}.$$

Proposition 3. *Suppose that $\varphi \in L^2(\mathbb{R})$ is such that $\{\varphi(\cdot - k) \,;\, k \in \mathbb{Z}\}$ is a Riesz basis of a closed subspace V_0 of $L^2(\mathbb{R})$. Let γ be such that $\hat{\gamma} = \hat{\varphi}/\sigma_\varphi$. Then, $\{\gamma(\cdot - k) \,;\, k \in \mathbb{Z}\}$ is an orthonormal basis for V_0.*

Proof. Let $\hat{\gamma} = \hat{\varphi}/\sigma_\varphi$. Since $1/\sigma_\varphi \in [1/\sqrt{B}, 1/\sqrt{A}]$, $\hat{\gamma} \in L^2(\mathbb{R})$ and so, $\gamma \in L^2(\mathbb{R})$. Note also that $1/\sigma_\varphi$ and σ_φ are both 2π periodic functions, and as elements of $L^2(\mathbb{T})$, there exist sequences (a_k) and (b_k) such that

$$\frac{1}{\sigma_\varphi(\xi)} = \sum_k a_k e^{-ik\xi} \qquad \sigma_\varphi(\xi) = \sum_k b_k e^{-ik\xi}.$$

Also,

$$\hat{\gamma}(\xi) = \hat{\varphi}(\xi) \sum_k a_k e^{-ik\xi}, \qquad \hat{\varphi}(\xi) = \hat{\gamma}(\xi) \sum_k b_k e^{-ik\xi},$$

and this implies that

$$\gamma(x) = \sum_k a_k \varphi(x-k), \qquad \varphi(x) = \sum_k b_k \gamma(x-k).$$

So, $\gamma \in \overline{\operatorname{span}\{\varphi_{0,k}\ ;\ k \in \mathbb{Z}\}}$ and $\varphi \in \overline{\operatorname{span}\{\gamma_{0,k}\ ;\ k \in \mathbb{Z}\}}$. Therefore, $V_0 = \overline{\operatorname{span}\{\gamma_{0,k}\ ;\ k \in \mathbb{Z}\}}$. Finally, by definition of $\hat{\gamma}$ and since σ_φ is 2π-periodic,

$$\sum_k |\hat{\gamma}(\xi + 2k\pi)|^2 = \frac{1}{\sigma_\varphi^2(\xi)} \sum_k |\hat{\varphi}(\xi + 2k\pi)|^2 = 1, \qquad a.e.\, \xi \in \mathbb{R}.$$

By Proposition 1, we may conclude that $\{\gamma_{0,k}\ ;\ k \in \mathbb{Z}\}$ is an orthonormal basis of V_0. $\square$

Spline wavelets The space V_j of degree $m \geqslant 0$ is the set of functions that are $m-1$ continuously differentiable and equal to a polynomial of degree m on the interval $[n2^{-j}, (n+1)2^{-j}]$ for all $n \in \mathbb{Z}$. Let $\Delta^0 = \mathbb{1}_{[0,1]}$, $\Delta^1 = \mathbb{1}_{[0,1]} \star \mathbb{1}_{[0,1]}$ and for $m \geqslant 2$:

$$\Delta^m = \Delta^{m-1} \star \mathbb{1}_{[0,1]}.$$

We call Δ^m the basic spline of order m, and one can check that $\{\Delta^m(\cdot - k) :\ k \in \mathbb{Z}\}$ forms a Riesz basis of V_0 of order m. Furthermore, Δ^m has support $[0, m+1]$ and $\widehat{\Delta^m}(\xi) = e^{-i(m+1)\xi/2} \left(\frac{\sin(\xi/2)}{\xi/2}\right)^{m+1}$.

From the Proposition 3, we can obtain a scaling function by choosing φ to be such that

$$\hat{\varphi}(\xi) = \frac{\widehat{\Delta^m}(\xi)}{\sqrt{\sum_{k\in\mathbb{Z}} \left|\widehat{\Delta^m}(\xi + 2\pi k)\right|^2}} = \frac{\exp(-i(m+1)\xi/2)}{\xi^{m+1} P_{2m+2}(\xi)}.$$

where $P_n(\xi) = \sqrt{\sum_{k\in\mathbb{Z}} \frac{1}{|\xi + 2k\pi|^n}}$.

From properties of the basic spline of degree m, the associated wavelet ψ^m and scaling function φ^m are both $m-1$ differentiable and the wavelet has $m+1$ vanishing moments. Furthermore, the functions

$$\frac{\mathrm{d}^j \psi^m}{\mathrm{d}x^j}, \qquad \frac{\mathrm{d}^j \varphi^m}{\mathrm{d}x^j}, \qquad j = 0, 1, \ldots, m-1,$$

all have exponential decay.

4 Construction of compactly supported wavelets

4.0.1 Construction using the low pass filter

The following theorem describes how to construct an MRA given a low pass filter.

Theorem 5. *Suppose that* $m(\xi)=\sum_{k=T}^{S} a_k e^{ik\xi}$ *satisfies*

(i) $|m(\xi)|^2+|m(\xi+\pi)|^2=1$, *for all* $\xi\in\mathbb{R}$.

(ii) $m(0)=1$.

(iii) $m(\xi)\neq 0$ *for all* $\xi\in[-\pi/2,\pi/2]$.

Then,

$$\Theta(\xi)=\prod_{j=1}^{\infty} m(2^{-j}\xi)$$

converges uniformly on compact sets, $\Theta(\xi)$ *is continuous, in* $L^2(\mathbb{R})$ *and* φ *such that* $\hat{\varphi}=\Theta$ *has support in* $[-S,-T]$ *and is a scaling function for an MRA. In this case,*

$$\psi(x)=2\sum_{k=T}^{S}\overline{a}_k\varphi(2x-(k-1))(-1)^k$$

is a compactly supported wavelet with support contained in $[(T-S-1)/2,(S-T-1)/2]$.

Remark 3. Some remarks about the conditions in Theorem 5:

(i) From Lemma 1, we know this is necessary if m is a low pass filter of an MRA.

(ii) If $\hat{\varphi}$ is a continuous scaling function (which is true if $\varphi\in L^1$), then $|\hat{\varphi}(0)|=1$. Assume that $\hat{\varphi}(0)=1$, then, from $\hat{\varphi}(\xi)=m(\xi/2)\hat{\varphi}(\xi/2)$, it follows that $m(0)=1$. Also, repeated application of $\hat{\varphi}(\xi)=m(\xi/2)\hat{\varphi}(\xi/2)$ yields $\hat{\varphi}(\xi)=\prod_{j\geqslant 1} m_0(2^{-j}\xi)$.

(iii) One can check that $m(\xi)=(1+e^{i\xi})/2$ satisfies conditions (i)-(iii), $\prod_{j\geqslant 1} m(2^{-j}\xi)=e^{i\xi/2}\frac{\sin(\xi/2)}{\xi/2}$ which is the Fourier transform of $\mathbb{1}_{[-1,0]}$, the Haar scaling function. Note however that $m_1(\xi):=m(3\xi)$ satisfies conditions (i), (ii) but not (iii). One can check that $\prod_{j\geqslant 1} m_1(2^{-j}\xi)$ is the Fourier transform of $\frac{1}{3}\mathbb{1}_{[-3,0]}$. So, a condition like (iii) is needed to ensure orthogonality of the translates.

Lecture 5

Lemma 3. *If* $m(\xi)$ *is a trigonometric polynomial and (i) and (ii) of Theorem 5 hold, then*

$$\Theta(\xi)=\prod_{j=1}^{\infty} m(2^{-j}\xi)$$

converges uniformly on compact sets and $\int|\Theta(\xi)|^2\,\mathrm{d}\xi\leqslant 2\pi$. *If we assume also (iii) of Theorem 5, then*

$$\frac{1}{2\pi}\int|\Theta(\xi)|^2 e^{-ik\xi}\mathrm{d}\xi=\delta_{0,k},$$

and if φ *is such that* $\hat{\varphi}=\Theta$, *then* $\{\varphi(\cdot-k):\ k\in\mathbb{Z}\}$ *forms an orthonormal system.*

Proof. First note that by (ii), there exists $C>0$ such that

$$|m(\xi)-1|\leqslant C|\xi|.$$

Let $\Pi_N=\prod_{j=1}^{N} m(2^{-j}\xi)$. Then,

$$|\Pi_{N+1}(\xi)-\Pi_N(\xi)|\leqslant|\Pi_N(\xi)|\left|m(\xi/2^{N+1})-1\right|\leqslant\frac{C|\xi|}{2^{N+1}},$$

where we have used the fact that $|m(\xi)| \leqslant 1$ for all $\xi \in \mathbb{R}$ (which follows from (i)). By applying the triangle inequality, for all $N, M \in \mathbb{N}$,

$$\begin{aligned}|\Pi_{N+M}(\xi) - \Pi_N(\xi)| &\leqslant |\Pi_{N+M}(\xi) - \Pi_{N+M-1}(\xi)| + \cdots + |\Pi_{N+2}(\xi) - \Pi_{N+1}(\xi)| + |\Pi_{N+1}(\xi) - \Pi_N(\xi)| \\ &\leqslant \sum_{j=1}^{M} \frac{C\,|\xi|}{2^{N+j}} \leqslant \frac{C\,|\xi|}{2^N}.\end{aligned}$$

So, Π_N converges uniformly to Θ compact sets. Moreover, as the uniform limit of continuous functions, Θ is also a continuous function.

Let $g_N(\xi) = \Pi_N(\xi)\mathbb{1}_{[-2^N\pi, 2^N\pi]}(\xi) = \Pi_N(\xi)\mathbb{1}_{[-\pi/2,\pi/2]}(2^{-N-1}\xi)$. Let

$$I_N^k = \int |g_N(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi = \int_{-2^N\pi}^{2^N\pi} |\Pi_N(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi.$$

Observe that Π_N is a $2^{N+1}\pi$-periodic function. So, by applying (i),

$$\begin{aligned}I_N^k &= \int_0^{2^{N+1}\pi} |\Pi_N(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi = \int_0^{2^N\pi} |\Pi_N(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi + \int_{2^N\pi}^{2^{N+1}\pi} |\Pi_N(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi \\ &= \int_0^{2^N\pi} \left|m(2^{-N}\xi)\right|^2 |\Pi_{N-1}(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi + \int_0^{2^N\pi} \left|m(2^{-N}\xi + \pi)\right|^2 |\Pi_{N-1}(\xi)|^2\, e^{-ik\xi} \mathrm{d}\xi \\ &= I_{N-1}^k = I_1^k = \int_{-2\pi}^{2\pi} |m(\xi/2)|^2\, e^{-ik\xi} \mathrm{d}\xi.\end{aligned}$$

By a change of variable, we see that

$$I_N^k = 2\int_{-\pi}^{\pi} |m(\mu)|^2\, e^{-2ik\mu} \mathrm{d}\mu = 2\int_{-\pi}^{0} (|m(\mu)|^2 + |m(\mu+\pi)|^2) e^{-2ik\mu} \mathrm{d}\mu = \begin{cases} 2\pi & k = 0 \\ 0 & k \neq 0. \end{cases}$$

Since $|m(\xi)| \leqslant 1$ for all $\xi \in \mathbb{R}$ by (i), we have that

$$\int_{-2^N\pi}^{2^N\pi} |\Theta(\xi)|^2 \leqslant I_N^0 = 2\pi \implies \|\Theta\|_{L^2}^2 \leqslant 2\pi.$$

To prove the last part, assume now that (iii) holds. Now, if we can show that for all $\xi \in \mathbb{R}$, $|g_N(\xi)| \leqslant \frac{1}{C'}\,|\Theta(\xi)|$, then applying the dominated convergence theorem to the equality $I_N^k = 2\pi\delta_{k,0}$ would yield the required result.

