Wavelet techniques for image processing

Gustavo Garrigós and Eugenio Hernández

Castro Urdiales, July 2-6 2007

Session 1: Discrete signals, Fourier transform and sampling theorem.

Session 2: The JPEG standard: discrete cosine transform, quantization and coding algorithms.

Session 3: Wavelets, multi-resolution and filter design.

Session 4: Wave-packets and best basis optimization.

Session 5: Wavelet image compression and non-linear approximation.

Objectives: Introduce basic mathematical tools underlying in modern image processing, and in particular wavelet based techniques. The course is aimed at a graduate student level.

Main reference: S. Mallat, *A wavelet tour of signal processing*, 2nd Ed., Academic Press 1999.

Session 1: Discrete signals, Fourier transform and sampling theorem

1.- Introduction and notation

DEF: By a *signal* we will mean a function or sequence, typically representing the measurement of a physical phenomenon.

• Functions $f(t)$ or $f(x,y) \longrightarrow$ *analog (or continuous) signals*

• Sequences $f[k]$ or $f[k_1,k_2] \longrightarrow$ *digital (or discrete) signals.*

Examples

• *Audio signals:* $f(t) =$ position of string or membrane of musical instrument at time t sec.

• *Images:* $f(x,y) =$ light intensity measured at point (x,y) of a screen.

• Signals typically belong to the space $L^2(\mathbb{R}^d)$ (or $\ell^2(\mathbb{Z}^d)$) and

$$\|f\|_2^2 = \int_{\mathbb{R}^d} |f(x)|^2\,dx \;=\; \text{Energy } (f).$$

• More general examples of signals (for mathematical manipulations) are contained in the space $\mathcal{S}'(\mathbb{R}^d)$ of tempered distributions.

2.- Basic results about Fourier transform

DEF: The *Fourier transform* (FT) of $f \in L^1(\mathbb{R}^d)$ is defined by

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

and the *inverse Fourier transform* by

$$\mathcal{F}^{-1}g(x) = \tfrac{1}{(2\pi)^d} \int_{\mathbb{R}^d} g(\xi)\, e^{ix\cdot\xi}\, d\xi.$$

DEF: A *convolution operator* is

$$f \longmapsto f * h(x) = \int_{\mathbb{R}^d} f(x-y)h(y)\, dy.$$

The three main properties of the Fourier transform are:

1.- **The inversion formula:** $f = \mathcal{F}^{-1}\hat{f}, \quad \forall\, f \in \mathcal{S}'$ (or $f, \hat{f} \in L^1$).

2.- **Plancherel theorem:** $\|f\|_{L^2(\mathbb{R}^d)} = \frac{1}{(2\pi)^{\frac{d}{2}}} \|\hat{f}\|_{L^2(\mathbb{R}^d)}, \quad \forall\, f \in L^2.$

3.- **The convolution theorem:** $\widehat{(f * h)} = \hat{f}\,\hat{h}, \quad \forall\, f, h \in L^1.$

Examples

1. *Gaussian:* $g(t) = e^{-\frac{t^2}{2}} \Longrightarrow \hat{g}(\xi) = \sqrt{2\pi}\, e^{-\frac{\xi^2}{2}}$.
2. *Rectangle:* $h(\xi) = \chi_{[-a,a]} \Longrightarrow \mathcal{F}^{-1}h(x) = \frac{\sin(ax)}{\pi x}$.
3. *Pure exponential:* $e_{\omega_0}(t) = e^{i\omega_0 t} \Longrightarrow \hat{e}_{\omega_0}(\xi) = \sqrt{2\pi}\, \delta_{\{\omega_0\}}$.
4. *Chirp:* $\gamma(t) = e^{i\frac{t^2}{2}} \Longrightarrow \hat{\gamma}(\xi) = \sqrt{2\pi i}\, e^{-i\frac{\xi^2}{2}}$.

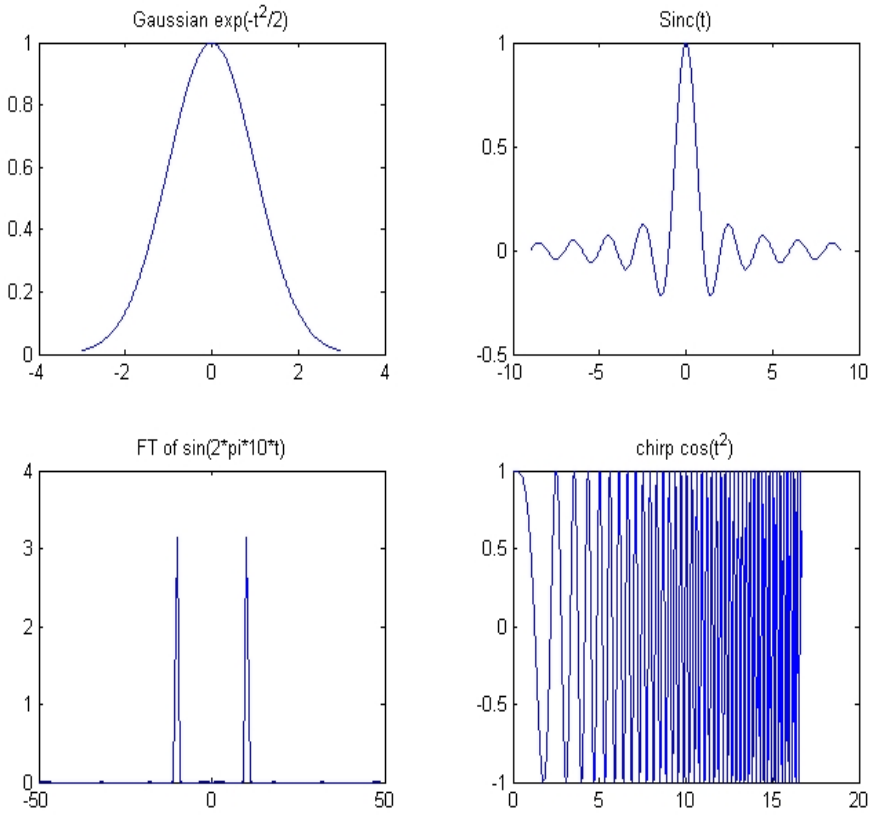


Figure 0.1: Examples of Fourier transforms of elementary functions.

.

3.- Discretization and sampling theorem

Given a signal $f(t)$, we consider the natural discretization

$$f(t) \longmapsto f_d = \{f(\tfrac{k}{N})\}_{k\in\mathbb{Z}},$$

for a suitable integer $N \geq 1$. The new signal f_d can be stored in a computer and manipulated numerically.

Question: Is there a general procedure to reconstruct $f(t)$ uniquely from the values $\{f_d[k]\}_{k\in\mathbb{Z}}$?

Theorem: Shannon-Whittaker.

If $f \in L^2(\mathbb{R})$ and Supp $\hat{f} \subset [-\pi N, \pi N]$, then

$$f(t) = \sum_{k\in\mathbb{Z}} f(\tfrac{k}{N})\, h_N(t - \tfrac{k}{N}), \quad t \in \mathbb{R},$$

where $h_N(t) = \frac{\sin(\pi N t)}{\pi N t}$.

Example

• $f(t) = \sin(2\pi 10t)$,

$$\text{Supp } \hat{f} = [-20\pi, 20\pi] \quad \Longrightarrow \quad \text{need at least} \quad N > 20 \quad \text{samples.}$$

• Sampling $f(t)$ with less points will produce a different signals ("alias signals"):

$$N = 9, \qquad f(\tfrac{k}{9}) = \sin(2\pi 10\,\tfrac{k}{9}) = \sin(2\pi\,\tfrac{k}{9}) \quad \longrightarrow \quad \tilde{f}(t) = \sin(2\pi t)$$
$$N = 10, \qquad f(\tfrac{k}{10}) = \sin(2\pi k) = 0 \quad \longrightarrow \quad \tilde{f}(t) = 0.$$

Observe that the reconstructed signals $\tilde{f}$ have spectrum in $[-\pi N, \pi N]$. This undersampling phenomenon is called **aliasing**.

Notes

• Audio signals are perceived by human ear at frequencies 20 Hz - 20 kHz, thus sampling must be performed with $N \geq 40.000$. In fact, audio CD's perform 44.100 samples/sec (of 16 bits each). Additional oversampling sometimes has important applications, like reducing the "quantization noise" (Super Audio CD's).

• Images may also be sampled in $2D$, $\left\{f(\frac{k_1}{N_1}, \frac{k_2}{N_2})\right\}$, but often these signals involve much higher frequencies, so the aliasing phenomenon is more frequently present (bricked walls, stripped clothes, turning wheels,...).

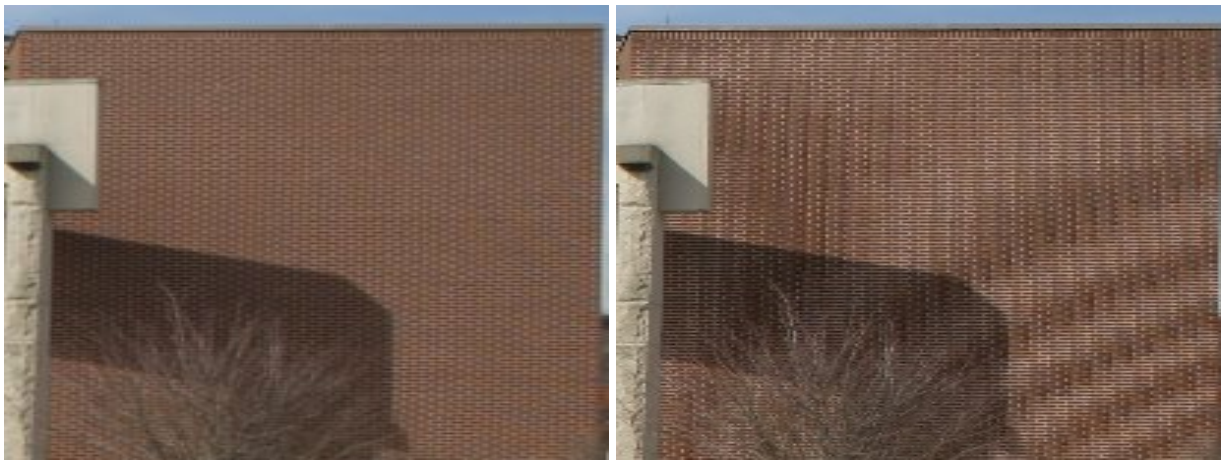


Figure 0.2: Aliasing phenomenon in bricked wall images

.

