Sparsity in imaging: Introduction to wavelets

Clarice Poon
University of Bath

January, 2019

Outline

Discrete representations

Task

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

Discrete representations

Task

Given a function $f(t)$ defined for $t \in \mathbb{R}$, (say, $f \in L^2(\mathbb{R})$), 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 store using finitely many values.

Discrete representations

Task

Given a function $f(t)$ defined for $t \in \mathbb{R}$, (say, $f \in L^2(\mathbb{R})$), 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 store using finitely many values.

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

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

The Fourier basis

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 $L^2(\mathbb{R})$.

From classical Fourier analysis, we know that

$$\left\{ \frac{1}{\sqrt{2B\pi}} e^{iB^{-1}k\cdot} \; ; \; k \in \mathbb{Z} \right\}$$

is an orthonormal basis of $L^2([-B\pi, B\pi])$. So, given any $f \in L^2([-B\pi, B\pi])$,

$$f(x) = \frac{1}{2B\pi} \sum_{k \in \mathbb{Z}} \hat{f}(kB^{-1}) e^{ikB^{-1}x}. \tag{1.1}$$

The Shannon-Nyquist-Whittake sampling theorem (1950)

Theorem

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

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

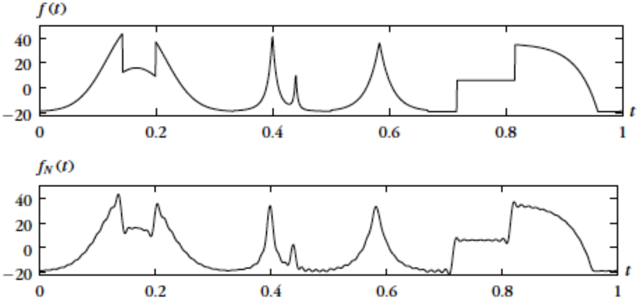
The Shannon-Nyquist-Whittaker theorem provides a discrete representation of functions and describes how one may approximate f with finitely many values. Forms the basis of modern signal processing and communication theory.

Drawbacks

– Fourier approximations or Shannon approximations are better for approximating smooth signals. Natural images have discontinuities...

– 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 $\hat{f}(kB^{-1})$.

Approximation with $N = 128$ Fourier coefficients:

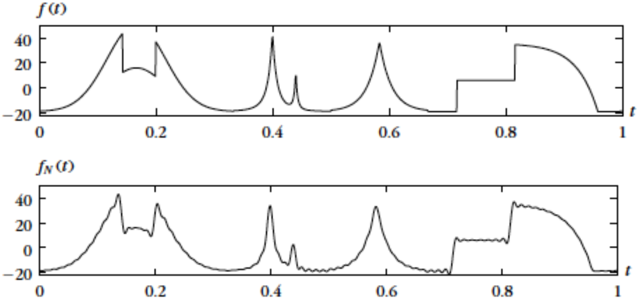


Drawbacks

– Fourier approximations or Shannon approximations are better for approximating smooth signals. Natural images have discontinuities...

– 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 $\hat{f}(kB^{-1})$.

Approximation with $N = 128$ Fourier coefficients:



Wavelets, developed between the 1990's and early 2000's form an alternative basis which are much better for approximating piecewise smooth signals.

This course: some aspects of sparsity in imaging

The Shannon-Nyquist theorem was an important development and forms the basis of much of modern signal processing.

However, in the last few decades, sparsity has played an increasing important role. We will only look at a **selection** of these developments.

- 1980's to early 2000's saw the development of wavelets give rise to sparse representations for natural images and signals.
- mid-2000's to early 2010 saw the development of compressed sensing – how can we exploit this sparsity for efficient image reconstruction and data acquisition?
- around 2012 onwards – development of infinite dimensional versions of compressed sensing. We will look at the Beurling LASSO.

Outline

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases} \qquad \begin{cases} A : 210.375, 124, \\ D : -21.25, 10, \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases} \qquad \begin{cases} A : 210.375, 124, \\ D : -21.25, 10, \end{cases} \qquad \begin{cases} A : 167.188, \\ D : -86.375 \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 10, 8, 4, -99, 0, 0, 1, -1, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : -4, -46.5, 26, -1, \end{cases} \quad \begin{cases} A : 210.375, 124, \\ D : -21.25, 10, \end{cases} \quad \begin{cases} A : 167.188, \\ D : -86.375 \end{cases}$$

Wavelet representations: idea

218,228,215,223,221,225,226,127,106,106, 132,132,129,130,129,128.

$$\begin{cases} \text{Averages} : 223, 219, 223, 176.5, 106, 132, 129.5, 128.5. \\ \text{Differences} : 0, 0, 0, -99, 0, 0, 0, 0, \end{cases}$$

$$\begin{cases} A : 221, 199.75, 119, 129, \\ D : 0, -46.5, 26, 0, \end{cases} \qquad \begin{cases} A : 210.375, 124, \\ D : -21.25, 0, \end{cases} \qquad \begin{cases} A : 167.188, \\ D : -86.375 \end{cases}$$