Since m is a continuous function, (ii) and (iii) imply that there exists $C > 0$ such that

$$|m(\xi)| > C, \qquad \forall \xi \in [-\pi/2, \pi/2].$$

Note also that $\Theta(0) = 1$ (as it is uniformly convergent on $[-\pi,\pi]$) and there exists $M > 0$ such that

$$\left|\Theta(\xi 2^{-M+1})\right| > 1/2, \qquad \forall \xi \in [-\pi, \pi].$$

From this, we have that for all $\xi \in [-\pi,\pi]$,

$$\Big|\prod_{j\geqslant M} m(2^{-j}\xi)\Big| = \Big|\prod_{j\geqslant 1} m(2^{-j+1-M}\xi)\Big| = \left|\Theta(\xi 2^{-M+1})\right| > 1/2.$$

So, for $\xi \in [-\pi,\pi]$,

$$|\Theta(\xi)| = \prod_{j=1}^{M-1} \left|m(2^{-j}\xi)\right| \cdot \prod_{j\geqslant M} \left|m(2^{-j}\xi)\right| > \frac{C^{M-1}}{2} =: C' > 0.$$

By writing $\Theta(\xi) = \Pi_N(\xi)\Theta(2^{-N}\xi)$, we see that for all $\xi \in [-2^N\pi, 2^N\pi]$,

$$C'\left|\Pi_N(\xi)\right| \leqslant \left|\Pi_N(\xi)\Theta(2^{-N}\xi)\right| = \left|\Theta(\xi)\right|.$$

So, for all $\xi \in \mathbb{R}$, $|g_N(\xi)| \leqslant \frac{1}{C'}|\Theta(\xi)|$. Now, $g_N(\xi) \to \Theta(\xi)$ a.e. $\xi \in \mathbb{R}$ and by the dominated convergence theorem,

$$\int |\Theta(\xi)|^2 e^{-ik\xi}\mathrm{d}\xi = \int \lim_{N\to\infty} |g_N(\xi)|^2 e^{-ik\xi}\mathrm{d}\xi = \lim_{N\to\infty} \int |g_N(\xi)|^2 e^{-ik\xi}\mathrm{d}\xi = \begin{cases} 2\pi & k = 0 \\ 0 & k \neq 0. \end{cases}$$

By splitting the integral on the LHS of the above inequality into an integral over intervals of length 2π, we see that all Fourier coefficients of the 2π-periodic function $\sum_{l\in\mathbb{Z}} |\Theta(\xi + 2l\pi)|^2$ are zero, except for the 0th coefficient. Therefore,

$$\sum_{l\in\mathbb{Z}} |\Theta(\xi + 2l\pi)|^2 = 1, \qquad a.e.$$

and by Proposition 1, $\{\varphi(\cdot - k) : \ k \in \mathbb{Z}\}$ is an orthonormal system.

$\square$

Lemma 4. *Let $m(\xi) = \sum_{k=T}^{S} a_k e^{ik\xi}$ satisfy (i) and (ii). Let $\Theta(\xi)$ be as in Theorem 5. Then,* $\mathrm{Supp}(\check{\Theta}) \subset [-S, -T]$. *If a_k's are real numbers, then $\check{\Theta}$ is a real valued function.*

Proof. Since m satisfies (i) and (ii), from Lemma 3, we know that Θ and $\Pi_N = \prod_{j=1}^{N} m(2^{-j}\cdot)$ are well defined functions in $L^2(\mathbb{R})$.

Now, as m is a trigonometric polynomial, we see that $\Pi_N(\xi)$ is a finite linear combination of terms of the form

$$\exp(i(\sum_{j=1}^{N} 2^{-j}p_j)\xi),$$

where $p_j \in [T, S]$. So, in a distributional sense, the inverse Fourier transform of Π_N is a finite linear combination of Dirac functions of the form δ_a where $a = -\sum_{j=1}^{N} 2^{-j}p_j$: By writing $\Pi_N(\xi) = \sum_{l=1}^{k} \beta_l \exp(ia_l\xi)$, for all $g \in C_c^\infty(\mathbb{R})$,

$$\langle \check{\Pi}_N,\, g\rangle = \frac{1}{2\pi}\langle \Pi_N,\, \hat{g}\rangle = \frac{1}{2\pi}\sum_{l=1}^{k} \beta_l \langle \exp(ia_l\cdot),\, \hat{g}\rangle = \sum_{l=1}^{k} \beta_l \overline{g}(-a_l) = \langle \sum_{l=1}^{k} \beta_l \delta_{-a_l},\, g\rangle.$$

In particular, $\check{\Pi}_N$ is supported on $[-S, -T]$. So, given any $f \in C_c^\infty$ such that $\mathrm{Supp}(f) \cap [-S, -T] = \emptyset$, $\hat{f}$ is in the Schwartz space, so by applying Plancherel, Lemma 3 and the above observation:

$$\begin{aligned}\langle \check{\Theta},\, f\rangle &= \frac{1}{2\pi}\langle \Theta,\, \hat{f}\rangle = \lim_{N\to\infty} \frac{1}{2\pi}\langle \Pi_N,\, \hat{f}(\xi)\rangle \\ &= \lim_{N\to\infty} \langle \sum_{l=1}^{k} \beta_l \delta_{-a_l},\, f\rangle = 0.\end{aligned}$$

Since f is arbitrary, this implies that $\mathrm{Supp}\,\check{\Theta} \subset [-S, -T]$. If a_k's are real, then the β_l's in Π_N are a real valued. So, if $f \in C^\infty$ is real valued, then $\int \check{\Theta}(t)f(t)\mathrm{d}t$ is also real valued. So, $\check{\Theta}$ is real. $\square$

Proof of Theorem 5. By Lemma 3 and Lemma 4, φ is well-defined and compactly supported. Furthermore, the following properties are satisfied:

(i) $\{\varphi_{0,k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis in $L^2(\mathbb{R})$.

(ii) $\varphi(\cdot/2) = \sum_k a_k \varphi(\cdot - k)$.

(iii) $\hat{\varphi}(\xi)$ is continuous and $\hat{\varphi}(0) \neq 0$.

Note that (ii) follows from taking the Fourier transform of $\hat{\varphi}(2\xi) = m(\xi)\hat{\varphi}(\xi)$. Let us define $V_j = \{f(2^j \cdot) : f \in V_0\}$ and consider the MRA properties (I)-(V): Clearly, (II) holds. Also, (i) implies (V) and (ii) implies (I). By Theorem 2, (III) follows. Finally, since $\hat{\varphi}(0) \neq 0$ and $\hat{\varphi}$ is continuous, Lemma 3 implies (IV).

□

4.0.2 Daubechies wavelets

Daubechies constructed a family of compacted supported wavelets with arbitrarily many vanishing moments. The constructed can be summarized as follows: First define a non-negative trigonometric polynomial M such that

(i) $M(\xi) + M(\xi + \pi) = 1$, for all $\xi \in \mathbb{R}$.

(ii) $M(0) = 1$.

(iii) $M(\xi) \neq 0$ for all $\xi \in [-\pi/2, \pi/2]$.

In particular, for each odd integer $N = 2K - 1$, M is chosen to be the first sum on the right hand side of the identity

$$1 = \left(\cos^2 \frac{\xi}{2} + \sin^2 \frac{\xi}{2}\right)^N = \left[\sum_{j=0}^{K-1} \binom{2K-1}{j} + \sum_{j=K}^{2K-1} \binom{2K-1}{j}\right] \left(\cos^2 \frac{\xi}{2}\right)^{2K-1-j} \left(\sin^2 \frac{\xi}{2}\right)^j .$$

It is clear that conditions (ii) and (iii) hold, and condition (i) holds by symmetry of the binomial coefficients and since $\cos^2(\xi + \pi/2) = \sin^2(\xi)$. One can then extract a trigonometric polynomial m (of the same degree) from M such that $M(\xi) = |m(\xi)|^2$. This polynomial m would therefore satisfies the condition of Theorem 5, thus leading to the construction of a compactly supported scaling function and wavelet can be constructed. This 'extraction of the square root' is possible by the following lemma:

Lemma 5 (Riesz). *If $g(\xi) = \sum_{k=-T}^{T} \gamma_k e^{ik\xi}$ is a nonnegative trigonometric polynomial with real coefficients, then there exists $m(\xi) = \sum_{k=0}^{T} a_k e^{ik\xi}$ with real coefficients such that $|m(\xi)|^2 = g(\xi)$.*

Remark 4. We will not prove Lemma 5, the reader can refer to [Chapter 2, Hernandez & Weiss] for a proof of this result. However, we will remark only that the proof of this lemma is constructive. Moreover, the function m is in general not unique, and different choices of m will lead to different wavelets.

Examples

1. In the case $K = 1$, $M(t) = \cos^2 \frac{t}{2} = (e^{it/2} \cos \frac{t}{2})(e^{-it/2} \cos \frac{t}{2}) = |m(t)|^2$, where $m(t) = e^{it/2} \cos \frac{t}{2} = (1 + e^{it})/2$ which is the low pass filter associated with the Haar wavelet.

2. In the case $K = 2$:

$$\begin{aligned} M(t) &= \cos^6\left(\frac{t}{2}\right) + 3\cos^4\left(\frac{t}{2}\right)\sin^2\left(\frac{t}{2}\right) = \cos^4 \frac{t}{2}\left(\cos^2\left(\frac{t}{2}\right) + 3\sin^2\left(\frac{t}{2}\right)\right) \\ &= \cos^4 \frac{t}{2} \left|\cos\left(\frac{t}{2}\right) + i\sqrt{3}\sin\left(\frac{t}{2}\right)\right|^2 . \end{aligned}$$

So,

$$m(t) = \cos^2\left(\frac{t}{2}\right)\left(\cos\left(\frac{t}{2}\right) + i\sqrt{3}\sin\left(\frac{t}{2}\right)\right)\alpha(t)$$

where α is such that $|\alpha(t)| = 1$ and will be determined later. By writing

$$\cos\left(\frac{t}{2}\right) = \frac{e^{it/2} + e^{-it/2}}{2}, \qquad \sin\left(\frac{t}{2}\right) = \frac{e^{it/2} - e^{-it/2}}{2i},$$

we have that

$$\begin{aligned} m(t) &= \frac{1}{8}\left(e^{it} + 2 + e^{-it}\right)\left(e^{it/2} + e^{-it/2} + \sqrt{3}e^{it/2} - \sqrt{3}e^{-it/2}\right)\alpha(t) \\ &= \frac{e^{it}}{8}\left(e^{-it} + 1\right)^2\left(e^{it/2} + e^{-it/2} + \sqrt{3}e^{it/2} - \sqrt{3}e^{-it/2}\right)\alpha(t) \end{aligned}$$

Choosing $\alpha(t) = e^{-3it/2}$ yields $m(t) = P(e^{-it})$ where P is such that

$$P(z) = \frac{1}{8}\left((1+\sqrt{3}) + (3+\sqrt{3})z + (3-\sqrt{3})z^2 + (1-\sqrt{3})z^3\right) = \left(\frac{1+\sqrt{3}}{8} + \frac{1-\sqrt{3}}{8}z\right)(1+z)^2.$$

Lecture 6

Vanishing moments The wavelet constructed via Daubechies' construction has K vanishing moments: In general, for each K, the low pass filter is $m_K(t) = P_K(e^{-it})$ where

$$P_K(z) = (z+1)^K \tilde{P}_K(z)$$

and $\tilde{P}_K$ is a polynomial of degree $K-1$ such that $\tilde{P}_K(-1) \neq 0$.

The associated scaling function φ_K and wavelet ψ_K are such that

$$\hat{\varphi}_K(\xi) = \prod_{j\geqslant 1} P_K(e^{-i\xi/2^j}), \qquad \hat{\psi}_K(\xi) = e^{i\xi/2} P_K(-e^{i\xi/2})\hat{\varphi}_K(\xi/2).$$

One can now check that

$$\hat{\psi}_K^{(j)}(0) = \begin{cases} 0 & j = 0, \dots, K-1 \\ \tilde{P}_K(-1)K!(i/2)^K & j = K. \end{cases}$$

It therefore follows that ψ_K has K vanishing moments. Note also that by Theorem 5, φ_K and ψ_K both have support of size $N = 2K-1$.

Remark 5 (Smoothness). Through her construction, Daubechies was able to show that: there exists a constant c_0 such that for all $r > 0$, there exists a wavelet $\psi \in C^r$ such that $|\mathrm{Supp}(\psi)| = N \leqslant c_0 r$.