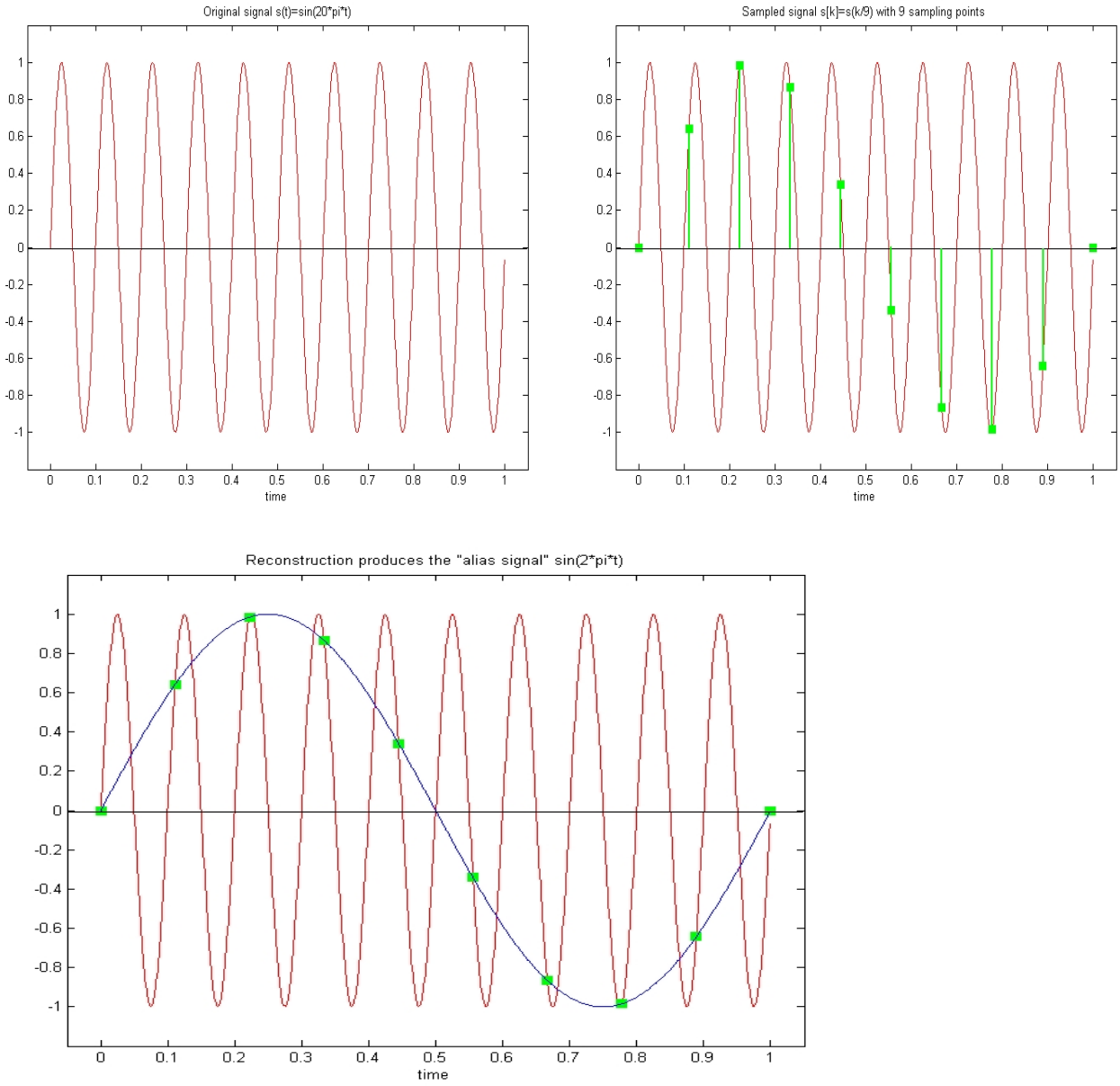


Figure 0.3: Sampling of $f(t) = \sin(2\pi 10t)$ with $N = 9$ points.

.

Proof of Shannon's theorem:

• Write $\hat{f}(\xi) = \hat{f}(\xi)\,\chi_{[-N\pi,N\pi]}(\xi)$ $\qquad (*)$

• Observe that $\mathcal{F}^{-1}\big(\chi_{[-N\pi,N\pi]}\big)(t) = \frac{\sin(\pi N t)}{\pi t} = N h_N(t)$

• Compute the Fourier series of $(*)$ in $[-N\pi, N\pi]$:

$$\hat{f}(\xi) \;=\; \sum_{k\in\mathbb{Z}} a_k\, e^{-\frac{ik\xi}{N}}\, \chi_{[-N\pi,N\pi]}(\xi)$$

which taking inverse Fourier transforms $\mathcal{F}^{-1}$ becomes

$$f(t) \;=\; \sum_{k\in\mathbb{Z}} a_k\, N\, h_N(t - \tfrac{k}{N})$$

• Finally, calculate the Fourier coefficients

$$a_k \;=\; \tfrac{1}{2\pi N} \int_{-\pi N}^{\pi N} \hat{f}(\xi)\, e^{\frac{ik\xi}{N}}\, d\xi \;=\; \tfrac{1}{N}\, \mathcal{F}^{-1}\hat{f}(\tfrac{k}{N}) \;=\; \tfrac{1}{N}\, f(\tfrac{k}{N}).$$

□

4.- The discrete Fourier transform

DEF: The *discrete Fourier transform* of a signal $\{f[n]\}_{0\le n<N}$ is:

$$DFT(f)[k] \;=\; \sum_{n=0}^{N-1} f[n]\, e^{-\frac{2\pi i k n}{N}}, \quad 0 \le k < N,$$

and the *inverse discrete Fourier transform*

$$IDFT(g)[n] \;=\; \tfrac{1}{N} \sum_{k=0}^{N-1} f[k]\, e^{\frac{2\pi i k n}{N}}, \quad 0 \le n < N,$$

NOTE: For simplicity we sometimes write $e_k[n] = e^{\frac{2\pi i k n}{N}}$ and $DFT(f) = \hat{f}$.

Proposition 1 *The system* $\{e_k\}_{0\le k<N}$ *is an orthogonal basis of* $\ell^2(N)$.

Corollary

(1) **Inversion formula:** $\quad f \;=\; IDFT\big[DFT(f)\big]$.

(2) **Plancherel formula:** $\sum_{0\le n<N}\big|f[n]\big|^2 \;=\; \frac{1}{N}\sum_{0\le k<N}\big|\hat{f}[k]\big|^2$.

DEF: The *circular convolution* of $\big\{f[n]\big\}_{0\le n<N}$ and $\big\{g[n]\big\}_{0\le n<N}$ is defined by

$$f \circledast g[n] \;=\; \sum_{\ell=0}^{N-1} \tilde{f}[\ell]\tilde{g}[n-\ell], \quad 0 \le n < N,$$

where $\tilde{f}$ and $\tilde{g}$ are the N-periodic extensions of f and g to $\mathbb{Z}$.

Exercise.

- Convolution formula:

$$DFT\big(f \circledast g\big)[n] \;=\; DFT(f)[n] \cdot DFT(g)[n], \quad 0 \le n < N.$$

The following proposition explains the relation between DFT and the continuous Fourier transform.

Proposition 2

• *Let* $f(t)$ *be an analog signal in* $L^2[0,2\pi]$, *and* $f_d = \left\{f(\frac{2\pi n}{N})\right\}_{0\le n<N}$. *Then*

$$DFT(f_d)[k] \;=\; \tfrac{N}{2\pi}\sum_{\ell\in\mathbb{Z}} \hat{f}(k+N\ell), \quad k\in\mathbb{Z}.$$

• *In particular, if Supp* $\hat{f}\subset[-\frac{N}{2},\frac{N}{2}]$, *then*

$$DFT(f_d)[k] \;=\; \tfrac{N}{2\pi}\,\hat{f}(k), \quad -\tfrac{N}{2}\le k<\tfrac{N}{2}.$$

Proof:

Here $\hat{f}(\ell)=\int_0^{2\pi} f(t)e^{-it\ell}\,dt$ denotes a Fourier coefficient of f as a function of $L^2[0,2\pi]$. Then

$$\begin{aligned}
DFT(f_d)[k] &= \sum_{j=0}^{N-1} f(\tfrac{2\pi j}{N})\, e^{-\frac{2\pi ijk}{N}} \\
\text{(by inversion formula)} &= \sum_{j=0}^{N-1}\Big(\tfrac{1}{2\pi}\sum_{\ell\in\mathbb{Z}} \hat{f}(\ell)\, e^{\frac{2\pi i\ell j}{N}}\Big)\, e^{-\frac{2\pi ikj}{N}} \\
&= \tfrac{1}{2\pi}\sum_{\ell\in\mathbb{Z}} \hat{f}(\ell)\,\underbrace{\Big(\sum_{j=0}^{N-1} e^{-\frac{2\pi i(k-\ell)j}{N}}\Big)}_{N\,\delta_{k+N\mathbb{Z}}}\, d\xi \\
&= \tfrac{N}{2\pi}\sum_{\ell\in\mathbb{Z}} \hat{f}(k+N\ell).
\end{aligned}$$

□

5.- The Fast Fourier transform (FFT)

Recall that

$$\hat{f}[k] \;=\; \sum_{n=0}^{N-1} f[n]\, e^{-\frac{2\pi i k n}{N}}, \quad 0 \le k < N.$$

Thus, computing $\hat{f}$ requires $O(N^2)$ complex multiplications.

Example.

A $3'$ music recording contains $N = 44.100 \times 60 \times 3 \approx 8 \cdot 10^6$ samples.

A modern computer performs 10^9 multiplications per second; therefore

$$DFT \text{ requires} > 17 \text{ hours!!}$$

Theorem: Cooley-Tuckey. *There exists an algorithm, called FFT, which computes DFT with $\le \kappa N \log_2 N$ operations (with $\kappa = \frac{3}{2}$).*

Back to the example, FFT requires ≤ 0.18 seconds!

Corollary: *The fast convolution computes*

$$f \circledast g \;=\; IFFT\big[FFT(f) \cdot FFT(g)\big]$$

with $\le 4\kappa N \log_2 N$ operations.

Proof: Exercise.

Note: The direct definition of $f \circledast g$ requires $2N^2$ operations.

Proof of theorem and description of the algorithm

Assume $N = 2^\ell$.

• Compute first the *even* frequencies:

$$\hat{f}[2k] \;=\; \sum_{n=0}^{N/2-1} f[n]\, e^{-\frac{2\pi i(2k)n}{N}} \;+\; \sum_{m=N/2}^{N-1} f[m]\, e^{-\frac{2\pi i(2k)m}{N}}$$

$$(m = n + \tfrac{N}{2}) \;=\; \sum_{n=0}^{N/2-1} \big(f[n] + f[n+\tfrac{N}{2}]\big)\, e^{-\frac{2\pi ikn}{N/2}} \;=\; DFT(f_{\text{even}}; N/2)[k],$$

where $f_{\text{even}}[n] = f[n] + f[n+\frac{N}{2}]$ is $\frac{N}{2}$-periodic.

• Compute similarly the *odd* frequencies (exercise):

$$\hat{f}[2k+1] \;=\; \sum_{n=0}^{N/2-1} e^{-\frac{2\pi in}{N}} \big(f[n] - f[n+\tfrac{N}{2}]\big)\, e^{-\frac{2\pi ikn}{N/2}} \;=\; DFT(f_{\text{odd}}; N/2)[k],$$

where $f_{\text{odd}}[n] = e^{-\frac{2\pi in}{N}} \big(f[n] - f[n+\frac{N}{2}]\big)$ is also $\frac{N}{2}$-periodic.

• **Iteration:** Thus, if $C(N) = \#$ operations for $FFT(\cdot, N)$ we have

$$\begin{aligned} C(N) \;&=\; 2C\big(\tfrac{N}{2}\big) \;+\; \tfrac{3}{2}\,N \quad \longleftarrow \text{ operations to construct } f_{\text{even}}, f_{\text{odd}} \\ &=\; \ldots \;=\; 2^\ell\, C(1) \;+\; \tfrac{3}{2}\,\ell\, N \;=\; \kappa\, N\, \log_2 N. \end{aligned}$$

□

Note: Variants of FFT, like the *split-radix algorithm* improve this bound to $N \log_2 N$ *real* multiplications and $3N \log_2 N$ sums.

6.- Sampling theorem and DFT for 2-D signals

The product theory is completely analogous, and can be carried out as an exercise.

• **Discretization:**

$$f(x,y) \quad \longrightarrow \quad f_d \,=\, \Big\{ f(\tfrac{k_1}{N_1}, \tfrac{k_2}{N_2}) \Big\}_{\substack{0\le k_1<N_1 \\ 0\le k_2<N_2}},$$

for suitable positive integers N_1, N_2.

• **Sampling theorem.** *If Supp* $\hat{f} \subset [-\pi N_1, \pi N_1] \times [-\pi N_2, \pi N_2]$, *then*

$$f(x_1,x_2) \,=\, \sum_{k_1,k_2\in\mathbb{Z}} f(\tfrac{k_1}{N_1}, \tfrac{k_2}{N_2})\, h_{N_1}(x_1 - \tfrac{k_1}{N_1})\, h_{N_2}(x_2 - \tfrac{k_2}{N_2}).$$

• **2-D DFT.** Given $f = \big\{ f[n_1,n_2] \big\}_{0\le n_1,n_2<N}$ we define

$$DFT(f)[k_1,k_2] \,=\, \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} f[n_1,n_2]\, e^{-\frac{2\pi i k_1 n_1}{N}}\, e^{-\frac{2\pi i k_2 n_2}{N}}.$$

This transform satisfies analogous inversion, Plancherel and convolution formulas.