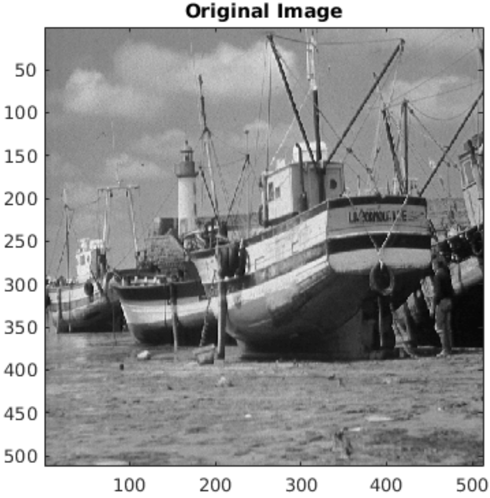
If we set all number $\leqslant 10$ to 0, then we retain only 6 values and the reconstructed sequence

$$221, 221, 221, 221, 223, 223, 226, 127, 111, 111, 137, 137, 124, 124, 124, 124$$

with absolute error:

$$3, 7, 6, 2, 2, 2, 0, 0, 5, 5, 5, 5, 5, 6, 5, 4$$

Wavelet decomposition



Wavelet decomposition

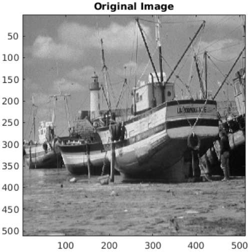
Wavelet decomposition

Wavelet decomposition

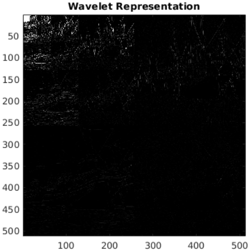
Wavelet decomposition

Wavelet decomposition

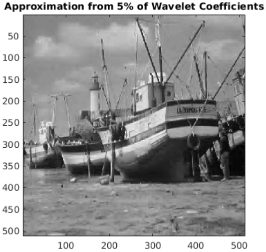
Sparse approximation with wavelets



I



$\{\langle I, \psi_{j,k}\rangle : k, j \in \mathbb{Z}\}$



$\sum_{k,j\in\Lambda}\langle I, \psi_{j,k}\rangle\psi_{j,k}$

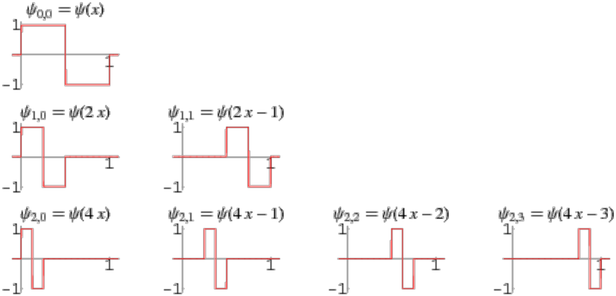
Wavelet definition

Wavelet

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

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

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



A (very) brief history

- 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.

A (very) brief history

- 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.

A (very) brief history

- 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.

A (very) brief history

- 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.

A (very) brief history

- 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.
- Today, JPEG 2000 standard provides compression architectures based on wavelets.

Multiresolution Analysis [Mallat '89, Meyer '92]

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.

Multiresolution Analysis [Mallat '89, Meyer '92]

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.

- (II) means that the spaces are scaled versions of each other.

Multiresolution Analysis [Mallat '89, Meyer '92]

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.

- (II) means that the spaces are scaled versions of each other.
- (I) means that the spaces become increasingly detailed as j increases. As $j \to +\infty$, we recover the entire signal $\lim_{j \to +\infty} \left\| P_{V_j} f - f \right\| = 0$; as $j \to -\infty$, we eventually lose all details as $\lim_{j \to -\infty} \left\| P_{V_j} f \right\| = 0$.

Multiresolution Analysis [Mallat '89, Meyer '92]

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.

- (II) means that the spaces are scaled versions of each other.
- (I) means that the spaces become increasingly detailed as j increases. As $j \to +\infty$, we recover the entire signal $\lim_{j\to+\infty} \left\| P_{V_j} f - f \right\| = 0$; as $j \to -\infty$, we eventually lose all details as $\lim_{j\to-\infty} \left\| P_{V_j} f \right\| = 0$.
- We can replace (V) with the requirement that there exists θ such that $\{\theta(\cdot - k)\ ;\ k \in \mathbb{Z}\}$ is a Riesz basis for V_0, i.e. there exists $A, B > 0$ such that for all $f \in V_0$,

$$A \|f\|^2 \leqslant \sum_k |\langle f, \theta(\cdot - k)\rangle|^2 \leqslant B \|f\|^2 .$$

So, signal expansions over $\{\theta(\cdot - k)\ ;\ k \in \mathbb{Z}\}$ are numerically stable.

Examples of MRA

Piecewise-constant: $\varphi = \chi_{[0,1)}$. and $V_j = \left\{ \chi_{[\frac{k}{2^j}, \frac{(k+1)}{2^j})} \ ; \ k \in \mathbb{Z} \right\}$. Not ideal for approximating smooth functions.

Shannon approximation: V_j is the set of functions with Fourier transform support inside $[-\pi 2^j, \pi 2^j]$. We know from the Shannon-Nyquist theorem that we can choose $\varphi(t) = \frac{\sin(\pi t)}{\pi t}$. Slow decay of $\left\| P_{V_j} f - f \right\|$ if f has compact support.

Spline approximations: V_j of degree m is the space of functions which are $m-1$ continuously differentiable and equal to polynomial of degree m on intervals $[n2^{-j}, (n+1)2^{-j}]$ for $n \in \mathbb{Z}$. $m = 0$ for piecewise-constant MRA, $m = 1$ for piecewise linear. One can construct a Riesz basis for V_0 using box splines,

$$\theta_m = \underbrace{\chi_{[0,1]} \star \cdots \star \chi_{[0,1]}}_{m+1 \text{ times}},$$

centering at 0 if m is odd and at $1/2$ otherwise. The resultant scaling function is $m-1$ differentiable and has exponential decay.

Constructing wavelets from MRA

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

$$V_1 = V_0 \oplus W_0.$$

Constructing wavelets from 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, we get $W_j \overset{\text{def.}}{=} \left\{\psi(2^j \cdot) \; ; \; \psi \in W_0\right\}$ such that

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

Constructing wavelets from 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, we get $W_j \overset{\text{def.}}{=} \left\{\psi(2^j \cdot) \; ; \; \psi \in W_0\right\}$ 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

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

Constructing wavelets from 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, we get $W_j \stackrel{\text{def.}}{=} \left\{\psi(2^j \cdot) \; ; \; \psi \in W_0\right\}$ 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 $j \to +\infty$,

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

Constructing wavelets from 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, we get $W_j \stackrel{\text{def.}}{=} \left\{ \psi(2^j \cdot) \; ; \; \psi \in W_0 \right\}$ 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. This implies that

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

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,

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

The 2π-periodic function m_0 is call the low pass filter of φ.