5 Fast wavelet transform

For $j, n \in \mathbb{Z}$, let

$$a_{j,n} = \langle f, \varphi_{j,n}\rangle, \qquad d_{j,n} = \langle f, \psi_{j,n}\rangle.$$

Recall that $V_J = V_0 \oplus W_0 \oplus \cdots \oplus W_{J-1}$. Given any $f \in V_J$,

$$f = \sum_n a_{J,n}\varphi_{J,n}, \qquad f = \sum_n a_{0,n}\varphi_{0,n} + \sum_{j=0}^{J-1}\sum_n d_{j,n}\psi_{j,n},$$

Definition 5. *For $L < J$, the orthogonal wavelet representation of $\{a_{J,n}\}_{n\in\mathbb{Z}}$ consists of the wavelet coefficients at scales between L and $J-1$, and the scaling coefficients at scale L:*

$$\{d_{l,n}\}_{L\leqslant l\leqslant J-1, n\in\mathbb{Z}} \cup \{a_{L,n}\}_{n\in\mathbb{Z}}$$

We first consider one level of the decomposition $V_j = V_{j-1} \oplus W_{j-1}$. In particular, we discuss how to efficiently obtain $\{a_{j-1,n}, d_{j-1,n}\}_n$ from $\{a_{j,n}\}_n$ and vice versa.

5.1 Decomposition and reconstruction

Recall that

$$\frac{1}{2}\varphi(\frac{\cdot}{2}) = \sum_k \alpha_k \varphi(\cdot + k), \qquad \alpha_k = \frac{1}{2}\int \varphi(\frac{x}{2})\overline{\varphi(x+k)}\mathrm{d}x.$$

So,

$$\varphi_{j-1,n} = \sqrt{2}\sum_{k\in\mathbb{Z}} \alpha_k \varphi_{j,2n-k}.$$

Also,

$$\frac{1}{2}\psi(\frac{\cdot}{2}) = \sum_k \beta_k \varphi(\cdot + k), \qquad \beta_k = \frac{1}{2}\int \psi(\frac{x}{2})\overline{\varphi(x+k)}\mathrm{d}x,$$

and

$$\psi_{j-1,n} = \sqrt{2}\sum_{k\in\mathbb{Z}} \beta_k \varphi_{j,2n-k}.$$

In fact, $\beta_k = \overline{\alpha}_{1-k}(-1)^{k+1}$.

Decomposition

$$a_{j-1,n} = \langle f,\, \varphi_{j-1,n}\rangle = \langle f,\, \sqrt{2}\sum_{k\in\mathbb{Z}} \alpha_k \varphi_{j,2n-k}\rangle = \sqrt{2}\sum_k \overline{\alpha}_k a_{j,2n-k} = \sqrt{2}(a_j \star \overline{\alpha})[2n].$$

$$d_{j-1,n} = \langle f,\, \psi_{j-1,n}\rangle = \langle f,\, \sqrt{2}\sum_{k\in\mathbb{Z}} \beta_k \varphi_{j,2n-k}\rangle = \sqrt{2}\sum_k \overline{\beta}_k a_{j,2n-k} = \sqrt{2}(a_j \star \overline{\beta})[2n].$$

Reconstruction

$$\begin{aligned}
\sum_n a_{j,n}\varphi_{j,n} &= \sum_n a_{j-1,}\varphi_{j-1,n} + \sum_n d_{j-1,n}\psi_{j-1,n} \\
&= \sum_n a_{j-1,n}\left(\sqrt{2}\sum_{k\in\mathbb{Z}}\alpha_k\varphi_{j,2n-k}\right) + \sum_n d_{j-1,n}\left(\sqrt{2}\sum_{k\in\mathbb{Z}}\beta_k\varphi_{j,2n-k}\right) \\
&\overbrace{=}^{l=2n-k} \sum_n a_{j-1,n}\left(\sqrt{2}\sum_{l\in\mathbb{Z}}\alpha_{2n-l}\varphi_{j,l}\right) + \sum_n d_{j-1,n}\left(\sqrt{2}\sum_{l\in\mathbb{Z}}\beta_{2n-l}\varphi_{j,l}\right) \\
&= \sum_{l\in\mathbb{Z}}\varphi_{j,l}\left(\sqrt{2}\sum_n a_{j-1,n}\alpha_{2n-l} + d_{j-1,n}\beta_{2n-l}\right)
\end{aligned}$$

Therefore,

$$a_{j,n} = \sqrt{2}\sum_k a_{j-1,k}\alpha_{2k-n} + d_{j-1,k}\beta_{2k-n}$$

Let $\tilde{\alpha}_j = \alpha_{-j}$, $\tilde{\beta}_j = \beta_{-j}$ and let

$$\tilde{a}_{j,n} = \begin{cases} a_{j,n/2} & n \text{ even,} \\ 0 & \text{otherwise,} \end{cases} \qquad \tilde{d}_{j,n} = \begin{cases} d_{j,n/2} & n \text{ even,} \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$a_{j,n} = \sqrt{2}(\tilde{a}_{j-1} \star \tilde{\alpha} + \tilde{d}_{j-1} \star \tilde{\beta})_n.$$

Initialisation

Lemma 6. *Suppose that f is continuous on $\mathbb{R}$ and $\varphi \in L^1(\mathbb{R})$, $\int \varphi = 1$ and $\mathrm{Supp}(\varphi) \subset [-R, R]$. Then, for all $x \in \mathbb{R}$*

$$\left| f(x) - 2^j \int f(x+y)\overline{\varphi(2^j y)} \mathrm{d}y \right| \to 0$$

as $j \to \infty$. Moreover, if f is uniformly continuous, then this convergence is uniform.

Proof. Observe that,

$$\left| f(x) - 2^j \int f(x+y)\overline{\varphi(2^j y)} \mathrm{d}y \right| = \left| 2^j \int (f(x) - f(x+y))\overline{\varphi(2^j y)} \mathrm{d}y \right|$$
$$= \left| \int (f(x) - f(x + 2^{-j} z))\overline{\varphi(z)} \mathrm{d}z \right| \leqslant \|\varphi\|_{L^1} \sup_{|t| \leqslant 2^{-j} R} |f(x) - f(x+t)|.$$

Since f is continuous, given any $\varepsilon > 0$, $\sup_{|t| \leqslant 2^{-j}R} |f(x) - f(x+t)| < \varepsilon$ by taking j sufficiently large. Moreover, if f is uniformly continuous, the choice of j is independent of x. ☐

Therefore, plugging in $x = n/2^j$ in the above lemma, we see that

$$2^j \int f(y)\varphi(2^j y - n) \approx f(n/2^j).$$

Computational complexity Suppose that $\mathrm{Supp}(f) \subset [0, 1]$, and a_j, d_j are of length 2^j. The forward and inverse wavelet transforms are respectively:

$$W : \{a_{J,n}\}_{n=0}^{2^J-1} \mapsto \{d_{j,n}\}_{L \leqslant j \leqslant J-1, 0 \leqslant n \leqslant 2^j - 1} \cup \{a_{L,n}\}_{0 \leqslant n \leqslant 2^L - 1}$$
$$W^{-1} : \{d_{j,n}\}_{L \leqslant j \leqslant J-1, 0 \leqslant n \leqslant 2^j - 1} \cup \{a_{L,n}\}_{0 \leqslant n \leqslant 2^L - 1} \mapsto \{a_{J,n}\}_{n=0}^{2^J-1}.$$

Now, if α and β have K nonzero coefficients (this is the case for compactly supported wavelets), then, the decomposition of a_j into a_{j-1} and d_{j-1} is computed with $2^j K$ elementary operations. So, the forward wavelet transform on a_J and $N = 2^J$ is calculated with at most $2KN$ elementary computations. Similarly, the reconstruction of a_j from a_{j-1} and d_{j-1} is obtained with $2^j K$ elementary operations. So, the cost of the inverse wavelet transform is at most $2KN$.

Wavelet graphs The reconstruction formula provides a way of approximating the graphs of the wavelet and scaling functions.

Let $f = \varphi$, then, the wavelet representation of f in $V_0 \oplus W_0 \oplus \cdots \oplus W_{J-1}$ is

$$\{a_{0,0}\} \cup \{d_{j,k}\}_{\substack{j=0,\ldots,J-1 \\ k=0,\ldots,2^j-1}}, \qquad a_{0,0} = 1, d_{j,k} = 0.$$

Computing the inverse wavelet transform on this sequence yields $\{a_{J,k}\}_{k=0}^{2^J-1}$ where $a_{J,k} \approx 2^{-J/2}\varphi(k/2^J)$.

Similarly, the wavelet representation of $f = \psi$ is

$$\{a_{0,0}\} \cup \{d_{j,k}\}_{\substack{j=0,\ldots,J-1 \\ k=0,\ldots,2^j-1}}, \qquad a_{0,0} = 0, d_{0,0} = 1, d_{j,k} = 0 \quad \forall j > 0.$$

Computing the inverse wavelet transform on this sequence yields $\{a_{J,k}\}_{k=0}^{2^J-1}$ where $a_{J,k} \approx 2^{-J/2}\psi(k/2^J)$

6 Wavelets on the interval

To decompose signals defined on $[0,1]$, it is necessary to construct wavelet bases for $L^2[0,1]$. A couple of simple approaches to constructing a basis for $L^2[0,1]$ from a compactly supported wavelet basis of $L^2(\mathbb{R})$ include:

- simply discarding those wavelets whose support do not intersect $[0,1]$. When considering the wavelet coefficients of $f \in L^2[0,1]$, this is equivalent to extending f to $\mathbb{R}$ by setting $f(t) = 0$ on $t \notin [0,1]$.
- periodizing the wavelets, so letting $\psi_{j,k}^{per}(x) = \sum_j \psi_{j,k}(x+j)$. When considering the wavelet coefficients of $f \in L^2[0,1]$, this is equivalent taking the wavelet coefficients of f^{per}, where f^{per} is the periodization of f such that $f^{per}(t) = f^{per}(t+1)$ for all $t \in \mathbb{R}$ and $f^{per}\mathbb{1}_{[0,1]} = f\mathbb{1}_{[0,1]}$.

The problem with these approaches is that the vanishing moments property is not preserved, and unless f is supported on $(0,1)$, we will effectively be treating f as discontinuous at $t = 0, 1$ and this will lead to large wavelet coefficients at the boundary.

Lecture 7

An alternative approach by Cohen, Daubechies and Vial is to keep only those wavelets with support inside $[0,1]$ and introduce modified wavelets near the boundary points 0 and 1. Note that if ψ has compact support, then there are a constant number of wavelets at each scale whose support overlaps $t = 0$ or $t = 1$.

Suppose that the scaling function φ and wavelet ψ have supports in $[-p+1, p]$. Then, at scale j so that $2^j \geqslant 2p$, there are $2^j - 2p$ scaling functions with support inside $[0,1]$, so these scaling functions are not modified:

$$\varphi_{j,n}^{int} = \varphi_{j,n}, \qquad n = p, \ldots, 2^j - p - 1.$$

To construct an approximation space V_j^{int} of dimension 2^j, we then add p scaling functions on the left boundary near $t = 0$,

$$\varphi_{j,n}^{int} = 2^{j/2}\varphi_n^{left}(2^j \cdot), \qquad n = 0, \ldots, p-1$$

and p scaling functions on the right boundary near $t = 1$:

$$\varphi_{j,n}^{int} = 2^{j/2}\varphi_n^{right}(2^j \cdot), \qquad n = 2^j - p, \ldots, 2^j - 1.$$

Similarly, at scale j, there are $2^j - 2p$ wavelets with support inside $[0,1]$, so these wavelets are not modified:

$$\psi_{j,n}^{int} = \psi_{j,n}, \qquad n = p, \ldots, 2^j - p - 1.$$

We then add p wavelets on the left boundary near $t = 0$,

$$\psi_{j,n}^{int} = 2^{j/2}\psi_n^{left}(2^j \cdot), \qquad n = 0, \ldots, p-1$$

and p wavelets on the right boundary near $t = 1$:

$$\psi_{j,n}^{int} = 2^{j/2}\psi_n^{right}(2^j \cdot), \qquad n = 2^j - p, \ldots, 2^j - 1.$$

The boundary wavelets and boundary scaling functions can be written as a finite linear combination of $\{\varphi(\cdot - n), \varphi(-\cdot + n) \; ; \; n \in \mathbb{Z}\}$. So, all the smoothness properties of the original scaling function also hold true for the boundary elements.

One can show that by choosing

$$V_j^{int} = \operatorname{span}\left\{\varphi_{j,n}^{int} \; ; \; n = 0, \ldots, 2^j - 1\right\}, \qquad W_j^{int} = \operatorname{span}\left\{\psi_{j,n}^{int} \; ; \; n = 0, \ldots, 2^j - 1\right\},$$

1. $\{\varphi_{j,n}^{int}\}_{n=0}^{2^j-1}$ is an orthonormal basis of V_j^{int}.