• **2-D FFT.** Consists is carrying out a 1-D FFT in each variable separately:

$$FFT(f[\cdot,\cdot]) \,:=\, FFT_{n_2}\Big(FFT_{n_1}\big(f[\cdot,n_2]\big)\Big).$$

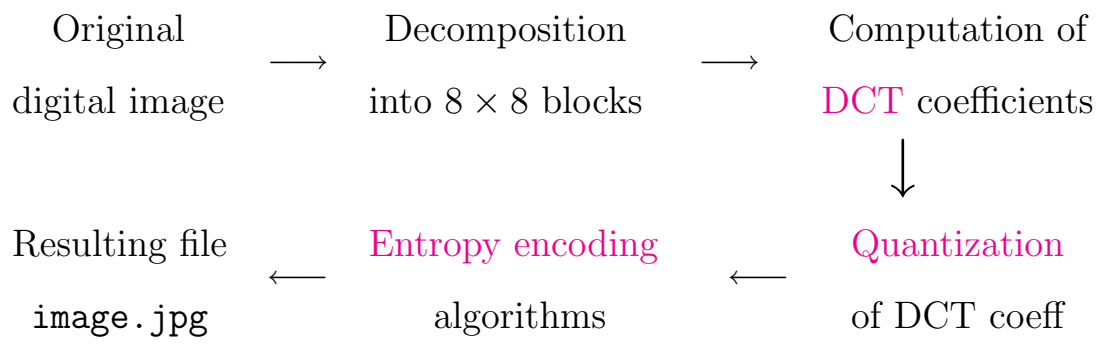
The algorithm performs $\le 2\kappa N^2 \log_2 N$ sums and multiplications.

Example. Image of 1024×1024 pixels. Then $N = 2^{10}$ and FFT requires $2\kappa 2^{20} \cdot 10 \,\approx\, 2^{25}$ operations. If the computer performs 10^9 multiplications/sec, this is less than 0.033 seconds!

Session 2: The JPEG format

1.- Introduction

• JPEG is a well-known format, designed in 1992, for compressing digital images. Good quality images are obtained at compression rates 1 : 8 to 1 : 16. The compression process can be schematically described is as follows:



Example: A typical 1024 × 1024 black & white photograph requires ≈ 1MB of storing space (8 bits/pixel). JPEG produces good quality compression at rates $0'5 - 1$ bits/pixel, i.e. the corresponding files occupy 65-130 KB.

Main JPEG features:

• *Lossy compression:* many high frequency DCT coefficients are small and can be set = 0. Visual resemblance remains, as human vision is less sensitive to high frequency variations than to low frequency changes.

• *Lossless compression:* JPEG profits from special properties of image DCT coefficients, such as long arrays of 0's, suitable size decomposition, similarity between neighboring blocks,...

1.- Cosine bases in $L^2[0,1]$

Motivation: Fourier series approximation of $f \in C[0,1]$ with $f(0) \neq f(1)$ produces oscillations near boundary points (*Gibbs phenomenon*).

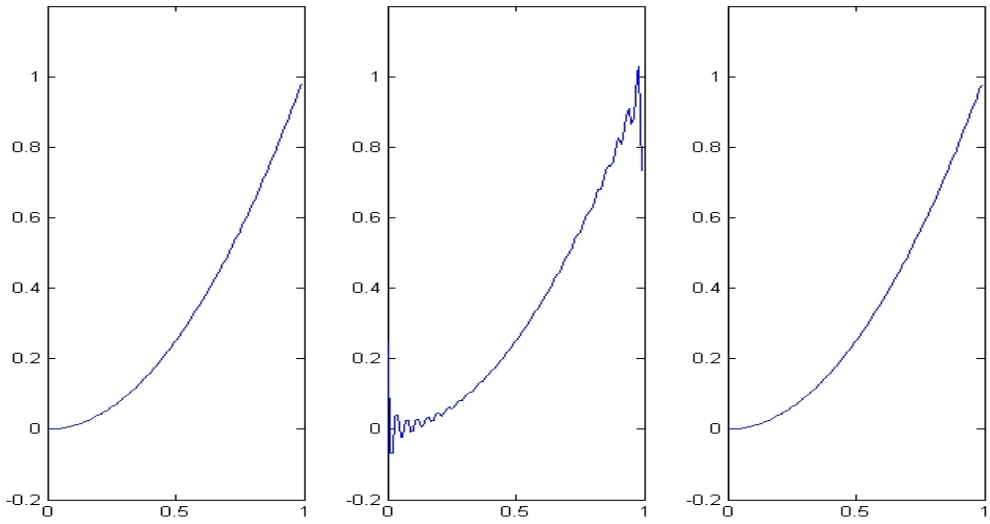


Figure 0.4: $f(t) = t^2$ approximated with Fourier and Cosine-I series.

• The phenomenon may be corrected by considering the even extension of f to $[-1,1]$, which is 2-periodic with $f(-1) = f(1)$. Thus

$$f_{\text{even}}(t) = \sum_{k=0}^{\infty} a_k \cos(\pi k t) + \sum_{k=1}^{\infty} b_k \sin(\pi k t), \quad \text{in } [-1,1].$$

By symmetry of f_{even}, the coefficients $b_k = 0$. Therefore

$$f(t) = \sum_{k=0}^{\infty} a_k \cos(\pi k t), \quad t \in [0,1],$$

which is called Cosine-I series of f.

Note: There is also a *Cosine-IV series*:

$$f(t) = \sum_{k=0}^{\infty} a_k \cos\big(\tfrac{\pi}{2}(2k+1)t\big), \quad t \in [0,1],$$

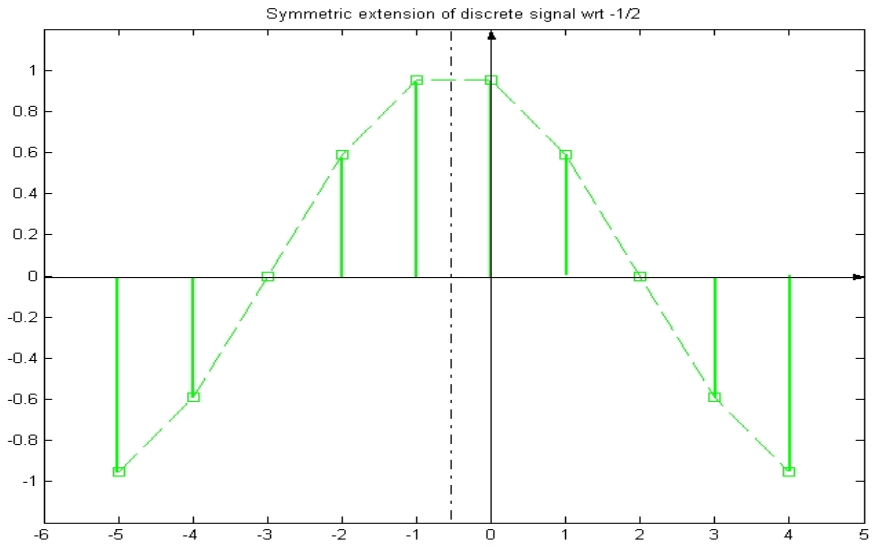
obtained from the extension of f to $[-2,2]$, evenly wrt 0 and oddly wrt ± 1. The Cosine-IV basis appears in various other contexts, such as the construction of ONB in $L^2(\mathbb{R})$ by partitioning the real line with "smooth windows" (local cosine bases).

2.- Discrete cosine bases

The discrete version of the Cosine-I basis is:

$$C_0^{(N)} \equiv 1 \quad \text{and} \quad C_k^{(N)}[n] = \sqrt{2}\,\cos\big(\tfrac{k\pi}{N}\,(n+\tfrac{1}{2})\big), \quad 1 \le k < N,$$

Note: $\cos(k\pi t)$ has been sampled at $t = (n+\frac{1}{2})/N$ (rather than $t = n/N$), $0 \le n < N$, because the even extension of a discrete signal $\{f[n]\}_{0\le n<N}$ is symmetric wrt $-1/2$.



Theorem. *The system* $\big\{\frac{1}{\sqrt{N}}\,C_k^{(N)}\big\}_{k=0}^{N}$ *is an ONB of* $\mathbb{C}^N$.

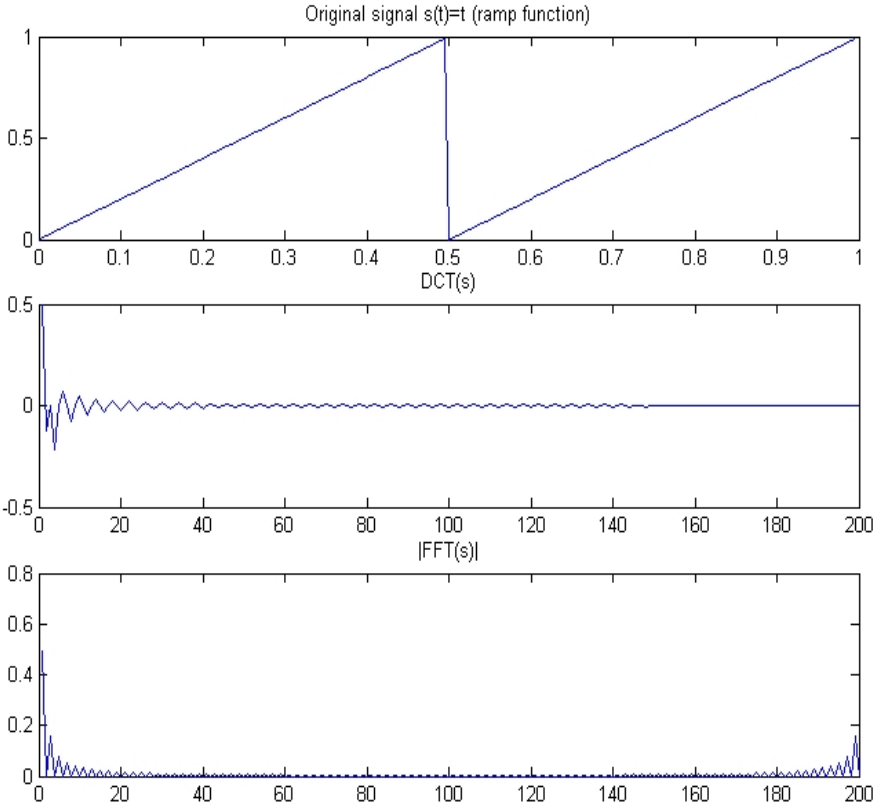
DEF: The $DCT^{(I)}$ and the $IDCT^{(I)}$ of $\{f[n]\}_{0\le n<N}$ are defined by

$$DCT^{(I)}(f)[k] \;=\; \tfrac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f[n]\, C_k^{(N)}[n], \quad 0 \le k < N,$$

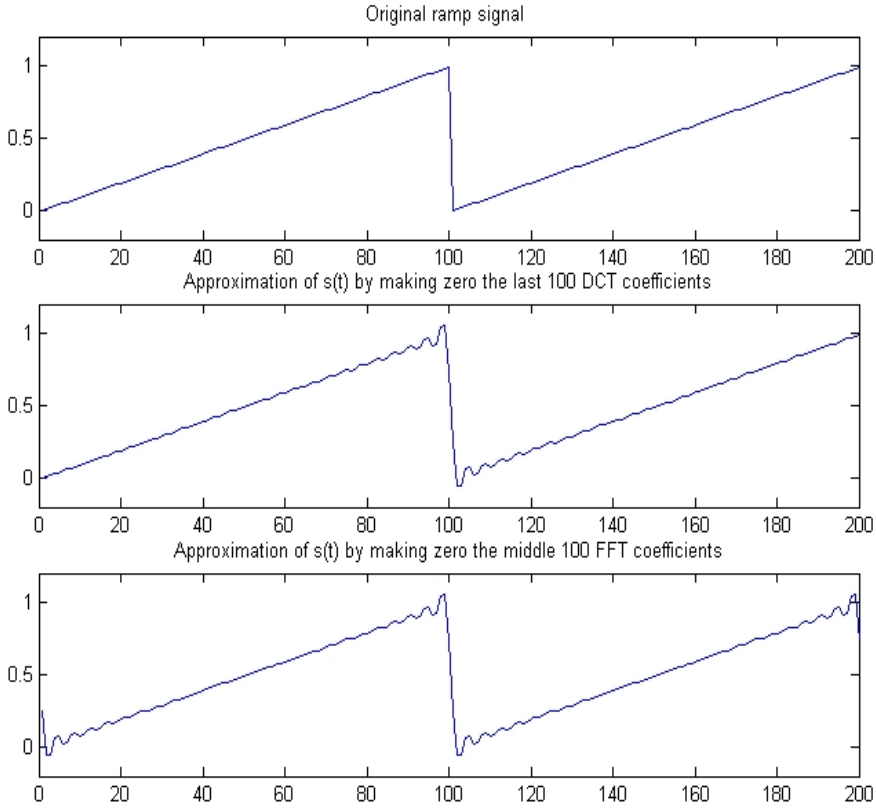
$$IDCT^{(I)}(f)[n] \;=\; \tfrac{1}{\sqrt{N}} \sum_{k=0}^{N-1} f[k]\, C_k^{(N)}[n], \quad 0 \le n < N.$$

As usual, the inversion and Plancherel formulas hold:

$$f = IDCT\big[DCT(f)\big] \quad \text{and} \quad \sum_{n=0}^{N-1} |f[n]|^2 = \sum_{k=0}^{N-1} |DCT(f)[k]|^2.$$


Original signal s(t)=t (ramp function)
DCT(s)
|FFT(s)|

.