Theorem

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 $W_0 = \overline{\mathrm{Span}\left\{\psi_{0,k}\ ;\ k\in\mathbb{Z}\right\}}$. Then, $\left\{\psi_{0,k}\ ;\ k\in\mathbb{Z}\right\}$ is an orthonormal basis of W_0 and $V_1 = V_0 \oplus W_0$.

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

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

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,

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

$\beta_k \stackrel{\text{def.}}{=} (-1)^k\overline{\alpha}_k$ are called the high pass filter coefficients.

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,

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

$\beta_k \stackrel{\text{def.}}{=} (-1)^k\overline{\alpha}_k$ are called the high pass filter coefficients.

Example: Piecewise constant MRA is linked to Haar wavelet

If $\varphi = 1_{[-1,0)}$, then $m_0(\xi) = \frac{1}{2}(1+e^{i\xi})$ and $\hat{\varphi}(\xi) = \frac{1-e^{i\xi}}{-i\xi} = e^{i\xi/2}\frac{\sin(\xi/2)}{\xi/2}$. So,

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

which is the Fourier transform of $1_{[-1,-1/2)} - 1_{[-1/2,0)}$.

Almost all wavelets are MRA wavelets...

Remark

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.

Discrete wavelet transform

Let $N = 2^J$. For a continuous signal f supported on $[0, 1]$, we have

$$2^{J/2}\langle f, \varphi_{J,k}\rangle = 2^J \int f(x)\varphi(2^J x - k)\mathrm{d}x \approx f(k/N), \qquad k = 1, \ldots, N.$$

Discrete wavelet transform

Let $N = 2^J$. For a continuous signal f supported on $[0, 1]$, we have

$$2^{J/2}\langle f,\, \varphi_{J,k}\rangle = 2^J \int f(x)\varphi(2^J x - k)\mathrm{d}x \approx f(k/N), \qquad k = 1, \ldots, N.$$

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

$$V_J = V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_{J-1}.$$

Discrete wavelet transform

Let $N = 2^J$. For a continuous signal f supported on $[0,1]$, we have

$$2^{J/2}\langle f,\, \varphi_{J,k}\rangle = 2^J \int f(x)\varphi(2^J x - k)\mathrm{d}x \approx f(k/N), \qquad k = 1, \dots, N.$$

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

$$V_J = V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_{J-1}.$$

Given any $f \in V_J$,

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

Discrete wavelet transform

Let $N = 2^J$. For a continuous signal f supported on $[0,1]$, we have

$$2^{J/2}\langle f, \varphi_{J,k}\rangle = 2^J \int f(x)\varphi(2^J x - k)\mathrm{d}x \approx f(k/N), \qquad k = 1, \ldots, N.$$

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

$$V_J = V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_{J-1}.$$

Given any $f \in V_J$,

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

You can convert between the following 2 representations in $\mathcal{O}(N)$ operations:

- $\{a_{J,k} : \ k = 1, \ldots, N\}$
- $\{a_{0,0}\} \cup \{d_{j,k} : \ 0 \leqslant j < J, \ 0 \leqslant k < 2^j\}$

Discrete wavelet transform [Mallat '89]

We first consider one level of the decomposition $V_j = V_{j-1} \oplus W_{j-1}$.

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

Recall that

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

Discrete wavelet transform [Mallat '89]

We first consider one level of the decomposition $V_j = V_{j-1} \oplus W_{j-1}$.

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

Recall that

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

Decomposition from $\{a_{j,n}\}_n$ to $\{a_{j-1,n}, d_{j-1,n}\}_n$

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

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

Discrete wavelet transform [Mallat '89]

We first consider one level of the decomposition $V_j = V_{j-1} \oplus W_{j-1}$.

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

Recall that

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

Reconstruction from $\{a_{j-1,n}, d_{j-1,n}\}_n$ to $\{a_{j,n}\}_n$

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

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

Wavelets in 2D

Given an ONB $\left\{\psi_{j,k}\ ;\ j,k \in \mathbb{Z}\right\}$ for $L^2(\mathbb{R})$, we can construct an ONB for $L^2(\mathbb{R}^2)$ by taking tensor products:

$$\left\{\Psi_{j_1,j_2,k_1,k_2}(x,y) \stackrel{\text{def.}}{=} \psi_{j_1,k_1}(x)\psi_{j_2,k_2}(y)\ ;\ j_1,j_2,k_1,k_2 \in \mathbb{Z}\right\}.$$

But we lose the MRA structure and we mix information at different scales.

Wavelets in 2D

Given an ONB $\{\psi_{j,k}\ ;\ j,k \in \mathbb{Z}\}$ for $L^2(\mathbb{R})$, we can construct an ONB for $L^2(\mathbb{R}^2)$ by taking tensor products:

$$\left\{\Psi_{j_1,j_2,k_1,k_2}(x,y) \stackrel{\text{def.}}{=} \psi_{j_1,k_1}(x)\psi_{j_2,k_2}(y)\ ;\ j_1,j_2,k_1,k_2 \in \mathbb{Z}\right\}.$$

But we lose the MRA structure and we mix information at different scales.

Take tensor product of 1D MRA's.