2. $V_j^{int} \subset V_{j+1}^{int}$.
3. $\lim_{j\to\infty} V_j^{int} = \overline{\bigcup_{j\geqslant \log_2(2p)} V_j^{int}} = L^2([0,1])$.
4. $V_j^{int} = V_{j-1}^{int} \oplus W_{j-1}^{int}$.
5. For any J with $2^J \geqslant 2p$, $L^2[0,1] = V_J^{int} \oplus \oplus_{j\geqslant J} W_j^{int}$. So,

$$\left[\{\varphi_{J,n}^{int}\}_{n=0}^{2^J-1}, \bigcup_{j\geqslant J} \{\psi_{j,n}^{int}\}_{n=0}^{2^j-1} \right]$$

   is an orthonormal basis of $L^2([0,1])$.

7 Wavelet bases in higher dimensions

For simplicity, we will discuss only the two-dimensional base. Higher dimensional cases can be treated in an analogous manner.

One way of constructing an orthonormal basis for $L^2(\mathbb{R}^2)$ starting from an orthonormal basis $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ is to consider the tensor product functions:

$$\Psi_{j_1,j_2,k_1,k_2}(x,y) = \psi_{j_1,k_1}(x)\psi_{j_2,k_2}(y).$$

Then,

$$\{\Psi_{j_1,j_2,k_1,k_2}\ ;\ j_1,j_2,k_1,k_2 \in \mathbb{Z}\}$$

is an orthonormal basis for $L^2(\mathbb{R}^2)$. Note that the two variables x and y are dilated independently.

An alternative construction which is more commonly used in practice is to dilate the two variables simultaneously. Instead of taking tensor product of the corresponding wavelet bases, we take tensor products of two 1D MRA's. More precisely, given an MRA $\{V_j\}_{j\in\mathbb{Z}}$, define for $j \in \mathbb{Z}$,

$$\mathbb{V}_j = V_j \otimes V_j = \{f(x)g(y)\ ;\ f,g \in V_j\}.$$

Then,

- $F \in \mathbb{V}_j$ if and only if $F(2\cdot,2\cdot) \in \mathbb{V}_{j+1}$.
- for all $j \in \mathbb{Z}$, $\mathbb{V}_j \subset \mathbb{V}_{j+1}$.
- $\bigcap_{j\in\mathbb{Z}} \mathbb{V}_j = \{0\}$ and $\overline{\bigcup_{j\in\mathbb{Z}} \mathbb{V}_j} = L^2(\mathbb{R}^2)$.

Finally, since $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0, for each $j \in \mathbb{Z}$, the functions

$$\Phi_{j,n_1,n_2}(x,y) = \varphi(2^j x - n_1)\varphi(2^j y - n_2), \qquad n_1,n_2 \in \mathbb{Z}$$

form an orthonormal basis of $\mathbb{V}_j$.

To construct the wavelet basis, just as in the 1D case, $\mathbb{W}_j$ is defined to be the orthogonal complement of $\mathbb{V}_j$ in $\mathbb{V}_{j+1}$. Observe that

$$\begin{aligned}\mathbb{V}_{j+1} &= V_{j+1} \otimes V_{j+1} = (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\ &= \underbrace{(V_j \otimes V_j)}_{\mathbb{V}_j} \oplus \underbrace{(W_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j)}_{\mathbb{W}_j}.\end{aligned}$$

So, we see that $\mathbb{W}_j$ consists of 3 parts and it has an orthonormal basis given by

$$\{\psi_{j,n_1}(x)\varphi_{j,n_2}(y), \varphi_{j,n_1}(x)\psi_{j,n_2}(y), \psi_{j,n_1}(x)\psi_{j,n_2}(y)\ ;\ n_1, n_2 \in \mathbb{Z}\}.$$

So, there are 3 generating wavelets (horizontal, vertical and diagonal)

$$\Psi^h(x,y) = \psi(x)\varphi(y), \quad \Psi^v(x,y) = \varphi(x)\psi(y), \quad \Psi^d(x,y) = \psi(x)\psi(y).$$

For a function $F \in L^2(\mathbb{R})$, by defining $F_{j,n_1,n_2}(x,y) = 2^j F(2^j x - n_1)F(2^j y - n_2)$, we have that

$$\left\{\Psi^\lambda_{j,n_1,n_2}\ ;\ \lambda \in \{h, v, d\}, j, n_1, n_2 \in \mathbb{Z}\right\},$$

forms an orthonormal basis of $L^2(\mathbb{R}^2)$.

Remark 6. The fast wavelet transform extends to higher dimensions for the separable wavelet basis described. In particular, define for $j \in \mathbb{N}$

$$a_j := (\langle f,\ \Phi_{j,n_1,n_2}\rangle)_{n_1,n_2=0}^{2^j-1} \in \mathbb{C}^{2^{2j}}$$

and for $\lambda \in \{h, v, d\}$,

$$d_j^\lambda := \left(\langle f,\ \Psi^\lambda_{j,n_1,n_2}\rangle\right)_{n_1,n_2=0}^{2^j-1} \in \mathbb{C}^{2^{2j}}.$$

Then one can check that the decomposition/reconstruction of a_j into/from $(a_{j-1}, d^h_{j-1}, d^v_{j-1}, d^d_{j-1})$ can be done in $O(2^{2j})$ elementary operations.

8 Linear approximation

The discretization of a compactly supported analogue function $f \in L^2[0,1]$ computes N samples

$$\{\langle f,\ \varphi_n\rangle\ ;\ n = 0, \ldots, N-1\}.$$

Typically, $\varphi_n = \varphi(\cdot - n)$ and forms a Riesz basis, the samples allows one to recover an approximation to f in the space

$$U_N = \operatorname{span}\{\varphi_n\ ;\ n = 0, \ldots, N-1\},$$

with approximation

$$f_N = \sum_{n=0}^{N-1} \langle f,\ \varphi_n\rangle \tilde{\varphi}_n,$$

where $\{\tilde{\varphi}_n\}_{n=0}^{N-1}$ is the biorthogonal basis. To compute the approximation error, we introduce an orthogonal basis $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$ of $L^2([0,1])$ such that $\{g_m\}_{m=0}^{N-1}$ is an orthonormal basis of U_N. Then,

$$f_N = \sum_{m=0}^{N-1} \langle f,\ g_m\rangle g_m,$$

and the linear approximation error is

$$\varepsilon_l(N, f) = \|f - f_N\|^2_{L^2} = \sum_{m=N}^{\infty} |\langle f,\ g_m\rangle|^2.$$

It is clear that since $f \in L^2[0,1]$, $\varepsilon_l(N,f) \to 0$ as $N \to \infty$. However, the rate of decay will depend on the decay of the coefficients $\langle f,\ g_n\rangle$:

Theorem 6. *Let $\{g_m\}_{m\in\mathbb{N}}$ be an orthonormal basis of $L^2[0,1]$. Suppose that for some $s > 1/2$,*

$$\sum_{m=0}^{\infty} m^{2s} \left|\langle f,\, g_m\rangle\right|^2 < \infty.$$

Then, there exist $A, B > 0$ such that

$$A\sum_{m=0}^{\infty} m^{2s} \left|\langle f,\, g_m\rangle\right|^2 \leqslant \sum_{N=0}^{\infty} N^{2s-1}\varepsilon_l(N,f) \leqslant B\sum_{m=0}^{\infty} m^{2s} \left|\langle f,\, g_m\rangle\right|^2.$$

It follows that $\varepsilon_l(N,f) = o(N^{-2s})$.

Proof. First observe that

$$\sum_{N=0}^{\infty} N^{2s-1}\varepsilon_l(N,f) = \sum_{N=0}^{\infty} N^{2s-1} \sum_{m=N}^{\infty} \left|\langle f,\, g_m\rangle\right|^2 = \sum_{m=0}^{\infty} \left|\langle f,\, g_m\rangle\right|^2 \sum_{N=0}^{m} N^{2s-1}.$$

For any $s > 1/2$,

$$Am^{2s} \leqslant \int_0^m x^{2s-1} \leqslant \sum_{N=1}^{m} N^{2s-1} \leqslant \int_1^{m+1} x^{2s-1} \leqslant Bm^{2s}.$$

This proves the first inequality.

To verify that $\lim_{N\to\infty} \varepsilon_l(N,f)N^{2s} = 0$, observe that $\varepsilon_l(m,f) \geqslant \varepsilon_l(N,f)$ for $m \leqslant N$. So,

$$\varepsilon_l(N,f)\sum_{m=N/2}^{N-1} m^{2s-1} \leqslant \sum_{m=N/2}^{N-1} m^{2s-1}\varepsilon_l(m,f) \leqslant \sum_{m=N/2}^{\infty} m^{2s-1}\varepsilon_l(m,f) \to 0, \qquad N\to\infty.$$

The conclusion follows because $\sum_{m=N/2}^{N-1} m^{2s-1} \geqslant CN^{2s}$. □

This theorem establishes that the linear approximation of f in a basis $\mathcal{B}$ decays faster than N^{-2s} if f belongs to

$$W_{\mathcal{B},s} = \left\{ f \in L^2[0,1]\,;\, \sum_{m=0}^{\infty} m^{2s} \left|\langle f,\, g_m\rangle\right|^2 < \infty \right\}.$$

In the following sections, we will show that in the case of Fourier or wavelet bases, this is in fact a Sobolev space.

8.1 Some function spaces

Definition 6. *Let $\Omega \subseteq \mathbb{R}^N$. The Sobolev space $W^{s,p}(\Omega)$ is the space of functions $f \in L^p(\Omega)$ such that for multi-index $\alpha = (\alpha_j)_{j=1}^N \in \mathbb{N}_0^N$, with $|\alpha| = \alpha_1 + \ldots + \alpha_N \leqslant s$, $D^\alpha f \in L^p(\Omega)$ exists in a weak sense, i.e.*

$$\int f\partial^\alpha\varphi = (-1)^{|\alpha|}\int D^\alpha f\varphi, \qquad \forall\varphi \in C_c^\infty(\Omega).$$

In the above, $\partial^\alpha\varphi = \frac{\partial^{|\alpha|}\varphi}{\partial x_1^{\alpha_1}\cdots x_N^{\alpha_N}}$.

$W^{s,p}(\Omega)$ is a Banach space equipped with the norm

$$\|f\|_{s,p} = \left(\sum_{|\alpha|\leqslant s}\int |D^\alpha f|^p\right)^{1/p}.$$

Note that any element in $W^{s,2}(\mathbb{R})$ is necessarily continuous: to give an informal argument, for any $f \in C_c^\infty$,

$$f(t) - f(s) = \int_s^t f'(r)\mathrm{d}r \leqslant \sqrt{t-s}\|f\|_{W^{s,2}},$$

so f is Hölder-1/2 continuous. This holds for all $f \in W^{s,2}$ since C_c^∞ is dense in $W^{s,2}$. So, Sobolev functions cannot have discontinuities.

To allow for functions such as $\mathbb{1}_{[0,1]}$, we consider instead the functions of bounded variations, which is an extension of $W^{1,1}$.

Definition 7. *Given $u \in L^1(\Omega)$, the total variation of u is*

$$\|u\|_V = \int_\Omega |Du| = \sup\left\{\langle \operatorname{div} z,\, u\rangle \;;\; z \in C_c^\infty(\Omega, \mathbb{R}^N), \|z\|_\infty \leqslant 1\right\}.$$

Definition 8. *The space $BV(\Omega)$ of functions of bounded variation is the set of functions $u \in L^1(\Omega)$ such that $\|u\|_V < \infty$ endowed with the norm $\|u\|_{BV} = \|u\|_{L^1} + \|u\|_V$.*

This space is a Banach space.

Examples

- if $u \in C^1(\Omega)$, then for all $z \in C_c^\infty(\Omega, \mathbb{R}^N)$, $\int_\Omega u \operatorname{div} z = -\int_\Omega \nabla u \cdot z$ and $\int_\Omega |Du| = \int_\Omega |\nabla u|$.
- Suppose that $\Omega = (-1,1)$, $u(x) = -1$ for $x \in (-1,0)$ and $u(x) = 1$ for $x \in [0,1)$. Then,

  $$\int_{-1}^1 uz' = -2z(0), \qquad \int_{-1}^1 |Du| = 2.$$

  In particular, Du is equal to the measure $2\delta_0$.