Original ramp signal
Approximation of s(t) by making zero the last 100 DCT coefficients
Approximation of s(t) by making zero the middle 100 FFT coefficients
0
0.5
1
0
20
40
60
80
100
120
140
160
180
200

.

3.- Fast algorithms for DCT

Proposition *DCT can be computed with* $O(N \log_2 N)$ *operations.*

Method 1: Write

$$DCT^{(I)}(f)[k] \;=\; \tfrac{1}{\sqrt{2N}}\, e^{-\frac{i\pi k}{2N}}\, FFT_{2N}\big(f_{\text{even}}\big)[k],$$

where as before $f_{\text{even}}[n] = f[-n-1]$ when $-N \le n \le -1$. This method requires $\approx 2\kappa N \log_2(2N)$ operations (some of them complex multiplications).

Method 2: The same type of algorithm as for FFT gives

$$DCT_N^{(I)}(f)[2k] \;=\; \tfrac{1}{\sqrt{2}}\, DCT_{N/2}^{(I)}\big(f^+\big)[k], \quad 0 \le k < N/2$$

$$DCT_N^{(I)}(f)[2k+1] \;=\; \tfrac{1}{\sqrt{2}}\, DCT_{N/2}^{(IV)}\big(f^-\big)[k], \quad 0 \le k < N/2$$

where $f^{\pm} = \{f[n] \pm f[N-n-1]\}_{0 \le n < N/2}$, and

$$\begin{aligned} DCT_M^{(IV)}(g)[k] \;&=\; \tfrac{1}{\sqrt{M}} \sum_{n=0}^{M-1} g[n]\, \sqrt{2}\, \cos\big(\tfrac{\pi}{M}\,(k+\tfrac{1}{2})(n+\tfrac{1}{2})\big) \\ &=\; \text{Discrete Cosine-IV transform.} \end{aligned}$$

This method can be refined to produce $\frac{N}{2}\log_2 N + 1$ real multiplications and $\frac{3}{2}N\log_2(N) - (N+1)$ sums.

4.- The JPEG use of 2-D DCT.

• For a 2-D signal $\{f[n_1, n_2]\}_{0 \le n_1, n_2 < N}$ we have

$$DCT(f)[k_1, k_2] = \frac{1}{N} \sum_{n_1, n_2 = 0}^{N-1} f[n_1, n_2]\, C_{k_1}^{(N)}[n_1]\, C_{k_2}^{(N)}[n_2], \quad 0 \le k_1, k_2 < N.$$

The fast algorithm involves $O(N^2 \log_2 N)$ operations. It is implemented in Matlab with the command `dct2(f)`.

Example: An 8×8 piece of "cameraman" image.

(i) Original signal (in gray scale $[-128, 127]$):

$$f = \begin{bmatrix} 54 & 54 & 57 & 57 & 57 & 60 & 58 & 60 \\ 52 & 55 & 54 & 58 & 61 & 63 & 63 & 64 \\ 55 & 58 & 57 & 66 & 52 & 4 & -43 & -44 \\ 53 & 59 & 63 & 22 & -84 & -112 & -112 & -111 \\ 55 & 62 & 27 & -101 & -115 & -116 & -116 & -115 \\ 57 & 53 & -91 & -114 & -114 & -113 & -113 & -112 \\ 67 & -32 & -112 & -114 & -114 & -113 & -114 & -114 \\ 31 & -104 & -115 & -115 & -116 & -115 & -114 & -114 \end{bmatrix}$$

(ii) Discrete cosine transform of f (rounded to the nearest integer):

$$\texttt{dct2(f)} = \begin{bmatrix} -183 & 302 & 105 & 31 & 14 & 1 & 2 & -4 \\ 456 & -109 & -130 & -72 & -26 & -12 & -1 & -2 \\ 66 & -186 & 5 & 70 & 51 & 25 & 10 & 4 \\ -9 & 1 & 83 & 14 & -40 & -36 & -22 & -5 \\ -53 & 36 & 1 & -49 & -19 & 26 & 27 & 17 \\ -19 & 44 & -19 & -6 & 23 & 16 & -7 & -16 \\ -4 & 6 & -24 & 17 & 9 & -18 & -12 & -8 \\ 9 & 2 & -10 & 16 & -5 & -13 & 17 & 20 \end{bmatrix}$$

(iii) One of the JPEG quantization matrices:

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 108 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 194 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 121 & 100 & 103 & 99 \end{bmatrix}$$

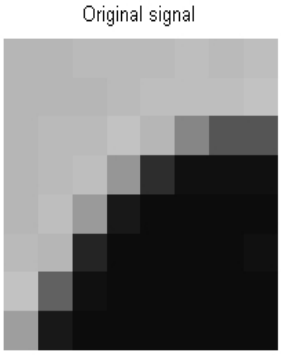
(iv) Resulting DCT matrix (encoding the remaining signal information):

$$Q(\texttt{dct2}(f)) = \begin{bmatrix} -11 & 27 & 10 & 2 & 1 & 0 & 0 & 0 \\ 38 & -9 & -9 & -4 & -1 & 0 & 0 & 0 \\ 5 & -14 & 0 & 3 & 1 & 0 & 0 & 0 \\ -1 & 0 & 4 & 0 & -1 & 0 & 0 & 0 \\ -3 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

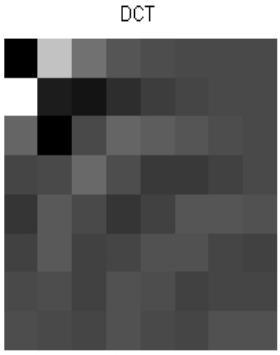
(v) Reconstructed signal (after reversing quantization and using `idct2`):

$$\tilde{f} = \begin{bmatrix} 51 & 60 & 62 & 53 & 47 & 54 & 63 & 68 \\ 55 & 48 & 54 & 72 & 81 & 71 & 57 & 51 \\ 60 & 53 & 57 & 63 & 42 & -1 & -29 & -33 \\ 55 & 64 & 55 & 9 & -59 & -109 & -119 & -109 \\ 56 & 54 & 13 & -66 & -124 & -132 & -119 & -112 \\ 70 & 20 & -59 & -121 & -129 & -105 & -98 & -110 \\ 59 & -30 & -113 & -129 & -109 & -103 & -113 & -119 \\ 22 & -75 & -133 & -108 & -89 & -117 & -129 & -107 \end{bmatrix}$$

Note: A measure of the quality of approximation is given by $\|f - \tilde{f}\|_2^2 = 0.1249\,\|f\|_2^2$.

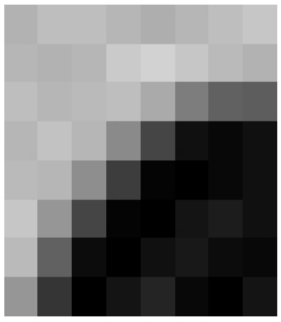


Original signal



DCT

DCT after quantization



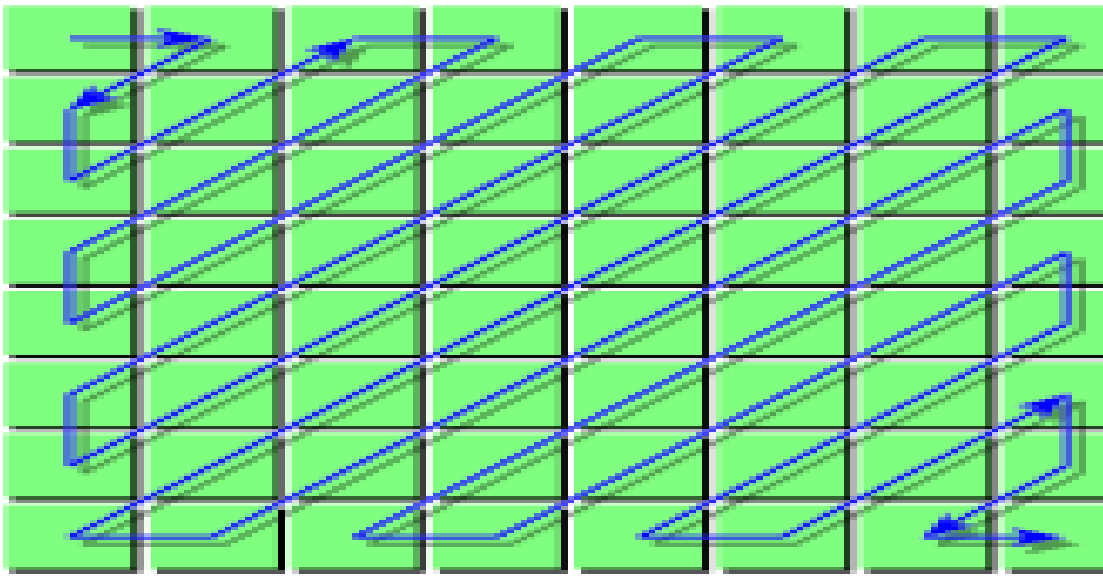
Recosntructed signal

.

.

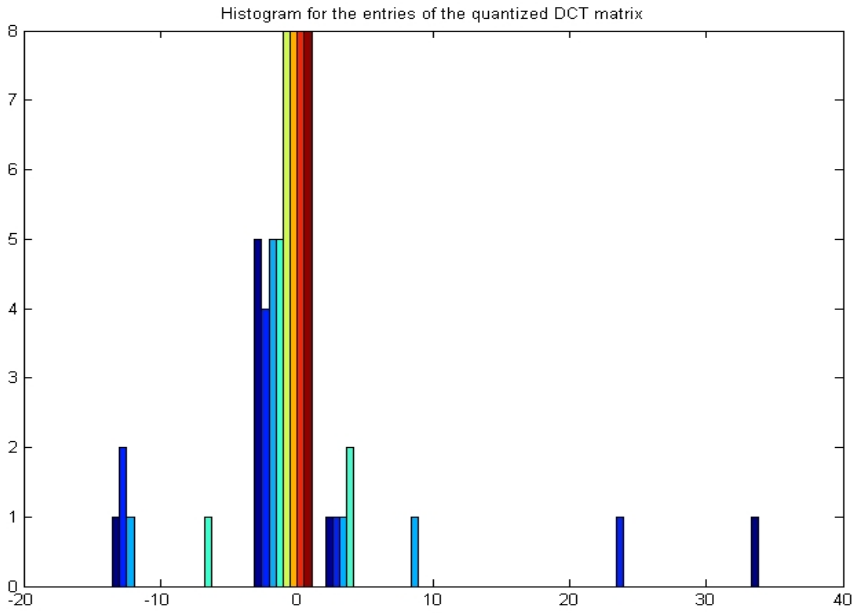
5.- Encoding the information

To encode the information in the quantized DCT matrix, JPEG uses the following sequential *zig-zag* reading of the matrix entries:



One expects long arrays of consecutive 0's at the end of the block, which can be coded with just one symbol EOB = (end of block). Most of the compression is due to this fact.

However, more compression is possible with an efficient reading of the remaining coefficients, which takes into account their probability distribution within the matrix (see histogram below). This suggests studying entropy coding techniques.


Histogram for the entries of the quantized DCT matrix
8
7
6
5
4
3
2
1
0
-20
-10
0
10
20
30
40

.

6.- Lossless compression

Given an finite alphabet $\mathcal{A} = \{x_1, x_2, \ldots, x_M\}$, we want to find a corresponding binary assignment $\{b_1, \ldots, b_M\}$ which minimizes the average #bits/letter:

$$\bar{L} = \sum_{m=1}^{M} \ell_m p_m,$$

where ℓ_m =Length(b_m) and p_m =Prob(x_m).

Example: $\mathcal{A} = \{a, b, c, d\}$. Suppose that a typical sentence is "*abacadabab*", which gives the probability distribution

$$p_a = 0.50, \quad p_b = 0.30, \quad p_c = 0.10, \quad p_d = 0.10.$$

Here are some possible codes:

	a	b	c	d	Rate $\bar{L}$
Code 1	00	01	10	11	2
Code 2	0	10	110	111	1.7
Code 3	0	10	110	101	

Code 2 is better than the standard Code 1 (15% of compression!). Code 3 is ambiguous: 1010 could be "*bb*" or "*da*". Ambiguous codes can be ruled out with the following

Prefix condition: *No code word may be the prefix (beginning) of another word.*

Note: Codes produced from the leafs of a binary tree always satisfy the prefix condition. Conversely, every such code can always be written in binary tree form.