Separable wavelet bases

Given an MRA $\{V_j\}_{j\in\mathbb{Z}}$, define for $j \in \mathbb{Z}$,

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

Separable wavelet bases

Given an MRA $\{V_j\}_{j\in\mathbb{Z}}$, define for $j \in \mathbb{Z}$,

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

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

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

Separable wavelet bases

Given an MRA $\{V_j\}_{j\in\mathbb{Z}}$, define for $j \in \mathbb{Z}$,

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

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

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

$\mathbb{W}_j$ consists of 3 parts and it has an orthonormal basis given by

$$\left\{\Psi^h_{j,n}, \Psi^v_{j,n}, \Psi^d_{j,n}\ ;\ n \in \mathbb{Z}^2\right\}$$

where given $\Psi(x,y)$, $\Psi_{j,n}(x,y) \stackrel{\text{def.}}{=} \Psi(2^j x - n_1, 2^j y - n_2)$.

Separable wavelet bases

Given an MRA $\{V_j\}_{j\in\mathbb{Z}}$, define for $j \in \mathbb{Z}$,

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

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

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

$\mathbb{W}_j$ consists of 3 parts and it has an orthonormal basis given by

$$\left\{\Psi^h_{j,n}, \Psi^v_{j,n}, \Psi^d_{j,n}\ ;\ n\in\mathbb{Z}^2\right\}$$

where given $\Psi(x,y)$, $\Psi_{j,n}(x,y) \overset{\text{def.}}{=} \Psi(2^j x - n_1, 2^j y - n_2)$.

There are 3 generating wavelets:

Horizonal: $\Psi^h(x,y) = \psi(x)\varphi(y)$,

Vertical: $\Psi^v(x,y) = \varphi(x)\psi(y)$,

Diagonal: $\Psi^d(x,y) = \psi(x)\psi(y)$.

Outline

Linear and nonlinear approximation (Fourier)

Reconstruct 1D periodic signal $f \in L^2[0,1]$ with basis elements: $e_m(x)$ for $m \in \mathbb{Z}$.

$$f_M^{lin} = \sum_{|m| \leqslant M/2} \langle f, e_m \rangle e_m, \qquad f_M^{nonlin} = \sum_{m \in \Lambda_M} \langle f, e_m \rangle e_m$$

where Λ_M indexes the largest M coefficients of f in magnitude.

Linear and nonlinear approximation (Fourier)

Reconstruct 1D periodic signal $f \in L^2[0,1]$ with basis elements: $e_m(x)$ for $m \in \mathbb{Z}$.

$$f_M^{lin} = \sum_{|m| \leqslant M/2} \langle f, e_m \rangle e_m, \qquad f_M^{nonlin} = \sum_{m \in \Lambda_M} \langle f, e_m \rangle e_m$$

where Λ_M indexes the largest M coefficients of f in magnitude.

Smooth functions (Fourier)

Let $e_m(x) = e^{2\pi i m x}$ with $m \in \mathbb{Z}$.

If $f \in W^{s,2}[0,1]$ with $\mathrm{Supp}(f) \subset (0,1)$, then $|\langle f, e_m \rangle| = \frac{1}{2\pi |m|^s} \left| \langle f^{(s)}, e_m \rangle \right|$.

So,

$$\left\| f - f_M^{lin} \right\|^2 = \sum_{|m| > M/2} \frac{\left| \langle f^{(s)}, e_m \rangle \right|^2}{4\pi^2 \left| m \right|^{2s}} \lesssim \frac{1}{M^{2s}} \sum_{|m| > M/2} \left| \langle f^{(s)}, e_m \rangle \right|^2 \leqslant \frac{1}{M^{2s}} \left\| f^{(s)} \right\|^2$$

Linear and nonlinear approximation (Fourier)

Reconstruct 1D periodic signal $f \in L^2[0,1]$ with basis elements: $e_m(x)$ for $m \in \mathbb{Z}$.

$$f_M^{lin} = \sum_{|m| \leqslant M/2} \langle f, e_m \rangle e_m, \qquad f_M^{nonlin} = \sum_{m \in \Lambda_M} \langle f, e_m \rangle e_m$$

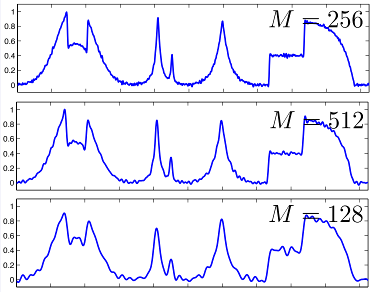
where Λ_M indexes the largest M coefficients of f in magnitude.

Piecewise-regular functions (Fourier)

If f is piecewise regular, then

$$\left\| f - f_M^{lin} \right\|^2 \lesssim M^{-1}.$$

NB: If $f = 1_{[0,1/2)}$, then for $m \neq 0$, $|\langle f, e_m \rangle| = \frac{1}{\pi |m|}$ if m odd and zero otherwise.



Linear and nonlinear approximation (Fourier)

Reconstruct 1D periodic signal $f \in L^2[0,1]$ with basis elements: $e_m(x)$ for $m \in \mathbb{Z}$.

$$f_M^{lin} = \sum_{|m| \leqslant M/2} \langle f, e_m \rangle e_m, \qquad f_M^{nonlin} = \sum_{m \in \Lambda_M} \langle f, e_m \rangle e_m$$

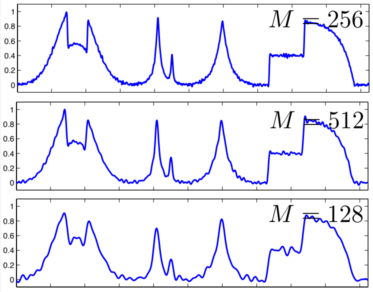
where Λ_M indexes the largest M coefficients of f in magnitude.