8.2 The Fourier case

Recall that $\left\{e^{-i2\pi m\cdot} \;;\; m \in \mathbb{Z}\right\}$ is an orthonormal basis of $L^2[0,1]$. Let

$$U_N = \left\{e^{-i2\pi m\cdot} \;;\; |m| \leqslant N/2\right\}.$$

Then,

$$\varepsilon_l(N, f) = \sum_{|m|>N/2} \left|\hat{f}(2\pi m)\right|^2.$$

Theorem 7. *Let $f \in L^2([0,1])$ with $\operatorname{Supp}(f) \subseteq (0,1)$. Then, $f \in W^{s,2}[0,1]$ if and only if*

$$\sum_{N=1}^\infty N^{2s} \frac{\varepsilon_l(N,f)}{N} < \infty. \tag{5}$$

So, $\varepsilon_l(N,f) = o(N^{-2s})$.

Proof. First, the Sobolev space H^s of periodic functions on $[0,1]$ is defined as

$$\left\{f \in L^2[0,1] \;;\; \sum_{m\in\mathbb{Z}} |m|^{2s} \left|\langle f,\, e^{i2\pi m\cdot}\rangle\right|^2 < \infty\right\}.$$

Since $\operatorname{Supp}(f) \subseteq (0,1)$, $f \in W^{s,2}$ if and only if $f \in H^s$, which in turn, is equivalent to (5) by Theorem 6.

□

Lecture 8

Theorem 8. *Let $f \in BV(0,1)$. Then, $\varepsilon_l(N,f) = O(\|f\|_V^2 N^{-1})$.*

Proof. If $\|f\|_V < \infty$, then

$$\left|\langle f,\, e^{i2\pi m\cdot}\rangle\right| = \Big|\int_0^1 f(u)e^{-i2\pi mu}\mathrm{d}u\Big| = \Big|\int_0^1 Df(u)\frac{e^{-i2\pi mu}}{-i2\pi m}\mathrm{d}u\Big| \leqslant \frac{\|f\|_V}{2\,|m|\,\pi}.$$

Thus,

$$\varepsilon_l(N,f) = \sum_{|m|>N/2} \left|\langle f,\, e^{i2\pi m\cdot}\rangle\right|^2 = \sum_{|m|>N/2} \frac{\|f\|_V^2}{4\,|m|^2\,\pi^2} = O(\|f\|_V^2 N^{-1}).$$

□

Remark 7. In general, one cannot expect faster decay than N^{-1} since if $f = \mathbb{1}_{[0,1/2]}$, then $\|f\|_V = 1$ and

$$\left|\langle f,\, e^{i2\pi m\cdot}\rangle\right| = \begin{cases} 0 & m \neq 0, \text{ even,} \\ 1/(\pi\,|m|) & m \text{ odd.} \end{cases}$$

So, $\varepsilon_l(N,f) \sim N^{-1}$.

8.3 The wavelet case

In this section, we will consider wavelet bases on the interval $[0,1]$ with

$$V_J = \operatorname{span}\left\{\varphi_{J,n}\ ;\ n = 0,\ldots,2^J-1\right\}, \qquad W_J = \operatorname{span}\left\{\psi_{J,n}\ ;\ n = 0,\ldots,2^J-1\right\}.$$

We will assume that $\psi \in C^q$ has compact support and q vanishing moments.

Let $N = 2^J$, and let $U_N := V_J$. Recall that $V_J = V_L \oplus \oplus_{j=L}^{J-1} W_j$ and the orthogonal projection of f onto V_J is

$$f_N = P_{V_J} f = \sum_{n=0}^{2^L-1} \langle f,\, \varphi_{L,n}\rangle\varphi_{L,n} + \sum_{j=L}^{J-1}\sum_{n=0}^{2^j-1} \langle f,\, \psi_{j,n}\rangle\psi_{j,n}.$$

The linear approximation error is

$$\varepsilon_l(N,f) = \sum_{j=J}^{\infty}\sum_{n=0}^{2^j-1} \left|\langle f,\, \psi_{j,n}\rangle\right|^2.$$

Theorem 9. *Let $s \in (0,q)$. $f \in L^2([0,1])$ is in $W^{s,2}([0,1])$ if and only if*

$$\sum_{j=J}^{\infty}\sum_{n=0}^{2^j-1} 2^{2sj}\left|\langle f,\, \psi_{j,n}\rangle\right|^2 < \infty. \tag{6}$$

Proof. We will not prove this result, but only give some intuition behind this result. First note that $f \in W^{s,2}(\mathbb{R})$ if and only if $\int |\omega|^{2s}\left|\hat{f}(\omega)\right|^2 \mathrm{d}\omega < \infty$. Note that the low-frequency part of this integral is always finite: For fixed J,

$$\int_{|\omega|\leqslant 2^J\pi} |\omega|^{2s}\left|\hat{f}(\omega)\right|^2 \mathrm{d}\omega \leqslant 2^{2Js}\int_{|\omega|\leqslant 2^J\pi}\left|\hat{f}(\omega)\right|^2 \mathrm{d}\omega < \infty.$$

Also, by Plancherel,

$$\sum_{j\geqslant J} 2^{2js}\sum_n \left|\langle f,\, \psi_{j,n}\rangle\right|^2 = \sum_{j\geqslant J}\sum_n \frac{2^{2js}}{4\pi^2}\left|\langle \hat{f},\, \hat{\psi}_{j,n}\rangle\right|^2.$$

So, it is enough to show that $g \in L^2$,

$$\int_{|\omega|>2^J\pi} |\omega|^{2s} |g(\omega)|^2 \,\mathrm{d}\omega < \infty \iff \sum_{j\geqslant J}\sum_n 2^{2js} \left|\langle g,\, \hat{\psi}_{j,n}\rangle\right|^2 < \infty.$$

Assume that ψ is bandlimited so that $\hat{\psi}$ is supported on $K = [-2\pi, -\pi] \cup [\pi, 2\pi]$ and $\inf_{\omega\in K} \left|\hat{\psi}(\omega)\right| > c > 0$. This is certainly true for the Shannon wavelet.

Now, if $\mathrm{Supp}(g) \subset [0, \infty)$, then

$$\begin{aligned}
\sum_n \left|\langle g,\, \hat{\psi}_{j,n}\rangle\right|^2 &= \sum_n \left| \int g(\omega)\hat{\psi}\left(\frac{\omega}{2^j}\right) e^{i\omega n/2^j} 2^{-j/2} \mathrm{d}\omega \right|^2 \\
&= \sum_n \left| \int_{2^j\pi}^{2^{j+1}\pi} g(\omega)\hat{\psi}\left(\frac{\omega}{2^j}\right)\sqrt{2\pi} \underbrace{e^{i\omega n/2^j}(2\pi 2^j)^{-1/2}}_{\text{Orthonormal basis element for } L^2[0,2^{j+1}\pi]} \mathrm{d}\omega \right|^2 \\
&= 2\pi \int_{2^j\pi}^{2^{j+1}\pi} |g(\omega)|^2 \left|\hat{\psi}(\frac{\omega}{2^j})\right|^2 \mathrm{d}\omega \sim \int_{2^j\pi}^{2^{j+1}\pi} |g(\omega)|^2 \,\mathrm{d}\omega.
\end{aligned}$$

Moreover, on $[2^j\pi, 2^{j+1}\pi]$, $|\omega| \sim 2^j$. So,

$$\sum_n 2^{2js} \left|\langle g,\, \hat{\psi}_{j,n}\rangle\right|^2 \sim \int_{2^j\pi}^{2^{j+1}\pi} |\omega|^{2js} |g(\omega)|^2 \,\mathrm{d}\omega.$$

Summing over j yields

$$\sum_{j\geqslant J}\sum_n 2^{2js} \left|\langle g,\, \hat{\psi}_{j,n}\rangle\right|^2 \sim \int_{|\omega|>2^J\pi} |\omega|^{2js} |g(\omega)|^2 \,\mathrm{d}\omega$$

The same argument can also be repeated for g with $\mathrm{Supp}(g) \subset (-\infty, 0]$.

To prove this result, one would extend this argument for bandlimited wavelets to more general wavelets.

□

Corollary 1. *Let $s \in (0, q)$. $f \in L^2([0,1])$ is in $W^{s,2}([0,1])$ if and only if*

$$\sum_{N=1}^{\infty} N^{2s} \frac{\varepsilon_l(N, f)}{N} < \infty.$$

Hence, $\varepsilon_l(N, f) = o(N^{-2s})$.

Proof. By writing $\psi_{j,n} = g_m$ with $m = 2^j + n$, (6) is equivalent to $\sum_{m=0}^{\infty} |m|^{2s} |\langle f,\, g_m\rangle|^2 < \infty$. The result then follows from Theorem 6. □

We see from the above result that the decay of wavelets characterize Sobolev spaces. We now show that wavelet coefficients can also be used to characterized Hölder spaces.

Definition 9. *f is uniformly Lipschitz-α over $[0,1]$ if there exists $K > 0$ such that for all $v \in [0,1]$, there exists a polynomial of degree $\lfloor \alpha \rfloor$, p_v, such that*

$$\forall t \in [0,1], \quad |f(t) - p_v(t)| \leqslant K\, |t - v|^{\alpha}.$$

The smallest such value of K is the homogeneous Hölder-α norm $\|f\|_{\tilde{C}^\alpha}$. The C^α-Hölder space are the functions such that

$$\|f\|_{C^\alpha} := \|f\|_{\tilde{C}^\alpha} + \|f\|_\infty < \infty.$$

Remark 8. The local polynomial approximation property is related to the differentiability of f. This can be seen through the Taylor expansion formula: If f is m-times continuously differentiable in $[v-h, v+h]$ and

$$p_v(t) = \sum_{k=0}^{m-1} \frac{f^{(k)}(v)}{k!}(t-v)^k$$

is the Taylor polynomial of f in $[v-h, v+h]$, then

$$|f(t) - p_v(t)| \leqslant \frac{|t-v|^m}{m!} \sup_{u\in[v-h,v+h]} \left| f^{(m)}(u) \right|.$$

In fact:

1. If f is uniformly Lipschitz α with $\alpha = m + \varepsilon$, $\varepsilon \in (0,1)$, then f is necessarily m-times continuously differentiable and $\left|f^{(m)}(t) - f^{(m)}(y)\right| \leqslant K\, |t-y|^{\varepsilon}$.
2. At each v, the polynomial $p_v(t)$ is uniquely defined. If f is $m = \lfloor \alpha \rfloor$ times continuously differentiable in a neighborhood of v, then p_v is the Taylor expansion of f at v.
3. A word of caution: Note that saying that f is uniformly Lipschitz-n for $n \in \mathbb{N}$ does not imply that f is n times continuously differentiable. It simply means that $f \in L^\infty$ is $n-1$-times continuously differentiable and there exists a constant K such that

$$\left| f^{(n-1)}(x) - f^{(n-1)}(y) \right| \leqslant K\, |x-y|\,.$$

   For example, $f(x) = |x|$ is Lipschitz-1, but not differentiable.