7.- Shannon entropy and Huffman algorithm

Theorem (Shannon). *Given an alphabet* $\{x_1, x_2, \ldots, x_M\}$ *with probability distribution* $\{p_1, \ldots, p_M\}$*, there exists a binary code so that*

$$\mathcal{E} := \sum_{m=1}^{M} p_m \log_2 \tfrac{1}{p_m} \leq \bar{L} < \mathcal{E} + 1.$$

Note: The number $\mathcal{E}$ is called entropy of the code. It somehow measures the uncertainty of the outcomes $\{x_j\}$ according to the probability distribution $\{p_j\}$. An easy exercise shows that $\mathcal{E}$ is maximum when all $p_j = 1/M$.

Example: In the previous example $\mathcal{E} = 1.68$, and the optimal code has $\bar{L} = 1.7$.

• When the probability distribution is known, there is a constructive method to find a code of optimal length: the Huffman algorithm. It is based on the following lemma:

Lemma: *Let* $p_1 \leq p_2 \leq \ldots \leq p_M$*. Suppose we know an optimal* $(M-1)$*-code*

$$\big\{(b_{1,2}, p_{1,2}), (b_3, p_3), \ldots, (b_M, p_M)\big\}$$

where $p_{1,2} = p_1 + p_2$*. Then an optimal* M*-code is given by*

$$\big\{(b_1, p_1), (b_2, p_2), (b_3, p_3), \ldots, (b_M, p_M)\big\},$$

where b_1 *and* b_2 *are the sons of* $b_{1,2}$ *in the corresponding binary tree.*

Exercise: Construct an optimal code for

$$p_1 = 0.05, \quad p_2 = 0.10, \quad p_3 = 0.10, \quad p_4 = 0.15, \quad p_5 = 0.20, \quad p_6 = 0.40$$

Proof of Shannon's theorem:

The proof uses the following lemma

Lemma (Kraft): *Every prefix code of lengths* $\{\ell_1, \ldots, \ell_M\}$ *satisfies*

$$\sum_{m=1}^{M} 2^{-\ell_m} \leq 1. \qquad (*)$$

Conversely, if $\{\ell_1, \ldots, \ell_M\}$ *are positive integers satisfying* $(*)$*, then there exists a prefix code* $\{b_j\}_{j=1}^{M}$ *with Length*$(b_j) = \ell_j$*,* $j = 1, \ldots, M$*.*

• To prove Shannon's theorem, one minimizes with Lagrange multipliers

$$\begin{cases} F(\ell_1, \ldots, \ell_M) &= \sum_{m=1}^{M} \ell_m p_m \\ G(\ell_1, \ldots, \ell_M) &= \sum_{m=1}^{M} 2^{-\ell_m} \leq 1. \end{cases}$$

Then, one must set $\nabla F = \lambda \nabla G$, for some $\lambda \in \mathbb{R}$. Differentiating, this is the same as $p_m = -\lambda 2^{-\ell_m} \ln 2$, or equivalently

$$\ell_m \;=\; \log_2 \tfrac{1}{p_m}, \quad m = 1, \ldots, M.$$

Thus every prefix code has $\bar{L} \geq \mathcal{E}$. Moreover, choosing $\ell_m = \lceil \log_2 \frac{1}{p_m} \rceil$, the Kraft lemma gives a code with $\bar{L} < \mathcal{E} + 1$.

□

Note: As the proof shows, an optimal code (i.e., a Huffman code) reaches the "entropy length" $\bar{L} = \mathcal{E}$ if and only if $\log \frac{1}{p_m} \in \mathbb{N}$, $\forall\, m$. However, it is possible to design codes with $\bar{L}$ as close as desired to $\mathcal{E}$, by replacing the alphabet $\mathcal{A}$ with alphabets $\mathcal{A}^K$ consisting of blocks of K letters. Indeed, an easy exercise shows that in such case

$$\mathcal{E} \;\leq\; \bar{L} \;<\; \mathcal{E} + \tfrac{1}{K}.$$

7.- The JPEG encoding

• The information in each 8×8 block is separated into the **DC** coefficient (in the upper-left corner), and the **AC** coefficients (the remaining 63 read sequentially in zig-zag order).

• The DC coefficients of all blocks are stored together, with a Huffman coding applied to

$$DC^1, \quad DC^2 - DC^1, \quad DC^3 - DC^2, \ldots$$

Note: Reading differences produces smaller numbers (because of frequent resemblances between neighboring blocks), which saves additional information.

• From the remaining 63 AC entries of each block, only **non-null coefficients** are codified. For each such coefficient JPEG uses the symbols

$$(Z, L)(A)$$

where

Z = number of consecutive 0's preceding the coefficient

L = length of the binary expression of the coefficient

A = binary expression of the coefficient.

The EOB symbol is denoted by $(0, 0)$. In practice, the (Z, L) symbol will not add redundancy, because of frequent arrays of 0's and small magnitude coefficients. This symbol is typically stored if a Huffman code (see example below).

• **The symbol A:** after quantization and round-off, one can prove that the matrix entries have absolute value $< 2^8$. Therefore, each entry can be stored with 9 binary digits (one of them counting the sign). One can improve on this by previously specifying the length, according to the table

L	Entries
1	± 1
2	$\pm 2, \pm 3$
3	$\pm 4, \ldots, \pm 7$
4	$\pm 8, \ldots, \pm 15$
5	$\pm 16, \ldots, \pm 31$
$\vdots$	$\vdots$

Observe that entries with length L can now be unambiguously stored in the symbol A with only L digits.

• **About Z and L:** JPEG constrains these numbers to take values in $[0, 15]$. When at some point an array of more than 15 zeros appears, it is denoted by $(15, 0)$ followed by the next symbols, e.g.:

$$(Z_{i-1}, L_{i-1})(A_{i-1}); \quad (15, 0); \quad (Z_{i+1}, L_{i+1})(A_{i+1})$$

Back to example:

(i) Zig-zag array

$$
\begin{array}{cccccccc}
\mathbf{-11} & & & & & & & \\
27 & 38 & & & & & & \\
10 & -9 & 5 & & & & & \\
2 & -9 & -14 & -1 & & & & \\
1 & -4 & 0 & 0 & -3 & & & \\
0 & -1 & 3 & 4 & 2 & -1 & & \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & \\
0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & 0 & & \\
0 & 0 & 0 & 0 & 0 & & & \\
0 & 0 & 0 & 0 & & & & \\
0 & 0 & 0 & & & & & \\
0 & 0 & & & & & & \\
0 & & & & & & &
\end{array}
$$

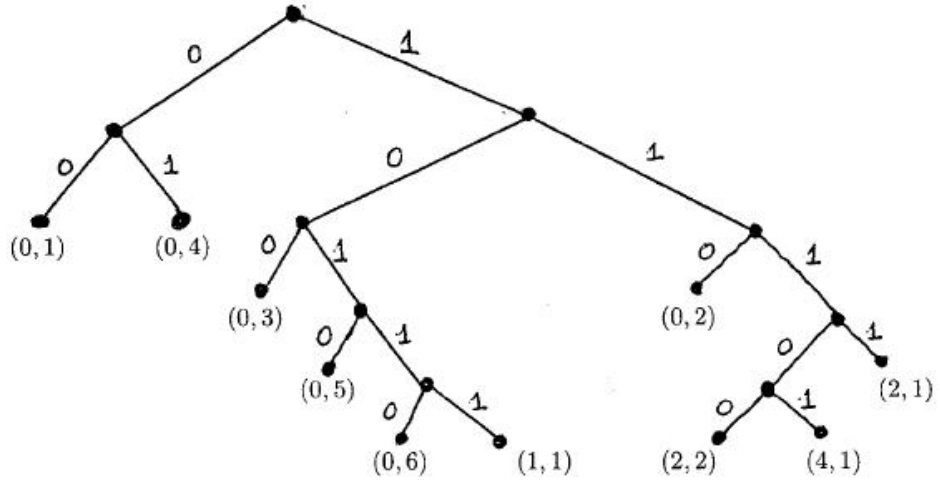
(ii) Thus $DC = -11$, while the AC coefficients are

$$
\begin{array}{cccccccc}
(0,5)(27) & (0,6)(38) & (0,4)(10) & (0,4)(-9) & (0,3)(5) & (0,2)(2) & (0,4)(-9) & \\
(0,4)(-14) & (0,1)(-1) & (0,1)(1) & (0,3)(-4) & (2,2)(-3) & (1,1)(-1) & (0,2)(3) & \\
(0,3)(4) & (0,2)(2) & (0,1)(-1) & (2,1)(1) & (2,1)(1) & (4,1)(-1) & (0,1)(-1) & (\mathbf{0},\mathbf{0})
\end{array}
$$

(iii) Probability distribution of symbols:

Symbols	$(0,1)$	$(0,2)$	$(0,3)$	$(0,4)$	$(0,5)$	$(0,6)$	$(1,1)$	$(2,1)$	$(2,2)$	$(4,1)$
Frequency	□	⊓	⊓	□	|	|	|	⌜	|	|

(iv) Computation of **Huffman tree** leads to the optimal entropy code:



Symbols	$(0,1)$	$(0,2)$	$(0,3)$	$(0,4)$	$(0,5)$	$(0,6)$	$(1,1)$	$(2,1)$	$(2,2)$	$(4,1)$
Binary code	00	110	100	01	1010	10110	10111	1111	11100	11101

Observe that this code has an *average length* of $66/21 = 3.14$ bits/symbol.

(v) Final bit sequence:

1010(27)	10110(38)	01(10)	01(−9)	100(5)	110(2)	01(−9)
01(−14)	00(−1)	00(1)	100(−4)	11100(−3)	10111(−1)	110(3)
100(4)	110(2)	00(−1)	1111(1)	1111(1)	11101(−1)	00(−1)

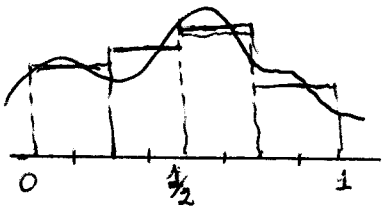
Here it is understood that $27, 38, 10, \ldots$ are represented with respective binary numbers of lengths $5, 6, 4$, etc... This makes a total of 118 bits, which produces a final encoded information with rate ≈ 1.84 bits/pixel.

Final remark: In order to decode this binary sequence, the `jpg` file should store the Huffman code of each 8×8 block. This may be expensive, so

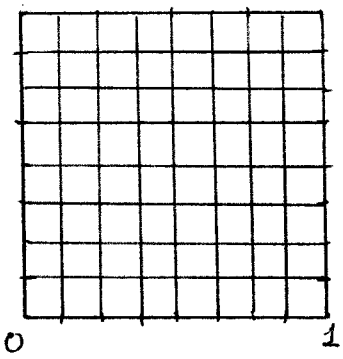
instead one can use a common Huffman code for all the 8×8 blocks, or the standard Huffman tables specified by JPEG.

• JPEG also allows the possibility of using *adaptive* entropy codes, such as the arithmetic code, which learns progressively the probability distribution of the source and adapts the encoding. These algorithms, although more efficient, do not always give substancial improvements and sometimes require patent agreements.

3. WAVELETS FOR SIGNALS AND IMAGES



Approximate representation of signals and images by piecewise constant functions



DEFINITION 3.1.

A finite collection $\{\psi^1, \ldots, \psi^L\} \subset L^2(\mathbb{R}^d)$ is an orthonormal (wavelets) system for $L^2(\mathbb{R}^d)$ if the set

$$\{\psi^\ell_{j,k}(x) = 2^{jd/2}\psi^\ell(2^j x - k) : j \in \mathbb{Z},\ k \in \mathbb{Z}^d,\ \ell = 1, \ldots, L\}$$

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

REMARK 3.2. Only interested in $d=1$ and $d=2$. It turns out that designing a single wavelet ψ in $L^2(\mathbb{R})$ is enough to obtain a system of wavelets in $L^2(\mathbb{R}^2)$ by tensor products.

3.1. MULTIRESOLUTION ANALYSIS IN $L^2(\mathbb{R})$.

DEFINITION 3.3.