Piecewise-regular functions (Fourier)

If f is piecewise regular, then

$$\left\| f - f_M^{lin} \right\|^2 \lesssim M^{-1}.$$

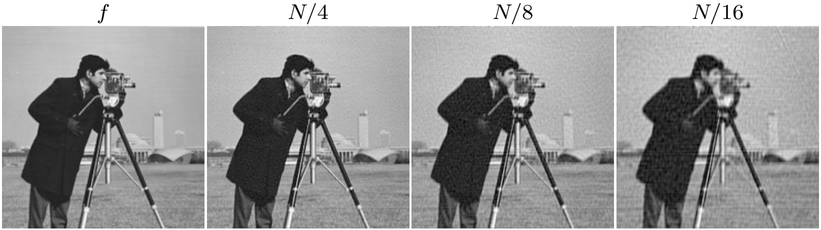
NB: If $f = 1_{[0,1/2)}$, then for $m \neq 0$, $|\langle f, e_m \rangle| = \frac{1}{\pi |m|}$ if m odd and zero otherwise.



For Fourier approximations, linear ≈ nonlinear.

Fourier approximations in 2D

- If $f \in W^{s,2}$, then linear/nonlinear error is $\mathcal{O}(M^{-s})$.
- If f is piecewise regular, then linear/nonlinear error is $\mathcal{O}(M^{-1/2})$.



Desirable properties of wavelets

1. Compact support.
2. We say that ψ has p vanishing moments if

$$\int \psi(x)x^k \mathrm{d}x = 0, \qquad k = 0, \ldots, p-1.$$

Note that if ψ has p vanishing moments, then $\langle f, \psi \rangle = 0$ whenever f is a polynomial of degree at most $p-1$.

Desirable properties of wavelets

1. Compact support.
2. We say that ψ has p vanishing moments if

$$\int \psi(x)x^k \mathrm{d}x = 0, \qquad k = 0, \dots, p-1.$$

Note that if ψ has p vanishing moments, then $\langle f, \psi \rangle = 0$ whenever f is a polynomial of degree at most $p-1$.

Tradeoffs

- An orthonormal wavelet with p vanishing moments must have support size at least $2p-1$.
- There does not exists a smooth wavelet with compact support.

Desirable properties of wavelets

1. Compact support.
2. We say that ψ has p vanishing moments if

$$\int \psi(x)x^k \mathrm{d}x = 0, \qquad k = 0, \dots, p-1.$$

Note that if ψ has p vanishing moments, then $\langle f, \psi \rangle = 0$ whenever f is a polynomial of degree at most $p-1$.

Tradeoffs

- An orthonormal wavelet with p vanishing moments must have support size at least $2p-1$.
- There does not exists a smooth wavelet with compact support.

Remarks:

- 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.

Desirable properties of wavelets

1. Compact support.
2. We say that ψ has p vanishing moments if

$$\int \psi(x) x^k \mathrm{d}x = 0, \qquad k = 0, \ldots, p-1.$$

Note that if ψ has p vanishing moments, then $\langle f, \psi \rangle = 0$ whenever f is a polynomial of degree at most $p-1$.

Tradeoffs

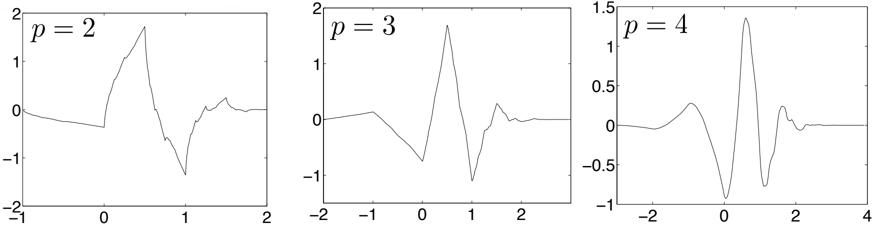
- An orthonormal wavelet with p vanishing moments must have support size at least $2p-1$.
- There does not exists a smooth wavelet with compact support.

Remarks:

- 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.

Daubechies wavelets

An entire family of compactly supported wavelets with arbitrarily many vanishing moments was constructed by Daubechies in 1992!



The Daubechies wavelet of p vanishing moments has support size $2p - 1$.

Magnitude of wavelet coefficients

In dimension d:

- If $f \in L^\infty$, then $\left|\langle f, \psi_{j,k}\rangle\right| \lesssim 2^{-jd/2}$.
- If $f \in C^\alpha$ and ψ has $p > \alpha$ vanishing moments, then $\left|\langle f, \psi_{j,k}\rangle\right| \lesssim 2^{-j(\alpha+d/2)}$.

Wavelets on the interval

We have so far constructed wavelets for $L^2(\mathbb{R})$, but what about $L^2[0,1]$?

If we simply restricted the wavelets to $[0,1]$, then there will be more than 2^j wavelets at the j^{th} scale, and vanishing moments properties will be lost.

Wavelets on the interval

We have so far constructed wavelets for $L^2(\mathbb{R})$, but what about $L^2[0,1]$?

If we simply restricted the wavelets to $[0,1]$, then there will be more than 2^j wavelets at the j^{th} scale, and vanishing moments properties will be lost.

Wavelets on the interval

We have so far constructed wavelets for $L^2(\mathbb{R})$, but what about $L^2[0,1]$?

If we simply restricted the wavelets to $[0,1]$, then there will be more than 2^j wavelets at the j^{th} scale, and vanishing moments properties will be lost.

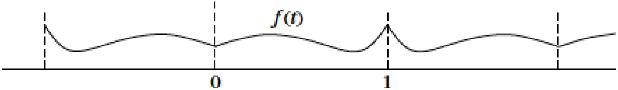
Possibility 1: Periodize. $\psi_{j,k}^{\mathrm{per}} \stackrel{\text{def.}}{=} \sum_{n\in\mathbb{Z}} \psi_{j,k}(x+n)$. There are 2^j elements at scale j, so $\left\{\psi_{j,k}^{\mathrm{per}}\ ;\ j\in\mathbb{Z}, n=0,\ldots,2^j-1\right\}$ is an ONB for $L^2[0,1]$.
Disadvantage: Large coefficients near the boundary. $\langle \psi_{j,k}^{\mathrm{per}},\, f\rangle = \langle \psi_{j,k},\, f^{\mathrm{per}}\rangle$. Vanishing moments are not preserved.
In general, for $f\in C^\alpha([0,1])$, we only have $\left|\langle f,\, \psi_{j,k}^{\mathrm{per}}\rangle\right| \lesssim 2^{-j/2}$.