Proposition 4. *Let $\lfloor \alpha \rfloor < q$, where q is the number of vanishing moments of our wavelet. Then, there exists $B \geqslant A > 0$ such that*

$$A\|f\|_{\tilde{C}^\alpha} \leqslant \sup\left\{ 2^{j(\alpha+1/2)} \left|\langle f,\, \psi_{j,n}\rangle\right| \;;\; j \in \mathbb{Z}, n = 0, \dots, 2^j - 1 \right\} \leqslant B\|f\|_{\tilde{C}^\alpha}.$$

Proof. We will only prove the upper bound, a proof of the lower bound can be found in [Mallat, Ch 9]: Let $p_{2^{-j}n}$ be the polynomial as described in Definition 9. Since $\psi_{j,n}$ has $q > \lfloor \alpha \rfloor$ vanishing moments, $\langle \psi_{j,n},\, p_{2^{-j}n}\rangle = 0$. So,

$$\begin{aligned}
\langle f,\, \psi_{j,n}\rangle &= |\langle f - p_{2^{-j}n},\, \psi_{j,n}\rangle| \\
&= \int (f(t) - p_{2^{-j}n}(t))\overline{\psi_{j,n}(t)}\mathrm{d}t \\
&\leqslant \|f\|_{\tilde{C}^\alpha} \int \left|t - 2^{-j}n\right|^\alpha \left|\psi(2^j t - n)\right| 2^{j/2}\mathrm{d}t \\
&= \|f\|_{\tilde{C}^\alpha} \int \left|x 2^{-j}\right|^\alpha |\psi(x)|\, 2^{-j/2}\mathrm{d}x \\
&= 2^{-j(\alpha+1/2)}\|f\|_{\tilde{C}^\alpha} \int |x|^\alpha\, |\psi(x)|\, \mathrm{d}x,
\end{aligned}$$

where we have used the fact that $\int |x|^\alpha\, |\psi(x)|\, \mathrm{d}x < \infty$ since ψ has compact support. □

Corollary 2. *If f is uniformly Lipschitz-α on $[0,1]$, then $\varepsilon_l(N, f) = \mathrm{O}(\|f\|^2_{\tilde{C}^\alpha} N^{-2\alpha})$.*

9 Nonlinear approximation

Given a basis $\{g_m\}_{m\in\mathbb{N}}$ of $L^2([0,1])$, the analogue to digital conversion of $f \in L^2([0,1])$ yields N samples and the linear approximation

$$f_N = \sum_{m=0}^{N-1} \langle f,\, g_m\rangle g_m.$$

We can further approximate f_N by a sparse representation: Let $\Lambda \subset [N] := \{0,\ldots,N-1\}$ index the largest (in absolute value) M coefficients of $\{\langle f,\, g_m\rangle\}_{m=0}^{N-1}$. The M-term nonlinear approximation is

$$f_\Lambda = \sum_{m\in\Lambda} \langle f,\, g_m\rangle g_m$$

and $\|f_N - f_\Lambda\|_{L^2}^2 = \sum_{m\in[N]\setminus\Lambda} |\langle f,\, g_m\rangle|^2$. The overall nonlinear approximation error is then

$$\varepsilon_n(M,f) = \|f - f_\Lambda\|^2 = \|f - f_N\|^2 + \|f_N - f_\Lambda\|^2.$$

Let T be such that

$$\Lambda = \{m \in [N]\ ;\ |\langle f,\, g_m\rangle| \geqslant T\}\,,$$

If N is sufficiently large such that all coefficients with absolute value above T are in the first N coefficients,

$$T \geqslant \max_{m\geqslant N} |\langle f,\, g_m\rangle| \quad \text{or equivalently} \quad N \geqslant \mathrm{argmax}_m\{|\langle f,\, g_m\rangle| \geqslant T\}, \tag{7}$$

then $\varepsilon_n(M,f) = \min\left\{\|f - f_\Lambda\|^2\ ;\ \Lambda \subset N, |\Lambda| = M\right\}$. In particular, if $|\langle f,\, g_m\rangle| \leqslant Cm^{-\beta}$ and $\beta > 0$, then choosing $N \geqslant C^{1/\beta}T^{-1/\beta}$ ensures (7).

Henceforth, given some orthonormal basis $\mathcal{B} = \{g_m\}_{m=0}^{\infty}$, we will study

$$\varepsilon_n(M,f) := \sum_{m\notin\Lambda} |\langle f,\, g_m\rangle|^2\,,$$

where Λ is of cardinality M and indexes the largest M coefficients of $\{|\langle f,\, g_m\rangle|\}_m$.

Lecture 9

9.1 Properties of the nonlinear approximation error

Notation Let $f_{\mathcal{B}}^r[k] = \langle f,\, g_{m_k}\rangle$ be the coefficient of rank k:

$$|f_{\mathcal{B}}^r[k]| \geqslant |f_{\mathcal{B}}^r[k+1]|\,, \qquad k > 0,$$

and nonlinear approximation and nonlinear approximation error are respectively

$$f_M = \sum_{k=1}^{M} f_{\mathcal{B}}^r[k] g_{m_k}, \qquad \varepsilon_n(M,f) = \sum_{k=M+1}^{\infty} |f_{\mathcal{B}}^r[k]|^2\,.$$

Theorem 10. *Let $s > 1/2$. If there exists $C > 0$ such that $|f_{\mathcal{B}}^r[k]| \leqslant Ck^{-s}$, then*

$$\varepsilon_n(M,f) \leqslant \frac{C^2}{2s-1} M^{1-2s}. \tag{8}$$

Conversely, if $\varepsilon_n(M,f)$ satisfies (8), then

$$|f_{\mathcal{B}}^r[k]| \leqslant \left(1 - \frac{1}{2s}\right)^{-s} Ck^{-s}$$

Proof. For the first assertion,

$$\varepsilon_n(M, f) = \sum_{k=M+1}^{\infty} |f^r_{\mathcal{B}}[k]|^2 \leqslant \sum_{k=M+1}^{\infty} C^2 k^{-2s} \leqslant C^2 \int_M^{\infty} x^{-2s} \leqslant \frac{C^2 M^{1-2s}}{2s-1}.$$

For the converse statement, let $\alpha < 1$. Then,

$$\varepsilon_n(\lfloor \alpha M \rfloor, f) \geqslant \sum_{k=\lfloor \alpha M \rfloor+1}^{M} |f^r_{\mathcal{B}}[k]|^2 \geqslant M(1-\alpha)\, |f^r_{\mathcal{B}}[M]|^2 .$$

So,

$$|f^r_{\mathcal{B}}[M]|^2 \leqslant M^{-1}(1-\alpha)^{-1} \frac{C^2}{2s-1} (\alpha M)^{1-2s} = \frac{\alpha C^2}{(1-\alpha)(2s-1)} (\alpha M)^{-2s}$$

Choosing $\alpha = 1 - 1/2s$ and $M = k$ yields the required result. □

Notation: For $p \geqslant 1$, let $\|f\|_{\mathcal{B},p} = \left(\sum_{m\in\mathbb{N}} |\langle f,\, g_m\rangle|^p\right)^{1/p}$.

Theorem 11. *Let $p \in [1, 2)$. If $\|f\|_{\mathcal{B},p} < \infty$, then*

$$|f^r_{\mathcal{B}}[k]| \leqslant \|f\|_{\mathcal{B},p} k^{-1/p},$$

and $\varepsilon_n(M, f) = o(M^{1-2/p})$.

Proof. To prove the inequality:

$$\|f\|^p_{\mathcal{B},p} = \sum_{m\in\mathbb{N}} |f^r_{\mathcal{B}}[m]|^p \geqslant \sum_{m=1}^{k} |f^r_{\mathcal{B}}[m]|^p \geqslant k\, |f^r_{\mathcal{B}}[k]|^p .$$

To show that $\lim_{M\to\infty} \varepsilon_n(M, f) M^{2/p-1} = 0$, let

$$S[k] = \sum_{n=k}^{2k-1} |f^r_{\mathcal{B}}[k]|^p \geqslant k\, |f^r_{\mathcal{B}}[2k]|^p .$$

Then,

$$\begin{aligned}\varepsilon_n(M, f) = \sum_{k=M+1}^{\infty} |f^r_{\mathcal{B}}[k]|^2 &\leqslant \sum_{k=M+1}^{\infty} S[k/2]^{2/p} (k/2)^{-2/p} \\ &\leqslant \sup_{k>M/2} S[k/2]^{2/p} \sum_{k=M+1}^{\infty} (k/2)^{-2/p}.\end{aligned}$$

Since $\|f\|^p_{\mathcal{B},p} < \infty$, $S[k/2] \to 0$ as $k \to \infty$ and

$$\sum_{k=M+1}^{\infty} (k/2)^{-2/p} \leqslant 2^{2/p} \frac{M^{-2/p+1}}{2/p-1}.$$

□

Remark 9. This theorem says that the functions in

$$\{f \in H \ ;\ \|f\|_{\mathcal{B},p} < \infty\}$$

are well approximated by their nonlinear approximation in $\mathcal{B}$.

9.2 Nonlinear approximation for wavelet bases

The following theorem shows that in practice, when the digital to analogue conversion of signal is in V_L for some L sufficiently large, the nonlinear approximation obtained from sparsifying the digital signal corresponds to approximation of the signal with the largest M coefficients with respect to all coefficients.

Theorem 12. *Let $T > 0$ and suppose that f is bounded. If j, n are such that $|\langle f, \psi_{j,n}\rangle| > T$, then $\psi_{j,n} \in V_L$ where*

$$N = 2^L = O(\|f\|_\infty^2 T^{-2}).$$

Proof. For each interior wavelet,

$$|\langle f, \psi_{j,n}\rangle| = \left| \int_0^1 f(t) 2^{j/2} \psi(2^j t - n) \mathrm{d}t \right| \leqslant \int_{-n}^{2^j - n} f\left(\frac{x+n}{2^j}\right) 2^{-j/2} \psi(x) \mathrm{d}x \leqslant \|f\|_\infty 2^{-j/2} \|\psi\|_1.$$

Similarly, for each boundary wavelet (note that there is a fixed number of these at each scale)

$$|\langle f, \psi_{j,n}\rangle| \leqslant 2^{-j/2} \|\psi_n^b\|_1 \|f\|_\infty,$$

where ψ_n^b is one of the boundary wavelets.

So, if $|\langle f, \psi_{j,n}\rangle| > T$, then $2^{j/2} \leqslant A\|f\|_\infty T^{-1}$, where the constant A depends only on the wavelet ψ and the fixed number of boundary wavelets ψ_n^b. □

Approximation of piecewise regular signals

Theorem 13. *If f has K discontinuities on $[0, 1]$ and is uniformly Lipschitz-α between these discontinuities with $\alpha \in (1/2, q)$, then*

$$\varepsilon_l(M, f) = O(K\|f\|_{C^\alpha} M^{-1}) \quad \text{and} \quad \varepsilon_n(M, f) = O(\|f\|_{C^\alpha} M^{-2\alpha}).$$

Proof. Let Δ denote the set of points where f is discontinuous. For $j \in \mathbb{N}$, let

$$\mathrm{I}_j := \{n \,;\, \mathrm{Supp}(\psi_{j,n}) \cap \Delta \neq \emptyset\},$$

and we will call the wavelets indexed by I_j type I wavelets of scale j. Let

$$\mathrm{II}_j := \{n \,;\, \mathrm{Supp}(\psi_{j,n}) \cap \Delta = \emptyset\},$$

and we will call the wavelets in II_j type II wavelets of scale j.

To deal with the linear approximation error, we first bound the number and coefficient magnitude of type I and type II wavelets at each scale: Since $\|f\|_\infty \leqslant \|f\|_{C^\alpha}$, by Theorem 12,

$$|\langle f, \psi_{j,n}\rangle| \leqslant B_1 \|f\|_{C^\alpha} 2^{-j/2}. \tag{9}$$

Moreover, if $C = |\mathrm{Supp}(\psi)|$, then $2^{-j}C = |\mathrm{Supp}(\psi_{j,n})|$ and $|\mathrm{I}_j| \leqslant KC$.

For type II wavelets of scale j, by Proposition 4,

$$|\langle f, \psi_{j,n}\rangle| \leqslant B_2 2^{-j(\alpha+1/2)} \|f\|_{\tilde{C}_\alpha} \tag{10}$$

and there are at most 2^j such wavelets.

Let $M = 2^J$, then

$$\begin{aligned}
\varepsilon_l(M,f) &= \sum_{j=J}^{\infty}\sum_{n=0}^{2^j-1} |\langle f,\, \psi_{j,n}\rangle|^2 = \sum_{j=J}^{\infty}\sum_{n\in \mathrm{I}_j} |\langle f,\, \psi_{j,n}\rangle|^2 + \sum_{j=J}^{\infty}\sum_{n\in \mathrm{II}_j} |\langle f,\, \psi_{j,n}\rangle|^2 \\
&= \sum_{j=J}^{\infty} KCB_1^2 \|f\|_{C_\alpha}^2 2^{-j} + \sum_{j=J}^{\infty} B_2^2 2^{-j(2\alpha+1)} \|f\|_{C_\alpha}^2 2^j \\
&\leqslant 2^{-J+1} KCB_1^2 \|f\|_{C_\alpha}^2 + B_2^2 \|f\|_{C_\alpha}^2 2^{-2J\alpha}(1-2^{-2\alpha})^{-1} \lesssim M^{-1} K \|f\|_{C^\alpha}^2.
\end{aligned}$$

Note that since $\alpha > 1/2$, the error above is dominated by the error due to the Type I wavelets.