A MULTIRESOLUTION ANALYSIS (MRA) consists of a sequence of closed linear subspaces V_j, $j \in \mathbb{Z}$, of $L^2(\mathbb{R})$ satisfying:

i) $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$; ii) $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$, $j \in \mathbb{Z}$

iii) $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$; iv) $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$

v) There exists a function $\varphi \in V_0$ such that $\{\varphi(x-k) : k \in \mathbb{Z}\}$ is an orthonormal basis for V_0. (φ is called scaling function)

REMARK 3.4. i), ii) and v) $\Rightarrow$ iii). See [HW], pages 45 and 46.

If i), ii) and v) holds, then iv) $\Leftrightarrow \hat{\varphi}(0) \neq 0$

A MRA can be defined by assuming only that $\{\varphi(x-k) : k \in \mathbb{Z}\}$ is a Riesz basis of V_0; that is any $f \in V_0$ can be written as $f(x) = \sum_n a_n \varphi(x-n)$ and

$$A \sum_n |a_n|^2 \leq \left\| \sum_{n \in \mathbb{Z}} a_n \varphi(x-n) \right\|_2^2 \leq B \sum_n |a_n|^2 .$$

LEMMA 3.5.

Let φ be a scaling function of an MRA and define $\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k)$, $j,k \in \mathbb{Z}$. Then, $\{\varphi_{j,k} : k \in \mathbb{Z}\}$ is an o.n. basis of V_j, $j \in \mathbb{Z}$.

Let $P_{V_j} : L^2(\mathbb{R}) \to V_j$ be the orthogonal projection from $L^2(\mathbb{R})$ onto V_j. By Lemma 3.5, for any $f \in L^2(\mathbb{R})$,

$$P_{V_j} f = \sum_{k \in \mathbb{Z}} \langle f, \varphi_{j,k} \rangle \varphi_{j,k} \qquad (3.1)$$

with convergence in $L^2(\mathbb{R})$. iii) $\Rightarrow \lim_{j \to -\infty} \|P_{V_j} f\|_2 = 0$; iv) $\Rightarrow$ $\lim_{j \to \infty} \|f - P_{V_j} f\|_2 = 0$. Thus (3.1) gives a better approximation of f as j increases to ∞.

3.2. HOW TO COMPUTE COARSE APPROXIMATIONS. FILTERS.

Let $c_j[k] = \langle f, \varphi_{j,k} \rangle$, $k \in \mathbb{Z}$. You can think that $\{c_j[k]\}_{k \in \mathbb{Z}}$ is a "good" digitalization of a signal f. Can we compute a "not-so-good" digitalization, that is $\{c_{j-1}[k]\}_{k \in \mathbb{Z}}$, of f at a coarser level ?.

FILTERS. Since $\varphi \in V_0$, $\frac{1}{2}\varphi(\frac{t}{2}) \in V_{-1}$ by ii). By (v), there exists $\{h[n] : n \in \mathbb{Z}\} \in \ell^2$ such that

$$\frac{1}{2}\varphi\left(\frac{t}{2}\right) = \sum_{n=-\infty}^{\infty} h[n]\varphi(t-n), \qquad h[n] = \int_{-\infty}^{\infty} \frac{1}{2}\varphi\left(\frac{t}{2}\right)\overline{\varphi(t-n)}\,dt \qquad (3.2)$$

Taking Fourier transforms in the first formula of (3.2) we obtain

$$\hat{\varphi}(2\omega) = \left(\sum_{n=-\infty}^{\infty} h[n] e^{-in\omega}\right)\hat{\varphi}(\omega) \equiv h(\omega)\hat{\varphi}(\omega) \qquad (3.3)$$

where $h \in L^2([0, 2\pi])$ is called LOW PASS FILTER OF THE MRA.

Since $\varphi_{j-1,p} \in V_{j-1} < V_j$ we can write

$$\varphi_{j-1,p} = \sum_{k=-\infty}^{\infty} \langle \varphi_{j-1,p}, \varphi_{j,k} \rangle \varphi_{j,k}$$

where

$$\langle \varphi_{j-1,p}, \varphi_{j,k} \rangle = \ldots\ldots = \sqrt{2}\, h[k-2p] .$$

Thus, $$\varphi_{j-1,p} = \sqrt{2} \sum_{k=-\infty}^{\infty} h[k-2p]\, \varphi_{j,k} \qquad (3.4)$$

Since $c_{j-1}[p] = \langle f, \varphi_{j-1,p} \rangle$ we deduce

$$c_{j-1}[p] = \sqrt{2} \sum_{k=-\infty}^{\infty} \overline{h[k-2p]} \langle f, \varphi_{j,k} \rangle = \sqrt{2} \sum_{k=-\infty}^{\infty} \overline{h[k-2p]}\, c_j[k] \qquad (3.5)$$

* MATRIX REPRESENTATION OF (3.5)

$$\begin{bmatrix} \vdots \\ c_{j-1}[1] \\ c_{j-1}[0] \\ c_{j-1}[1] \\ \vdots \end{bmatrix} = \sqrt{2} \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \\ \cdots \overline{h[0]} & \overline{h[1]} & \overline{h[2]} & \overline{h[3]} & \overline{h[4]} & \cdots \\ \cdots \overline{h[-2]} & \overline{h[-1]} & \overline{h[0]} & \overline{h[1]} & \overline{h[2]} & \cdots \\ \cdots \overline{h[-4]} & \overline{h[-3]} & \overline{h[-2]} & \overline{h[-1]} & \overline{h[0]} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} \vdots \\ c_j[1] \\ c_j[0] \\ c_j[1] \\ \vdots \end{bmatrix}$$

$$[c_{j-1}] = \sqrt{2}\, H\, [c_j]$$

REMARK 3.6. In practice it is convenient to have filters with a finite number of non-zero coefficients. This filters will produce wavelets with compact support.

REMARK 3.7. Formula (3.5) can be written as a convolution. For a sequence $x[n]$ write $x^*[n] = \overline{x[-n]}$. Then, (3.5) is equivalent to

$$c_{j-1}[p] = \sqrt{2}\, c_j * h^*[2p]. \qquad (3.6).$$

---

* COMPLEXITY

If the filter $\{h[n]\}$ has exactly k non-zero coefficients, the number of (complex) operations needed to compute $c_{j-1}[p]$ (p fixed) from (3.5) is $\approx 2k$. If the original data has N non-zero coefficients $\{c_j[k]\}$, we have only $\frac{N}{2}$ non-zero values of $c_{j-1}[p]$. Thus, the total number of operations needed to compute the coarse approximation $c_{j-1}[p]$ is $\approx 2k \cdot \frac{N}{2} = kN$.

3.3. DESIGN OF WAVELETS FROM A MRA (S. MALLAT, 1987)

Let W_0 be the orthogonal complement of V_0 in V_1, $V_0 \oplus W_0 = V_1$. Define $W_j = \{ g(t) = f(2^j t) : f \in W_0 \}$ for all $j \in \mathbb{Z}$. It can be proved that $V_j \oplus W_j = V_{j+1}$ for all $j \in \mathbb{Z}$. Thus

$$V_{j+1} = V_j \oplus W_j = V_{j-1} \oplus W_{j-1} \oplus W_j = \cdots = \bigoplus_{\ell=-\infty}^{j} W_\ell \qquad (3.7)$$

Since $\bigcap_{j \in \mathbb{Z}} V_j = \{0\}$. Since $\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$ we have

$$L^2(\mathbb{R}) = \bigoplus_{\ell=-\infty}^{\infty} W_\ell \qquad (3.8)$$

LEMMA 3.8
~~CONSEQUENCE~~: To obtain a wavelet for $L^2(\mathbb{R})$ it is enough to find $\psi \in W_0$ such that $\{ \psi(\cdot - k) : k \in \mathbb{Z} \}$ is an o.n. basis of W_0.

PROOF: Starting with an o.n. basis of W_0 of the form $\{ \psi(\cdot - k) : k \in \mathbb{Z} \}$, the argument of lemma 3.5 shows that $\{ \psi_{j,k} : k \in \mathbb{Z} \}$ is an o.n. basis of W_j. The result follows from (3.8).

THEOREM 3.9 (S. MALLAT, 1987)

Let $\{ V_j : j \in \mathbb{Z} \}$ be a MRA with scaling function φ and low pass filter $h(\omega)$. If ψ is a function defined by

$$\hat{\psi}(\omega) = g\left(\frac{\omega}{2}\right) \hat{\varphi}\left(\frac{\omega}{2}\right) \quad \text{with} \quad g(\omega) = e^{-i\omega} \overline{h(\omega + \pi)}\, \nu(2\omega),$$

where $\nu(\omega)$ is 2π-periodic with $|\nu(\omega)| = 1$ a.e. $\omega \in \mathbb{R}$, then ψ is an orthonormal wavelet for $L^2(\mathbb{R})$.

We take $\nu(\omega) = 1$. From $g(\omega) = e^{-i\omega} \overline{h(\omega + \pi)}$ the Fourier coefficients $\{ g[n] \}$ of g can be obtained from the Fourier coefficients of $\{ h[n] \}$. In fact,

$$\sum_{n=-\infty}^{\infty} g[n] e^{-in\omega} = g(\omega) = e^{-i\omega} \overline{h(\omega+\pi)} = e^{-i\omega} \sum_{n=-\infty}^{\infty} \overline{h[n]}\, e^{in(\omega+\pi)}$$

$$= \sum_{n=-\infty}^{\infty} \overline{h[n]}\, e^{in\omega} (-1)^n e^{-i\omega n} \overset{n-1=-k}{=\!=} \sum_{k=-\infty}^{\infty} \overline{h[1-k]}\, (-1)^{1-k} e^{-ik\omega}.$$

Thus,

$$g[n] = (-1)^{1-n} \overline{h[1-n]} \tag{3.9}$$

EXERCISE 1. Theorem 3.9 gives a wavelet in terms of its Fourier transform. Use (3.9) to show that

$$\psi(t) = 2 \sum_{n=-\infty}^{\infty} (-1)^{1-n} \overline{h[1-n]}\, \varphi(2t-n] \tag{3.10}$$

giving an expression of ψ in the time domain.

3.4. HOW TO COMPUTE THE DETAILS

The coarse coefficients $c_{j-1}[k] = \langle f, \varphi_{j-1,k} \rangle$, $k \in \mathbb{Z}$, of a signal f are obtained from the coefficients $c_j[k] = \langle f, \varphi_{j,k} \rangle$, $k \in \mathbb{Z}$, using the low-pass filter coefficients $h[k]$ as described in (3.5). The details not captured by this coarse approximation are stored in the coefficients

$$d_{j-1}[k] = \langle f, \psi_{j-1,k} \rangle, \tag{3.11}$$

which are the coefficients of the orthogonal projection $P_{W_{j-1}} : L^2(\mathbb{R}) \longrightarrow W_{j-1}$. These details can be obtained from the data $\{c_{j-1}[k] : k \in \mathbb{Z}\}$ by means of a formula similar to (3.5), replacing the low pass filter coefficients $h[k]$ by the high-pass filter coefficients $g[k]$.

To be precise, since $\psi_{j-1,p} \in W_{j-1} \subset V_j$ we can write

$$\psi_{j-1,p} = \sum_{k=-\infty}^{\infty} \langle \psi_{j-1,p}, \varphi_{j,k} \rangle \varphi_{j,k}$$

Whore

$$\langle \psi_{j-1,p}, \varphi_{j,k} \rangle = \ldots\ldots\ldots = \sqrt{2}\, g[k-2p],$$

Thus since $\frac{1}{2}\psi(\frac{t}{2}) = \sum_{k=-\infty}^{\infty} g[k]\,\varphi(t-k)$ with $g[k] = \langle \frac{1}{2}\psi(\frac{t}{2}), \varphi(t-k)\rangle$.