Wavelets on the interval

We have so far constructed wavelets for $L^2(\mathbb{R})$, but what about $L^2[0,1]$?

If we simply restricted the wavelets to $[0,1]$, then there will be more than 2^j wavelets at the j^{th} scale, and vanishing moments properties will be lost.

Possibility 2: Reflect. $\psi_{j,k}^{\text{fold}} \stackrel{\text{def.}}{=} \sum_{n\in\mathbb{Z}} \psi_{j,k}(-x+2n) + \psi_{j,k}(x+2n)$.



Slightly smaller coefficients near boundary as $\langle \psi_{j,k}^{\text{fold}}, f\rangle = \langle \psi_{j,k}, f^{\text{fold}}\rangle$ and f^{fold} is continuous at 0 and at 1. One vanishing moment preserved.

For $\alpha \in (0,1)$ and $f \in C^\alpha([0,1])$, we have $\left|\langle f, \psi_{j,k}^{\text{fold}}\rangle\right| \lesssim 2^{-j(\alpha+1/2)}$.

Wavelets on the interval

Possibility 3 by [Cohen, Daubechies, Vial (1993)]: modify the wavelets whose support intersect 0 or 1 .

$$\left\{\psi_{j,k}^{\text{int}}\ ;\ j \in \mathbb{Z},\ k = 0, \ldots, 2^j - 1\right\}$$

where

$$\psi_{j,k}^{\text{int}}(x) = \begin{cases} \psi_{j,k} & k = p, \ldots, 2^j - p - 1 \\ \psi_{j,k}^{\text{left}}(x) = 2^{j/2}\psi_k^{\text{right}}(2^j x) & k = 0, \ldots, p-1 \\ \psi_{j,k}^{\text{right}}(x) = 2^{j/2}\psi_k^{\text{right}}(2^j x) & k = 2^j - p, \ldots, 2^j - 1 \end{cases}$$

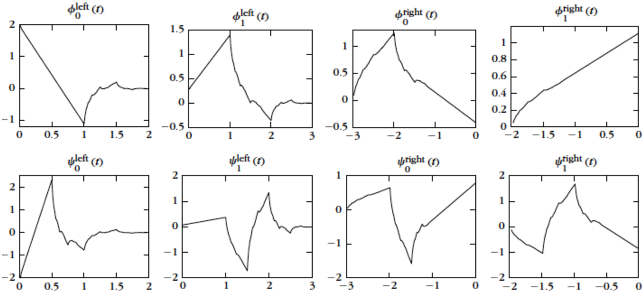


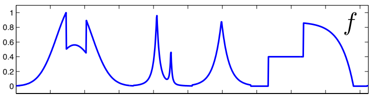
Figure: Modified scaling functions and wavelets for the Daubechies wavelet of $p = 2$ vanishing moments.

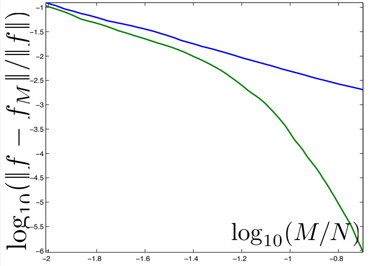
All vanishing moments preserved. For any $\alpha > 0$, if $f \in C^{\alpha}([0,1])$ and ψ has $p > \alpha$ vanishing moments, then $\left|\langle f,\, \psi_{j,k}^{\text{int}}\rangle\right| \lesssim 2^{-j(\alpha+1/2)}$.

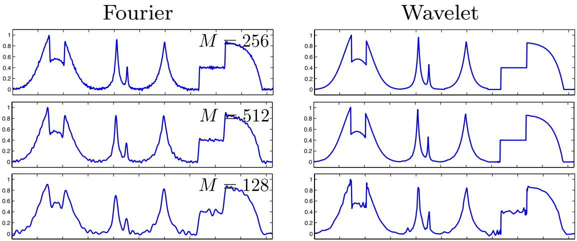
Nonlinear wavelet approximations (1D)

If f is C^α except outside a finite set of discontinuities, then

$$\left\| f - f_M^{nonlin} \right\|^2 = \begin{cases} \mathcal{O}(M^{-1}) & (Fourier) \\ \mathcal{O}(M^{-2\alpha}) & (wavelets) \end{cases}$$







Nonlinear wavelet approximations (1D)

Suppose that f is C^α except at K points.

Nonlinear wavelet approximations (1D)

Suppose that f is C^α except at K points.

Idea: Let $S = \mathrm{Supp}(\psi)$, then for each j, there are at most $K\,|S|$ elements of $\{\psi_{j,k}\}_{0\leqslant k<2^j}$ whose support intersects the K discontinuities.

Nonlinear wavelet approximations (1D)

Suppose that f is C^α except at K points.

Idea: Let $S = \mathrm{Supp}(\psi)$, then for each j, there are at most $K\,|S|$ elements of $\{\psi_{j,k}\}_{0 \leqslant k < 2^j}$ whose support intersects the K discontinuities.

For linear approximation: let $N = 2^J$.

$$\sum_{j \geqslant J} \sum_k \left|\langle f, \psi_{j,k}\rangle\right|^2 \lesssim \underbrace{\sum_{j \geqslant J} k\,|S|\,2^{-j}}_{disct\ intersecting} + \underbrace{\sum_{j \geqslant J} 2^j 2^{-j(1+2\alpha)}}_{nondisct\ intersecting} = \mathcal{O}(2^{-J} + 2^{-2\alpha J}) = \mathcal{O}(N^{-1}).$$

Nonlinear wavelet approximations (1D)

Suppose that f is C^α except at K points.