For the nonlinear approximation error, we consider type I and type II coefficients separately and consider the decay of rank r coefficients of each type: Order the coefficients $\{\langle f,\, \psi_{j,n}\rangle : n \in \mathrm{I}_j, j \in \mathbb{N}\}$ in decreasing order of magnitude and let $f^r_{\mathcal{B},I}[k] = \langle f,\, \psi_{j_k,n_k}\rangle$ denote the kth largest coefficient in magnitude. Similarly, order the coefficients $\{\langle f,\, \psi_{j,n}\rangle : n \in \mathrm{II}_j, j \in \mathbb{N}\}$ in decreasing order of magnitude and let $f^r_{\mathcal{B},II}[k] = \langle f,\, \psi_{j_k,n_k}\rangle$ denote the kth largest coefficient in magnitude.

There are at most jKC type I wavelets at scales less than j, and given any type I wavelet at scale $l \geqslant j$, (9) holds. Therefore

$$f^r_{\mathcal{B},I}[jKC] \leqslant B_1 \|f\|_{C^\alpha} 2^{-j/2} \implies f^r_{\mathcal{B},I}[m] \lesssim \|f\|_{C^\alpha} 2^{-m/(2KC)}.$$

On the other hand, there are at most 2^j type II wavelets at scales less than j and given any type II wavelet at scale $l \geqslant j$, (10) holds. So,

$$f^r_{\mathcal{B},II}[2^j] \leqslant B_2 2^{-j(\alpha+1/2)} \|f\|_{\tilde{C}_\alpha} \implies f_{\mathcal{B},II}[m] \lesssim m^{-(\alpha+1/2)} \|f\|_{\tilde{C}_\alpha}$$

Since type I wavelets yield much faster decay, we have that

$$f^r_{\mathcal{B}}[m] = O(m^{-(\alpha+1/2)} \|f\|_{\tilde{C}_\alpha}),$$

and Theorem 10 yields the required result.

□

Remark 10. Notice that the error from Type I wavelets (i.e. error arising from the discontinuities) dominates the linear approximation error, but becomes negligible in the nonlinear approximation error.

9.3 Besov and BV spaces

Definition 10 (Modulus of smoothness)**.** *Let $T_h f = f(\cdot + h)$ for $h > 0$ and let $\Delta_h^r := (T_h - I)^r$ be the r^{th} difference operator with step h. Given f defined on $\Omega \subset \mathbb{R}$, and $p \in [1,\infty]$,*

$$\omega_r(f,t)_p := \sup_{|h|\leqslant t} \|\Delta_h^r f\|_{L_p(\Omega)}$$

is the r^{th} modulus of smoothness of f.

Remark 11.

1. $\omega_r(f,t)_p$ is non-decreasing in t.
2. $\omega_r(f,t)_p \to 0$ as $t \to 0$ for $f \in L^p(\Omega)$ if $p \in [1,\infty)$ and for f uniformly continuouos if $p = \infty$. This is clear for the uniformly continuous case. To see this for $p \in [1,\infty)$, the convergence to 0 is clear if $f \in C_c(\Omega)$ is continuous and compactly supported, and note that $C_c(\Omega)$ is dense in L^p. The faster the convergence to 0, the smoother f is.

The Besov spaces allows us to characterize different smoothness spaces using this modulus of smoothness:

Definition 11 (Besov spaces)**.** *Let $\alpha > 0$ and let $\beta, \gamma \in [1, \infty]$. Let $r = [\alpha]+1$ be the smallest integer greater than α. The homogeneous Besov norm is*

$$\|f\|^*_{\alpha,\beta,\gamma} := \begin{cases} \left(\int_0^\infty [t^{-\alpha}\omega_r(f,t)_\beta]^\gamma \frac{\mathrm{d}t}{t}\right)^{1/\gamma} & \gamma \in [1,\infty) \\ \sup_{t>0} t^{-\alpha}\omega_r(f,t)_\beta & \gamma = \infty. \end{cases}$$

This is a seminorm since $\|f\|^*_{\alpha,\beta,\gamma} = 0$ for constant functions. The Besov space with parameters $\alpha > 0$, $\beta, \gamma \in [1, \infty]$, denoted by $\mathbb{B}^\alpha_{\beta,\gamma}$, is the set of functions such that

$$\|f\|_{\alpha,\beta,\gamma} := \|f\|_{L^\beta} + \|f\|^*_{\alpha,\beta,\gamma} < \infty.$$

Remark 12. α relates to the smoothness and β tells us the space f and its derivatives live in. γ is an extra fine-tuning parameter and is such that

$$\mathbb{B}^\alpha_{\beta,\gamma_1} \subset \mathbb{B}^\alpha_{\beta,\gamma_2}, \qquad \gamma_1 < \gamma_2,$$

however, there is in general not much distinction between these spaces and γ is less important.

Theorem 14. *Let φ and ψ be a compactly supported scaling function and wavelet generating an orthonormal wavelet basis on $L^2[0, 1]$, i.e.*

$$\{\varphi_{J,n}\}_{n=0}^{2^J-1} \cup \{\psi_{j,n}\}_{\substack{j\geqslant J \\ n=0,\ldots,2^j-1}}.$$

Suppose that ψ is q times differentiable and has q vanishing moments. Then, for $s \in (0, q)$, by denoting $\psi_{J-1,n} := \varphi_{J,n}$,

$$\|f\|_{s,\beta,\gamma} \sim \left[\sum_{j=J-1}^{\infty} \left(2^{j(s+1/2-1/\beta)} \left(\sum_{n=0}^{2^j-1} |\langle f,\, \psi_{j,n}\rangle|^\beta\right)^{1/\beta}\right)^\gamma\right]^{1/\gamma},$$

and the homogeneous Besov norm is obtained by ignoring the scaling coefficients, that is:

$$\|f\|^*_{s,\beta,\gamma} \sim \left[\sum_{j=J}^{\infty} \left(2^{j(s+1/2-1/\beta)} \left(\sum_{n=0}^{2^j-1} |\langle f,\, \psi_{j,n}\rangle|^\beta\right)^{1/\beta}\right)^\gamma\right]^{1/\gamma},$$

- The space $\mathbb{B}^s_{\beta,\gamma}$ does not depend on the particular choice of the wavelet basis, provided that ψ has $q > s$ vanishing moments and is in C^q.
- Removing the first term involving the coarse scale coefficients gives the homogeneous Besov norm, which we denote by $\|f\|^*_{s,\beta,\gamma}$.
- For $\beta = \gamma = 2$, from Theorem 9, $\mathbb{B}^s_{2,2}([0, 1]) = W^{s,2}([0, 1])$, and by Corollary 1, elements in this space have a linear approximation error of $\varepsilon_l(N, f) = o(N^{-2s})$. Note that there exists elements $f \in W^{s,2}([0, 1])$ for which $\varepsilon_l(N, f) \sim \varepsilon_n(N, f)$.
- For $\beta = \gamma = \infty$, we saw from Proposition 4 that

  $$\|f\|^*_{s,\infty,\infty} = \sup_{j\geqslant J, n\in\mathbb{Z}} 2^{j(s+1/2)} |\langle f,\, \psi_{j,n}\rangle| \sim \|f\|_{\tilde{C}^s},$$

  is equivalent to the homogeneous Hölder-s norm.
- For $\beta < 2$, the functions in $\mathbb{B}^\alpha_{\beta,\gamma}$ are not necessarily uniformly regular, and for such functions, nonlinear approximations have significant improvements over linear approximations. The following proposition sheds further light on this remark.

Proposition 5. *Let $p \in [1, 2)$ and let $\alpha = 1/p - 1/2$. Then, given any $f \in \mathbb{B}^\alpha_{p,p}$, $\varepsilon_n(N, f) = o(N^{1-2/p})$. Moreover, the space $\mathbb{B}^\alpha_{p,p}$ contains functions with discontinuities.*

Proof. For our choice of p and α,

$$\|f\|_{\alpha,p,p} \sim \left(\sum_{j \geqslant J-1} \sum_{n} |\langle f,\, \psi_{j,n} \rangle|^p \right)^{1/p}.$$

From Theorem 11, $\varepsilon_n(N, f) = o(N^{1-2/p})$.

Suppose that f is uniformly Lipschitz-s between a finite number of dicontinuities. Then, from the proof of Theorem 13, we have that

$$|f^r_{\mathcal{B}}[k]| = O(k^{-s-1/2}).$$

So, if $p \in (1/(s+1/2), 2)$, then $f \in \mathbb{B}^{\alpha}_{p,p}$, so $\mathbb{B}^{\alpha}_{p,p}$ contains functions which are not necessarily regular at all points. $\square$

The following theorem shows that

$$\mathbb{B}^1_{1,1} \subset BV([0,1]) \subset \mathbb{B}^1_{1,\infty}.$$

Theorem 15. *Suppose that $\|\psi\|_V < \infty$. Then, there exists $A, B > 0$ such that*

$$A\|f\|^*_{1,1,\infty} \underbrace{\leqslant}_{(a)} \|f\|_V \underbrace{\leqslant}_{(b)} B\|f\|_{1,1,1}.$$

Proof. We first prove the upper bound: By writing out the wavelet expansion of f,

$$f = \sum_{n=0}^{2^J-1} \langle f,\, \varphi_{J,n} \rangle \varphi_{J,n} + \sum_{j=J}^{\infty} \sum_{n=0}^{2^j-1} \langle f,\, \psi_{j,n} \rangle \psi_{j,n},$$

we see that

$$\|f\|_V \leqslant \sum_{n=0}^{2^J-1} |\langle f,\, \varphi_{J,n} \rangle|\, \|\varphi_{J,n}\|_V + \sum_{j=J}^{\infty} \sum_{n=0}^{2^j-1} |\langle f,\, \psi_{j,n} \rangle|\, \|\psi_{j,n}\|_V.$$

Now, for interior wavelets $\psi_{j,n}$,

$$\|\psi_{j,n}\|_V = \int_0^1 \left| 2^{j/2} \psi'(2^j t - n) 2^j \right| \mathrm{d}t = 2^{j/2} \|\psi\|_V,$$

Similarly, $\|\varphi_{J,n}\|_V = 2^{J/2}\|\varphi\|_V$. Also, each boundary element is of the form $2^{j/2}\psi^b_n(2^j\cdot)$, we have that

$$\|\psi_{j,n}\|_V \leqslant 2^{j/n} \|\psi^b_n\|_V.$$

Note that there is a fixed number of boundary elements, say p, at each scale. So, the upper bound holds with A being the maximum of

$$\max\{\max\{\|\psi^b_n\|_V\}^q_{n=1}, \|\varphi\|_V, \|\psi\|_V\}.$$

For the lower bound, let $\theta = \int_{-\infty}^{x} \psi(t)\mathrm{d}t$. Then, $\theta' = \psi$ and if we assume that ψ has support on $[-K, K]$, then since $\int \psi = 0$, θ has also support in $[-K, K]$. Also, each of the p boundary elements ψ^b_n is supported on $[0, 1]$ and $\int_0^1 \psi^b_n = 0$. By defining $\theta^b_n(x) = \int_0^x \psi^b_n(t)\mathrm{d}t$, we have that θ^b_n has support in $[0, 1]$, ${\theta^b_n}' = \psi^b_n$ and $\theta^b_n(0) = \theta^b_n(1) = 0$.