Thus,

$$\psi_{j-1,p} = \sqrt{2} \sum_{k=-\infty}^{\infty} g[k-2p]\,\varphi_{j,k} \qquad (3.12)$$

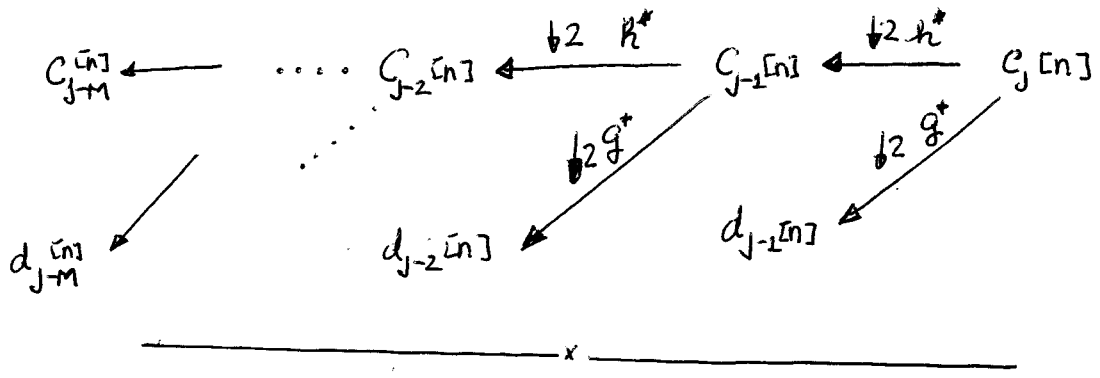
Since $d_{j-1}[p] = \langle f, \psi_{j-1,p}\rangle$ we deduce

$$d_{j-1}[p] = \sqrt{2} \sum_{k=-\infty}^{\infty} \overline{g[k-2p]}\, \langle f, \varphi_{j,k}\rangle = \sqrt{2} \sum_{k=-\infty}^{\infty} \overline{g[k-2p]}\, c_j[k] \qquad (3.13)$$

REMARK 3.10. As in the case of (3.5), formula (3.13) can be written in a matrix form $[d_{j-1}] = \sqrt{2}\, G\,[c_j]$ and also as a convolution

$$d_{j-1}[p] = \sqrt{2}\, c_j * g^*[-p] \qquad (3.14)$$

As in the case of (3.5), the number of operations needed to compute the $\frac{N}{2}$ detail coefficients from the N original non-zero coefficients, using (3.13), is $\le KN$, for a filter of "length" K. Thus, the "decomposition" algorithms (3.5) and (3.13) require a total of $2KN$ (complex) operations.



Starting with $N = 2^q$ coefficients, M steps, with $M \leq q$, can be performed of the decomposition algorithms (3.5) and (3.13) to obtain $N/2^M$ coarse coefficients $\{c_{j-M}[n]\}$ plus

$$\frac{N}{2^M} + \frac{N}{2^{M-1}} + \dots + \frac{N}{2} = N\left(1 - \frac{1}{2^M}\right)$$

details at different leveles. Thus, a total of $N(1 - \frac{1}{2^M}) + N/2^M = N$ numbers are stored at the end of this process.

The ability of wavelets to compress data depends on the number of small details that can be set to zero before starting the reconstruction process.

3.5. THE IMPORTANCE OF WAVELETS WITH ZERO MOMENTS

A function ψ has p zero moments if

$$\int_{\mathbb{R}} \psi(t)\,dt = \int_{\mathbb{R}} t\psi(t)\,dt = \dots = \int_{\mathbb{R}} t^{p-1}\psi(t)\,dt = 0 .$$

Zero moments of a wavelet are important to have small wavelet coefficients. To see this, suppose that ψ has compact support on $[0,M]$ and $f \in C^p$ in a neighborhood of $k2^{-j}$. By the Taylor formula

$$f(t) = \sum_{n=0}^{p-1} \frac{f^{(n)}(k2^{-j})}{n!}(t - 2^{-j}k)^n + \frac{f^{(p)}(\xi_t)}{p!}(t-2^{-j}k)^p$$

valid on $[k2^{-j}, (k+M)2^{-j})$ choosing j large. Thus

$$|d_{j,k}| = |\langle f, \psi_{j,k}\rangle| = \left| \int_{k/2^j}^{\frac{k+M}{2^j}} f(t)\, 2^{j/2}\overline{\psi(2^j t - k)}\,dt \right| =$$

$$= \left| \int_{k/2^j}^{(k+M)/2^j} \frac{f^{p}(\xi_t)}{p!}(t-2^{-j}k)^p\, 2^{j/2}\overline{\psi(2^j t-k)}\,dt \right|$$

$$\leq 2^{j/2}\|\psi\|_\infty \frac{1}{p!}\left[\sup_{\left[\frac{k}{2^j}, \frac{k+M}{2^j}\right)} |f^p(t)|\right] \int_{k/2^j}^{\frac{k+M}{2^j}} (t-2^{-j}k)^p\,dt$$

$$= C(p,\psi,f)\left(\frac{M}{2^j}\right)^{p+1} 2^{j/2} = C\,2^{-j(p+1/2)} \qquad (3.15)$$

Thus, wavelets with a large amount of zero moments produce small details coefficients on smooth parts of a signal. The number of zero moments of a wavelet ψ is reflected on the behaviour of the derivatives of the low-pass filter h at the point π.

PROPOSITION 3.10.

Let ψ be an orthonormal wavelet in $L^2(\mathbb{R})$. If $\hat{\psi}(\omega) \in C^p$ in a neigh. of $\omega = 0$, the following statements are equivalent:

i) ψ has p zero moments

ii) $\dfrac{d^k \hat{\psi}}{d\omega^k}(0) = 0$, $k = 0, 1, \dots, p-1$

iii) For the low-pass filter $h(\omega)$, $\dfrac{d^k h}{d\omega^k}(\pi) = 0$, $k = 0, 1, \dots, p-1$.

PROOF. We have

$$\frac{d^k \hat{\psi}}{d\omega^k} = \frac{d^k}{d\omega^k}\left(\int_{-\infty}^{\infty} \psi(t) e^{-it\omega} dt\right) = \int_{-\infty}^{\infty} \psi(t)(-it)^k e^{-it\omega} dt \quad \Rightarrow$$

$$i^k \frac{d^k \hat{\psi}}{d\omega^k}(0) = \int_{-\infty}^{\infty} t^k \psi(t)\, dt .$$

This proves i) $\Leftrightarrow$ ii). By the S. Mallat recipe (theorem 3.9) with $N = 1$ we have $\hat{\psi}(2\omega) = e^{-i\omega}\, \overline{h(\omega+\pi)}\, \hat{\varphi}(\omega)$. Since $\hat{\varphi}(0) = 1$, $\overline{h(\pi)} = \hat{\psi}(0)$. Taking derivatives (3.16)

$$2\frac{d\hat{\psi}}{d\omega}(2\omega) = -i e^{-i\omega}\overline{h(\omega+\pi)}\,\hat{\varphi}(\omega) + e^{-i\omega}\overline{\frac{dh}{d\omega}(\omega+\pi)}\,\hat{\varphi}(\omega) + e^{-i\omega}\overline{h(\omega+\pi)}\,\frac{d\hat{\varphi}}{d\omega}(\omega) \qquad (3.17)$$

If we assume ii), since we already have $\overline{h(\pi)} = \hat{\psi}(0) = 0$ we deduce

$$0 = 0 + \overline{\frac{dh}{d\omega}(\pi)}\,\hat{\varphi}(0) + 0 .$$

Thus, $\dfrac{dh}{d\omega}(\pi) = 0$ since $\hat{\varphi}(0) = 1$. Proceed by induction up to $p-1$.

Conversely, if we assume iii), (3.16) gives $\hat{\psi}(0) = 0$, and (3.11) produces $\dfrac{d\hat{\psi}}{d\omega}(0) = 0$. Take more derivatives to prove the result. ∎

3.6. THE RECONSTRUCTION ALGORITHM

The decomposition algorithms of sections 3.2 and 3.4 have an inverse algorithm that recovers the signal f precisely from the coarse coefficients and the detail ones. This is due to the decomposition $W_{j-1} \oplus V_{j-1} = V_j$. In fact, this equality shows that $\{\varphi_{j-1,n}\}_{n\in\mathbb{Z}} \cup \{\psi_{j-1,n}\}_{n\in\mathbb{Z}}$ is an o.n. basis of V_j. Hence, for $\varphi_{j,p} \in V_j$,

$$\varphi_{j,p} = \sum_{n=-\infty}^{\infty} \langle \varphi_{j,p}, \varphi_{j-1,n} \rangle \varphi_{j-1,n} + \sum_{n=-\infty}^{\infty} \langle \varphi_{j,p}, \psi_{j-1,n} \rangle \psi_{j-1,n}$$

$$\underset{\text{(3.4) \& (3.12)}}{=\!=\!=} \sqrt{2} \sum_{n=-\infty}^{\infty} \overline{h[p-2n]}\, \varphi_{j-1,n} + \sqrt{2} \sum_{n=-\infty}^{\infty} \overline{g[p-2n]}\, \psi_{j-1,n}.$$

Thus,

$$c_j[p] = \langle f, \varphi_{j,p} \rangle = \sqrt{2} \sum_{n=-\infty}^{\infty} h[p-2n]\, c_{j-1}[n] + \sqrt{2} \sum_{n=-\infty}^{\infty} g[p-2n]\, d_{j-1}[n] \qquad (3.18)$$

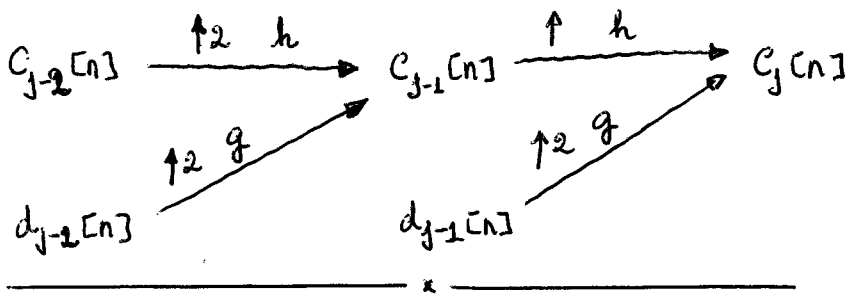
Writing

$$\overset{\vee}{x}[n] = \begin{cases} x[p] & \text{if } n = 2p \\ 0 & \text{if } n = 2p+1 \end{cases}$$

we have

$$c_j[p] = \sqrt{2}\, \overset{\vee}{c}_{j-1} * h[p] + \sqrt{2}\, \overset{\vee}{d}_{j-1} * g[p] \qquad (3.19)$$

COMPLEXITY. For discrete signals of length N the number of operations needed to compute the reconstruction with (3.18) is $\leq 4KN$ with filters of length K.



3.7. THE HAAR WAVELET

A simple MRA in $L^2(\mathbb{R})$ is built by considering V_j as the subspace of $L^2(\mathbb{R})$ that consists of functions that are constant on $[\frac{k}{2^j}, \frac{k+1}{2^j})$, $k \in \mathbb{Z}$. A scaling function of this MRA is $\varphi = \chi_{[0,1]}$.

Since $\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k) = 2^{j/2}\chi_{[\frac{k}{2^j}, \frac{k+1}{2^j})}(t)$ we have

$$P_{V_j} f = \sum_{k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle \varphi_{j,k} = \sum_{k\in\mathbb{Z}} 2^{j/2}\left(\int_{k/2^j}^{\frac{k+1}{2^j}} f(t)\,dt\right) 2^{j/2}\chi_{[\frac{k}{2^j}, \frac{k+1}{2^j})} \qquad (3.20)$$

so that the coefficients $c_{j,k} = \langle f, \varphi_{j,k}\rangle$ are essentially the average of f over $[\frac{k}{2^j}, \frac{k+1}{2^j})$.

Formula (3.2), that is $h[n] = \int_{-\infty}^{\infty} \frac{1}{2}\varphi(\frac{t}{2})\overline{\varphi(t-n)}\,dt$ allows us to obtain

$$h[0] = \frac{1}{2}, \quad h[1] = \frac{1}{2} \quad \text{and} \quad h[n] = 0 \text{ if } n \neq 0, 1. \qquad (3.21)$$

Formula (3.5) that gives coarse approximations is

$$c_{j-1}[p] = \sqrt{2}\left[\frac{1}{2}c_j[2p] + \frac{1}{2}c_j[2p+1]\right], \qquad (3.22)$$

which is "essentially" the average of two consecutive sampling values.

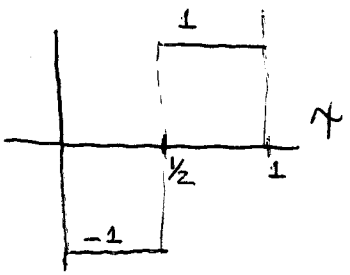
The high-pass filter values are obtained from (3.9):

$$g[0] = -\frac{1}{2}, \quad g[1] = \frac{1}{2} \quad \text{and} \quad g[n] = 0 \text{ if } n \neq 0, 1. \qquad (3.23)$$

REMARK: By (3.10) the wavelet (Haar) is

$$\psi(t) = 2\left[-\frac{1}{2}\varphi(2t) + \frac{1}{2}\varphi(2t-1)\right]$$

Notice that ψ has only one moment zero.