Idea: Let $S = \mathrm{Supp}(\psi)$, then for each j, there are at most $K\,|S|$ elements of $\{\psi_{j,k}\}_{0 \leqslant k < 2^j}$ whose support intersects the K discontinuities.

For linear approximation: let $N = 2^J$.

$$\sum_{j \geqslant J} \sum_k |\langle f, \psi_{j,k} \rangle|^2 \lesssim \underbrace{\sum_{j \geqslant J} k\,|S|\,2^{-j}}_{disct\ intersecting} + \underbrace{\sum_{j \geqslant J} 2^j 2^{-j(1+2\alpha)}}_{nondisct\ intersecting} = \mathcal{O}(2^{-J} + 2^{-2\alpha J}) = \mathcal{O}(N^{-1}).$$

For nonlinear approximation: let $f_r[k]$ be the k^{th} largest coefficient in magnitude.
$\exists$ at most $jK\,|S|$ wavelets at scales up to j which intersect discontinuities.

$$\left| f_r^{disc}[jK\,|S|] \right| = \mathcal{O}(2^{-j/2}) \implies \left| f_r^{disc}[m] \right| = \mathcal{O}(2^{-m/(2K|S|)}).$$

$\exists$ at most 2^j wavelets not intersecting the discontinuities at scales up to j,

$$\left| f_r^{nondisc}[2^j] \right| = \mathcal{O}(2^{-j(\alpha+1/2)}) \implies \left| f_r^{nondisc}[m] \right| = \mathcal{O}(m^{-(\alpha+1/2)}).$$

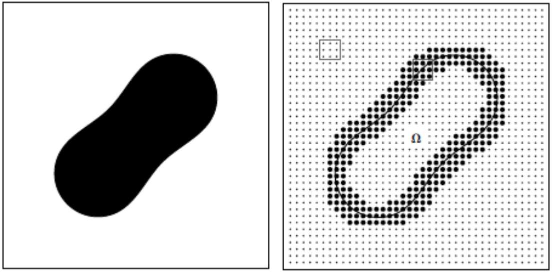
Contribution of the discontinuity intersecting wavelets is negligible!

Wavelet approximations in 2D

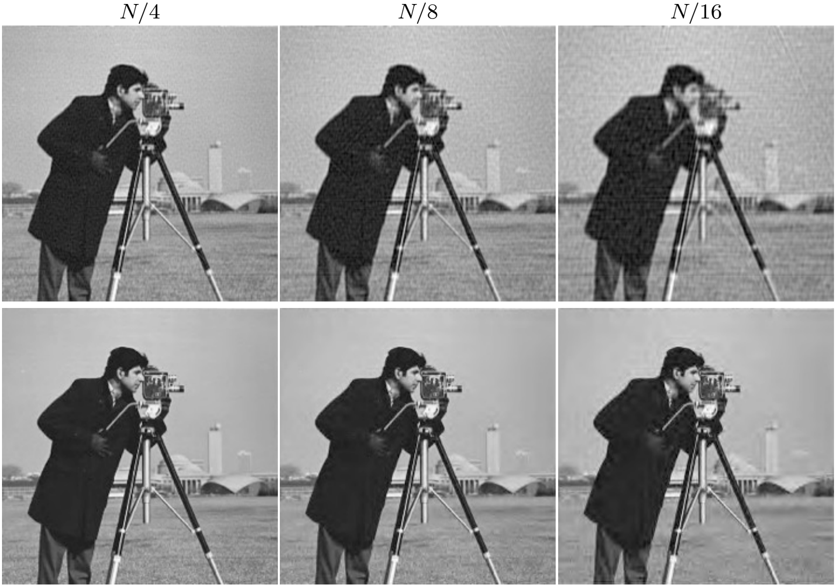
If f is C^α outside a set of finite length edge curves, then

$$\left\| f - f_M^{nonlin} \right\|^2 = \begin{cases} M^{-1/2} & Fourier \\ M^{-1} & Wavelets \end{cases}$$

- if f is discontinuous along some curve, there will be $\mathcal{O}(2^j)$ wavelets at each scale whose support intersect this curve.
- Wavelets better than Fourier, but suboptimal.
- For BV functions, same rate of decay (optimal).



Fourier vs Wavelet approximations

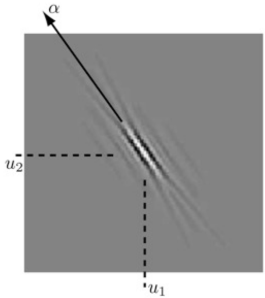


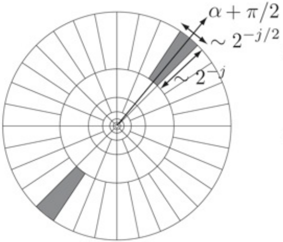
Curvelets

[Candès, Donoho], [Candès, Demanet, Ying, Donoho]

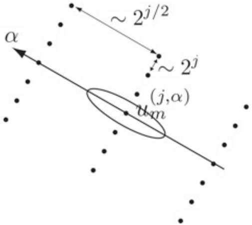
- Parabolic dyadic scaling: $c_{2^j}(x_1, x_2) \approx 2^{-3j/4} c(2^{-j/2} x_1, 2^{-j} x_2)$.
- Rotation: $c^{\alpha}_{2^j,u}(x_1, x_2) = c^{\alpha}_{2^j}((x-u))$ where $c^{\alpha}_{2^j} = c_{2^j}(R_\alpha \cdot)$ and R_α is the rotation matrix with angle α.