To simplify notation, we simply write the wavelets as if they were all generated from a single wavelet ψ. Then,

$$\sum_{n=0}^{2^j-1}|\langle f,\,\psi_{j,n}\rangle| = \sum_{n=0}^{2^j-1}\Big|\int_0^1 f(t)2^{j/2}\psi(2^jt-n)\mathrm{d}t\Big| = \sum_{n=0}^{2^j-1}\Big|\int_0^1 Df(t)2^{-j/2}\theta(2^jt-n)\mathrm{d}t\Big|$$
$$\leqslant \sum_{n=0}^{2^j-1} 2^{-j/2}\int_0^1 |Df(t)|\,\big|\theta(2^jt-n)\big|\,\mathrm{d}t \leqslant \sum_{n=0}^{2^j-1} 2^{-j/2}\|\theta\|_\infty \int_{\mathrm{Supp}(\theta_{j,n})} |Df(t)|\,\mathrm{d}t.$$

Note that the second equality follows via integration by part, and there is no boundary term since the wavelet $\theta_{j,n}$ is supported on $[0,1]$ and is zero at 0 and 1. Note that $\theta_{j,n}$ has support in $[(-K+n)/2^j, (K+n)/2^j]$ and

$$\sum_{n=0}^{2^j-1}\|\theta\|_\infty\int_{\mathrm{Supp}(\theta_{j,n})}|f'(t)|\,\mathrm{d}t \leqslant 2K\|\theta\|_\infty\|f\|_V.$$

Since this is true for all j, we have that

$$\|f\|_V \geqslant \frac{1}{2K}\sup_j 2^{j/2}\sum_{n=0}^{2^j-1}|\langle f,\,\psi_{j,n}\rangle| = \frac{1}{2K}\|f\|^*_{1,1,\infty}.$$

□

Theorem 16. *Let $f \in BV([0,1])$. For $M > 2q$,*

$$\varepsilon_l(M,f) = O(\|f\|_V^2 M^{-1}) \quad \textit{and} \quad \varepsilon_n(M,f) = O(\|f\|_V^2 M^{-2}).$$

Proof. For the linear approximation rate, note that there exists 2^j wavelet coefficients at scale 2^j and in V_L, there are $\sum_{j=0}^{L-1} 2^j = 2^L$ wavelets. So,

$$\varepsilon_l(2^L,f) = \sum_{j=L}^{\infty}\sum_{n=0}^{2^j-1}|\langle f,\,\psi_{j,n}\rangle|^2.$$

From Theorem 15,

$$\sum_{n=0}^{2^j-1}|\langle f,\,\psi_{j,n}\rangle|^2 \leqslant \left(\sum_{n=0}^{2^j-1}|\langle f,\,\psi_{j,n}\rangle|\right)^2 \leqslant A^{-2}2^{-j}\|f\|_V^2.$$

Therefore,

$$\varepsilon_l(2^L,f) \leqslant \sum_{j=L}^{\infty} A^{-2}2^{-j}\|f\|_V^2 = O(\|f\|_V^2 2^{-L}),$$

as required.

For the nonlinear approximation error, let $f^r_{\mathcal{B}}[k]$ denote the wavelet coefficient of rank k, excluding the scaling coefficients $\langle f,\,\varphi_{J,n}\rangle$ (since their values cannot be controlled by $\|f\|_V$). Suppose for now that

$$|f^r_{\mathcal{B}}[k]| \leqslant B_0\|f\|_V k^{-3/2}, \qquad \forall f \in BV([0,1]). \tag{11}$$

Then, we have the required conclusion:

$$\varepsilon_n(M,f) = \sum_{k=M-2^J+1}^{\infty}|f^r_{\mathcal{B}}[k]|^2 \lesssim \|f\|_V^2 M^{-2}.$$

It remains to prove (11). For each $j \in \mathbb{N}$, let $f^r_{\mathcal{B}}[j,k]$ denote the wavelet coefficient of rank k at scale j. Then, by Theorem 15,

$$\sum_{n=0}^{2^j-1} |\langle f,\, \psi_{j,n}\rangle| \leqslant A^{-1} 2^{-j/2} \|f\|_V.$$

So, we may apply Theorem 11 to deduce that

$$|f^r_{\mathcal{B}}[j,k]| \leqslant A^{-1} 2^{-j/2} \|f\|_V k^{-1} =: C 2^{-j/2} k^{-1}.$$

So, at scale j, the number of coefficients k_j with absolute value greater than or equal to T satisfy

$$k_j \leqslant \min\{C 2^{-j/2} T^{-1}, 2^j\}$$

So, the total number of coefficients with absolute value at least T is

$$\begin{aligned}
\sum_{j \geqslant J} k_j &\leqslant \sum_{j \geqslant J} \min\{C 2^{-j/2} T^{-1}, 2^j\} \\
&= \sum_{2^j \leqslant C^{2/3} T^{-2/3}} 2^j + \sum_{C^{2/3} T^{-2/3} < 2^j} C 2^{-j/2} T^{-1} \leqslant 2 C^{2/3} T^{-2/3} + C T^{-1} \frac{C^{-1/3} T^{1/3}}{1 - 2^{-1/2}} \\
&\leqslant 6 (C T^{-1})^{2/3} = 6 A^{-2/3} \|f\|_V^{2/3} T^{-2/3}.
\end{aligned}$$

So, if $T = |f^r_{\mathcal{B}}[k]|$, then there are at least k coefficients with absolute value at least T and therefore, it follows that

$$|f^r_{\mathcal{B}}[k]| \leqslant 6^{3/2} A^{-1} \|f\|_V k^{-3/2}.$$

$\square$

9.4 Remarks for the 2 dimensional case

We make some remarks about approximation with wavelet bases in two dimensions. More details can be found in Chapter 9 of *A wavelet tour of signal processing: the sparse way.* In 2D, again considering wavelet bases of $L^2[0,1]^2$ with q vanishing moments and C^q:

1. f is uniformly Lipschitz-α with $\alpha \in (0,q)$ implies that $\varepsilon_l(N,f) = \mathcal{O}(\|f\|_{\tilde{C}_\alpha} N^{-\alpha})$. This is in fact the optimal decay rate for such functions.
2. If $f \in BV \cap L^\infty$, then $\varepsilon_l(N,f) = \mathcal{O}(\|f\|_V \|f\|_\infty N^{-1/2})$.
3. If $f \in BV$, then $\varepsilon_n(N,f) = \mathcal{O}(\|f\|_V^2 N^{-1})$.
4. If f is piecewise C^α, then $\varepsilon_n(N,f) = \mathcal{O}(N^{-1})$. This is different from what happend in the 1D case. In 1D, given K discontinuities, there are $\mathcal{O}(1)$ coefficients at each scale whose support intersect this discontinuity and when considering the nonlinear error, the error inccured by these coefficients become insignificant. On the other hand, in 2D, one can check that if f is discontinuous along some curve in the plane, then there will be $\mathcal{O}(2^j)$ wavelets at scale j whose support intersects this curve. Thus, for the approximation of piecewise C^α functions, wavelets are no longer optimal.

A Some terminology and results from Analysis

- Let H be a linear space over either $\mathbb{R}$ or $\mathbb{C}$. A scalar product on H is a function $\langle \cdot,\, \cdot \rangle$ from $\mathcal{H} \times \mathcal{H}$ into scalars such that

$$\begin{aligned}
\langle x,\, y \rangle &= \overline{\langle y,\, x \rangle} \\
\langle \alpha x_1 + \beta x_2,\, y \rangle &= \alpha \langle x_1,\, y \rangle + \beta \langle x_2,\, y \rangle \\
\langle x,\, x \rangle &\geqslant 0 \\
\langle x,\, x \rangle &= 0 \iff x = 0.
\end{aligned}$$

  In this case, $\|x\| = \sqrt{\langle x,\, x \rangle}$ is a norm, i.e.

$$\begin{aligned}
\|x + y\| &\leqslant \|x\| + \|y\| \\
\|\alpha x\| &= |\alpha|\, \|x\| \\
\|x\| &= 0 \iff x = 0.
\end{aligned}$$

  A linear space $\mathcal{H}$ equipped with a scalar product is called a Hilbert space if $\mathcal{H}$ is complete as a metric space with the metric $d(x, y) = \|x - y\|$.

- $L^2(\mathbb{R})$ is a Hilbert space with inner product $\langle f,\, g \rangle = \int f(x)\overline{g(x)}\mathrm{d}x$ and norm $\|f\|_{L^2} = \sqrt{\int |f(x)|^2}$.
- Given $g, f \in \mathcal{H}$ some Hilbert space, we have the Cauchy-Schwarz inequality $|\langle f,\, g \rangle| \leqslant \|f\|\|g\|$.
- Let X be a linear space over $\mathbb{R}$ or $\mathbb{C}$. X is a Banach space if it is equipped with a norm $\|\cdot\|$ and is complete with the metric $d(x, y) = \|x - y\|$.
- We say that $f \perp g$ if $\langle f,\, g \rangle = 0$.
- $\{f_j\}_{j \in \mathbb{Z}}$ is an orthonormal sequence if $\langle f_n,\, f_m \rangle = \delta_{n,m} = \begin{cases} 1 & n = m, \\ 0 & n \neq m. \end{cases}$
- Given any orthonormal system $\{f_j\}_{j \in \mathbb{Z}}$ of a Hilbert space $\mathcal{H}$, $\sum_{k \in \mathbb{Z}} |\langle f,\, f_j \rangle|^2 \leqslant \|f\|$ with equality if it is a basis of $\mathcal{H}$.

Properties of the Fourier transform

- For $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the Fourier transform of f is defined by

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-ix\xi} \mathrm{d}x.$$

- The inverse Fourier transform is defined by

$$\check{g}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} g(\xi) e^{ix\xi} \mathrm{d}\xi.$$

- If $\hat{f}, f \in L^1$, then $\check{\hat{f}} = f$.
- The Plancherel equality is $\langle f,\, g \rangle = \frac{1}{2\pi} \langle \hat{f},\, \hat{g} \rangle$.
- The Fourier transform extends to all $f \in L^2(\mathbb{R})$, and $U : f \mapsto \frac{1}{\sqrt{2\pi}} \hat{f}$ is a unitary operator, i.e. $\|Uf\| = \|f\|$.

- If $f, h \in L^1(\mathbb{R})$, then $g = h \star f \in L^1$ and $\hat{g}(\omega) = \hat{f}(\omega)\hat{h}(\omega)$.
- f is bounded and p times differentiable if

$$\int \left|\hat{f}(\omega)\right| (1 + |\omega|^p)\mathrm{d}\omega < \infty$$

  In particular, if there exists $K, \varepsilon > 0$ such that

$$\left|\hat{f}(\omega)\right| < \frac{K}{1 + |\omega|^{p+1+\varepsilon}},$$

  then $f \in C^p$.
- If $\hat{f}$ has compact support, then $f \in C^\infty$.
- If $f = \mathbb{1}_{[-T,T]}$, then $\left|\hat{f}(\omega)\right| \sim |\omega|^{-1}$.
- (Poisson Summation Formula) For all $f \in \mathcal{S}(\mathbb{R})$ (the space of Schwarz functions), $\sum_n f(x + 2\pi n/T)$ is a $2\pi/T$-periodic function and

$$\sum_{n\in\mathbb{Z}} \hat{f}(nT)e^{iTnt} = \frac{2\pi}{T}\sum_{k\in\mathbb{Z}} f\left(t + \frac{2\pi k}{T}\right).$$

  Letting $t = 0$ yields the Poisson Summation Formula

$$\sum_{n\in\mathbb{Z}} \hat{f}(nT) = \frac{2\pi}{T}\sum_{k\in\mathbb{Z}} f\left(\frac{2\pi k}{T}\right).$$

  Phrased differently, in the sense of distributions,

$$\sum_{n\in\mathbb{Z}} e^{-inT\cdot} = \frac{2\pi}{T}\sum_{k\in\mathbb{Z}} \delta\left(\cdot - \frac{2\pi k}{T}\right),$$

  where δ is the Dirac measure.

Lebesgue integration theorems

Theorem A.1 (Fatou's lemma)**.** *Suppose that $\{f_k\}_{k=1}^\infty$ and f are nonnegative and integrable. Then,*

$$\int_{\mathbb{R}^n} \liminf_{k\to\infty} f_k(x)\mathrm{d}x \leqslant \liminf_{k\to\infty} \int_{\mathbb{R}^n} f_k(x)\mathrm{d}x.$$

Theorem A.2 (Monotone Convergence Theorem)**.** *Assume that $\{f_k\}_{k=1}^\infty$ are measurable, with*

$$f_1 \leqslant f_2 \leqslant \cdots \leqslant f_k \leqslant f_{k+1} \leqslant \cdots,$$

then

$$\int_{\mathbb{R}^n} \lim_{k\to\infty} f_k(x)\mathrm{d}x = \lim_{k\to\infty} \int_{\mathbb{R}^n} f_k(x)\mathrm{d}x.$$

Theorem A.3 (Dominated Convergence Theorem)**.** *Assume that the functions $\{f_k\}_{k=1}^\infty$ are integrable and*

$$f_k \to f \qquad a.e.$$

Suppose that

$$|f_k| \leqslant g \qquad a.e.,$$

for some integrable function g. Then,

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} f_k(x)\mathrm{d}x = \int_{\mathbb{R}^n} f(x)\mathrm{d}x.$$