"width ≈ length^2"





Frequency tiling



Spatial sampling

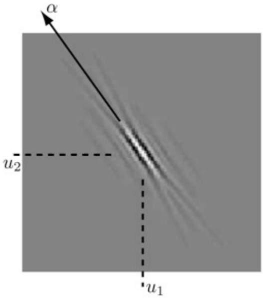
Curvelet Tight Frame

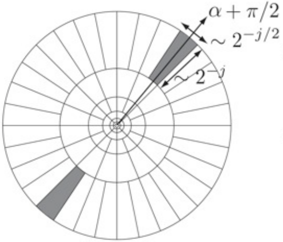
Angular sampling: $\Theta_j = \left\{\alpha = k\pi 2^{\lfloor j/2 \rfloor - 1} \; ; \; 0 \leqslant k < 2^{-\lfloor j/2 \rfloor + 2}\right\}$

Spacial sampling: $\forall m \in \mathbb{Z}^2$, $u_m^{(j,\alpha)} = R_\alpha(2^{j/2} m_1, 2^j m_2)$.

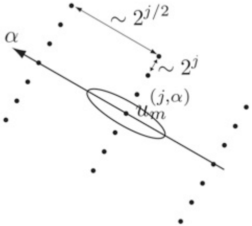
Tight frame of $L^2(\mathbb{R}^2)$: $\{c_{j,m}^{\alpha} = c_{2^j}^{\alpha}(x - u_m^{j,\alpha})\}_{j\in\mathbb{Z}, \alpha\in\Theta_j, m\in\mathbb{Z}^2}$.

$$\|f\| = \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta_j} \sum_{m\in\mathbb{Z}^2} \left|\langle f, c_{j,m}^{\alpha}\rangle\right|^2, \quad f(x) = \sum_{j\in\mathbb{Z}} \sum_{\alpha\in\Theta_j} \sum_{m\in\mathbb{Z}^2} \langle f, c_{j,m}^{\alpha}\rangle c_{j,m}^{\alpha}.$$





Frequency tiling

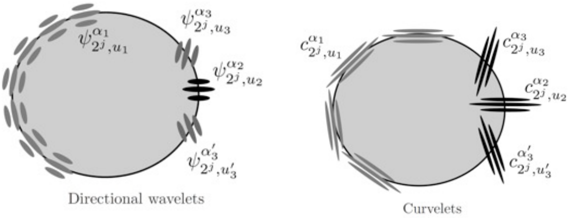


Spatial sampling

Curvelet Approximation

$$f_M = \sum_{(\theta,j,m)\in\Lambda_T} \langle f,\, c^{\alpha}_{j,m}\rangle c^{\alpha}_{j,m} \text{ with } \Lambda_T = \Big\{(j,\theta,m)\ ;\ \Big|\langle f,\, c^{\alpha}_{j,m}\rangle\Big| > T\Big\}.$$

Theorem: If f is C^2 outside a set of C^2 edges, then $\|f - f_M\|^2 = \mathcal{O}(M^{-2}(\log M)^3)$.

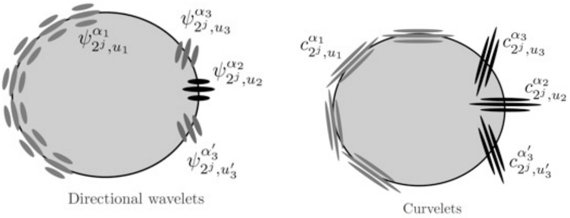


Discrete curvelets: $\mathcal{O}(N\log(N))$ algorithm. `www.curvelet.org`

Curvelet Approximation

$$f_M = \sum_{(\theta,j,m)\in\Lambda_T} \langle f,\, c^{\alpha}_{j,m}\rangle c^{\alpha}_{j,m} \text{ with } \Lambda_T = \left\{(j,\theta,m)\ ;\ \left|\langle f,\, c^{\alpha}_{j,m}\rangle\right| > T\right\}.$$

Theorem: If f is C^2 outside a set of C^2 edges, then $\|f - f_M\|^2 = \mathcal{O}(M^{-2}(\log M)^3)$.

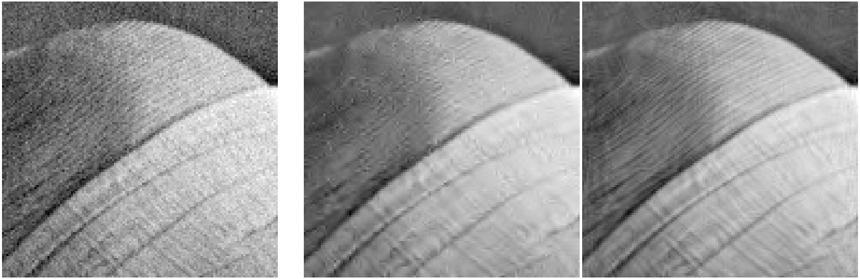


Discrete curvelets: $\mathcal{O}(N\log(N))$ algorithm. `www.curvelet.org`

- Other representation systems based on anisotropic scaling: Contourlets (Vetterli, 2005), Shearlets (Kutynoik, 2006),...
- Curvelets are near-optimal if f is piecewise C^α for $\alpha = 2$, but not any other α! If $\alpha > 2$, the error decay exponent at 2.
- Other adaptive geometric representations such as bandlets [Le Pennec, Mallat, Peyré, '05, '08] (dictionaries of warped wavelet basis) which exhibit optimal error decay of $\mathcal{O}(C^{-\alpha})$ for piecewise C^α functions.

Denoising via thresholding

Noisy | Wavelets | Curvelets



Better at restoring elongated edges, parallel textures. But, irregular textures or pointwise singularities have a representation that is more sparse with wavelets than with curvelets, and are thus better estimated by a wavelet thresholding.

Summary

- We discussed the construction of wavelets via a multiresolution analysis. This forms the basis of the (fast) discrete wavelet transform.
- Linear vs nonlinear approximations.
- Wavelet approximations are optimal for piecewise-smooth 1D signals.
- Sub-optimal for cartoon images, but still better than Fourier representations.
- Improved approximations for piecewise C^2 images via curvelets, shearlets.

Sources

- *A Wavelet Tour of Signal Processing: The Sparse Way* by Stephane Mallat.
- Matlab code: `https://statweb.stanford.edu/~wavelab/`