The details are given by (3.13), which applied to this case gives

$$d_{j-1}[p] = \sqrt{2}\left[-\frac{1}{2}c_j[2p] + \frac{1}{2}c_j[2p+1]\right] \qquad (3.24)$$

The reconstruction algorithm can be computed from (3.18) or from (3.22) and (3.24):

$$c_j[2p] = \sqrt{2}\left[\frac{1}{2}c_{j-1}[p] - \frac{1}{2}d_{j-1}[p]\right]; \quad c_j[2p+1] = \sqrt{2}\left[\frac{1}{2}c_{j-1}[p] + \frac{1}{2}d_{j-1}[p]\right]. \qquad (3.25)$$

3.8. FILTER BANKS FOR PERFECT RECONSTRUCTION

Start with two sets of filters $\{h[n]\}_{n\in\mathbb{Z}}\}$ and $\{g[n]: n\in\mathbb{Z}\}$ (real values). For a discrete signal $\{c_j[n]: n\in\mathbb{Z}\}$ follow formulas (3.5) and (3.13) to write (here $h^*[n] = h[-n]$)

$$c_{j-1}[p] = \sqrt{2}\, c_j * h^*[2p] \quad \text{and} \quad d_{j-1}[p] = \sqrt{2}\, d_j * g^*[2p] \qquad (3.26)$$

Take now two other sets of filters $\{\tilde{h}[n]: n\in\mathbb{Z}\}$ and $\{\tilde{g}[n]: n\in\mathbb{Z}\}$ (also real values) and follow (3.19) to write

$$\tilde{c}_j[p] = \sqrt{2}\, \check{c}_{j-1} * \tilde{h}[p] + \sqrt{2}\, \check{d}_{j-1} * \tilde{g}[p] \qquad (3.27)$$

where $\check{x}[n] = \begin{cases} x[p] & \text{if } n = 2p \\ 0 & \text{if } n = 2p+1 \end{cases}$.

QUESTION: Find conditions on h, g, $\tilde{h}$ and $\tilde{g}$ so that we obtain perfect reconstruction, that is $\tilde{c}_j[p] = c_j[p]$.

REMARK 3.11. For a discrete signal $\{x[n]: n\in\mathbb{Z}\}$ we write $X(\omega) = \sum_{n\in\mathbb{Z}} x[n]\, e^{-in\omega}$ and $y[n] = x[2n]$ (subsample). Then

$$y(2\omega) = \frac{1}{2}\left[X(\omega) + X(\omega+\pi)\right].$$

REMARK 3.12. Upsampling by 2 has also a simple expression with Fourier series: if $z[n] = \check{x}[n]$ we have

$$Z(\omega) = \sum_{n\in\mathbb{Z}} z[n]\, e^{-in\omega} = \sum_{p\in\mathbb{Z}} x[p]\, e^{-i2p\omega} = X(2\omega)$$

TEOREMA 3.13. (Vetterli, 1986)

Perfect reconstruction of any discrete signals $\{c_j[n]: n\in\mathbb{Z}\}$ with filter banks h, g, $\tilde{h}$ and $\tilde{g}$ is obtained if and only if

$$\overline{\hat{h}(\omega+\pi)}\,\hat{\tilde{h}}(\omega) + \overline{\hat{g}(\omega+\pi)}\,\hat{\tilde{g}}(\omega) = 0 \qquad (3.28)$$

and

$$\overline{\hat{h}(\omega)}\,\hat{\tilde{h}}(\omega) + \overline{\hat{g}(\omega)}\,\hat{\tilde{g}}(\omega) = 1 \qquad (3.29)$$

Moreover, if the filters are of finite length, there exists $a\in\mathbb{R}$ and $l\in\mathbb{Z}$ such that

$$\hat{g}(\omega) = a\, e^{-i(2l+1)\omega}\,\overline{\hat{\tilde{h}}(\omega+\pi)} \quad \text{and} \quad \hat{\tilde{g}}(\omega) = \frac{1}{a}\, e^{-i(2l+1)\omega}\,\overline{\hat{h}(\omega+\pi)} \qquad (3.30)$$

PROOF. We give the proof of the first part. We have

$$(c_j * h^*)(\omega) = c_j(\omega)\cdot h^*(\omega) = c_j(\omega)\,\overline{h(\omega)}$$

Since $h^*(\omega) = \sum_{n\in\mathbb{Z}} h^*[n]\, e^{-in\omega} = \sum_{n\in\mathbb{Z}} h[-n] e^{-in\omega} = \overline{h(\omega)}$ ($h[n]$ real).

Thus, from (3.26) and Remark 3.11 we obtain

$$c_{j-1}(2\omega) = \frac{\sqrt{2}}{2}\left[c_j(\omega)\overline{h(\omega)} + c_j(\omega+\pi)\overline{h(\omega+\pi)}\right]. \qquad (3.31)$$

Similarly, one obtains

$$d_{j-1}(2\omega) = \frac{\sqrt{2}}{2}\left[c_j(\omega)\overline{g(\omega)} + c_j(\omega+\pi)\overline{g(\omega+\pi)}\right]. \qquad (3.32)$$

Using (3.27) and Remark 3.12 we obtain

$$\tilde{c}_j(\omega) = \sqrt{2}\, c_{j-1}(2\omega)\,\tilde{h}(\omega) + \sqrt{2}\, d_{j-1}(2\omega)\,\tilde{g}(\omega) \qquad (3.33)$$

Replace $c_{j-1}(2\omega)$ and $d_{j-1}(2\omega)$ by the values obtained in (3.31) and (3.32) to conclude the first part of the result. ∎

3.9. DESIGN OF FILTERS

We want to find finite filters $\{h[n] : n\in\mathbb{Z}\}$ and $\{\tilde{h}[n] : n\in\mathbb{Z}\}$ such that

$$\overline{h(\omega)}\,\tilde{h}(\omega) + \overline{h(\omega+\pi)}\,\tilde{h}(\omega+\pi) = 1 \qquad (3.34)$$

and

$$g(\omega) = e^{-i\omega}\overline{\tilde{h}(\omega+\pi)}, \quad \tilde{g}(\omega) = e^{-i\omega}\overline{h(\omega+\pi)}. \qquad (3.35)$$

Condition (3.35) implies (3.28). Thus, these two conditions are sufficient to have perfect reconstruction filters according to theorem (3.13). We have choosen $\ell=1$, $a=0$ in (3.30).

In order to have $\int_{\mathbb{R}} \psi(t)dt = 0 = \int_{\mathbb{R}} \tilde{\psi}(t)dt = 0$, we must have $g(0) = \tilde{g}(0) = 0$, and consecuently $h(\pi) = \tilde{h}(\pi) = 0$. From (3.34) we obtain $\overline{h(0)}\,\tilde{h}(0) = 1$; this is satisfied if we choose $h(0) = 1 = \tilde{h}(0)$.

In order to the wavelet ψ associated to the filter g to have p zero moments, we must suppose

$\frac{d^k \tilde{h}}{d\omega^k}(\pi)=0$ if $k=0,1,2,-,p-1$. Similarly, for the wavelet $\tilde{\psi}$ associated to the filter $\tilde{g}$ to have $\tilde{p}$ zero moments, we must assume $\frac{d^k h}{d\omega^k}(\pi)=0$ if $p=0,1,2,-,p-1$.

We restrict ourselves to symmetric filters $h[n]=h[-n]$ of finite length. Write $h(\omega)=\sum_{n=-N}^{N} h[n]e^{-in\omega}$. Thus

$$h(-\omega)=\sum_{n=-N}^{N} h[n]e^{in\omega}=\sum_{n=-N}^{N} h[-n]e^{-in\omega}=\sum_{n=-N}^{N} h[n]e^{-in\omega}=h(\omega)$$

so that the transfer function $h(\omega)$ is also symmetric. Observe that the number of non-zero coefficients of h is odd.

For this case

$$h(\omega)=\sum_{-N}^{-1} h[n]e^{-in\omega}+h[0]+\sum_{1}^{N} h[n]e^{-in\omega}$$

$$=\sum_{1}^{N} h[n]e^{in\omega}+h[0]+\sum_{1}^{N} h[n]e^{-in\omega}$$

$$=h[0]+\sum_{n=1}^{N} h[n]2\cos n\omega=p(\cos\omega), \qquad (3.36)$$

where p denotes a polynomial. The next lemma shows that a solution $\tilde{h}$ of (3.34) can also be taken to be symmetric with respect to $n=0$.

LEMMA 3.14.

Let $h(\omega)=h(-\omega)$ and suppose that $\tilde{h}$ satisfies (3.34). Then, $\tilde{h}^{\#}(\omega)=\frac{1}{2}[\tilde{h}(\omega)+\tilde{h}(-\omega)]$ also satisfies (3.34) and $\tilde{h}^{\#}(\omega)=\tilde{h}^{\#}(-\omega)$.

PROOF. Replace $\tilde{h}$ by $\tilde{h}^{\#}$ in (3.34) and compute the result. ∎

From (3.36) and $h(\pi)=0$ we deduce $p(-1)=h(\pi)=0$. Thus, $x=-1$ is a root of $p(x)$ and we can write $p(x)=(1+x)^{\tilde{p}}q(x)$ for some $\tilde{p}=1,2,\ldots$. Since $1+\cos\omega=2\cos^2\frac{\omega}{2}$ we have

$$h(\omega)=p(\cos\omega)=(1+\cos\omega)^{\tilde{p}}q(\cos\omega)=2^{\ell}\left(\cos\frac{\omega}{2}\right)^{2\tilde{p}}q(\cos\omega)=\left(\cos^2\frac{\omega}{2}\right)^{\tilde{p}}p_0(\cos\omega)$$

$$p_0(-1)\neq 0 \qquad (3.37)$$

The filter (3.37) gives a wavelet $\tilde{\psi}$ with $2\tilde{p}$ zero moments. Similarly,

$$\tilde{h}(\omega) = (\cos^2 \tfrac{\omega}{2})^{p} \tilde{p}_0(\cos\omega) \quad , \tilde{p}_0(-1) \neq 0 \tag{3.38}$$

gives a wavelet ψ with $2p$ zero moments.

Replace (3.37) and (3.38) in (3.34). We obtain

$$1 = (\cos \tfrac{\omega}{2})^{2\tilde{p}} p_0(\cos\omega)(\cos \tfrac{\omega}{2})^{2p} \tilde{p}_0(\cos\omega) +$$

$$(\cos \tfrac{\omega+\pi}{2})^{2\tilde{p}} p_0(\cos(\omega+\pi))(\cos \tfrac{\omega+\pi}{2})^{2p} \tilde{p}_0(\cos(\omega+\pi))$$

$$= (\cos \tfrac{\omega}{2})^{2(\tilde{p}+p)} p_0(\cos\omega)\tilde{p}_0(\cos\omega) + (\operatorname{sen} \tfrac{\omega}{2})^{2(\tilde{p}+p)} p_0(-\cos\omega)\tilde{p}_0(-\cos\omega).$$

Let $Q(x) = p_0(x)\tilde{p}_0(x) = P(\frac{1-x}{2})$ (use Taylor). Then

$$Q(\cos\omega) = P(\tfrac{1-\cos\omega}{2}) = P(\operatorname{sen}^2 \tfrac{\omega}{2})$$

and

$$Q(-\cos\omega) = P(\tfrac{1+\cos\omega}{2}) = P(\cos^2 \tfrac{\omega}{2}).$$

Thus, we have to find P such that

$$(\cos \tfrac{\omega}{2})^{2(\tilde{p}+p)} P(\operatorname{sen}^2 \tfrac{\omega}{2}) + (\operatorname{sen} \tfrac{\omega}{2})^{2(\tilde{p}+p)} P(\cos^2 \tfrac{\omega}{2}) = 1. \tag{3.39}$$

REMARK 3.15. Once P is computed, different filters can be obtained by writing different factorizations of $P(\operatorname{sen}^2 \frac{\omega}{2}) = p_0(\cos\omega)\tilde{p}_0(\cos\omega)$.

With $x = \operatorname{sen}^2 \frac{\omega}{2}$, $1-x = 1-\operatorname{sen}^2 \frac{\omega}{2} = \cos^2 \frac{\omega}{2}$, and (3.39) is

$$(1-x)^{\tilde{p}+p} P(x) + x^{\tilde{p}+p} P(1-x) = 1. \tag{3.40}$$

Recall that $p_0(-1) \neq 0$ and $\tilde{p}_0(-1) \neq 0$, so that $P(+1) \neq 0$. Since $(1-x)^{\tilde{p}+p}$ and $x^{\tilde{p}+p}$ have no common factors, by Bezout's identity the polynomial P can be found. There is another way to find $P(x)$. From (3.40),

$$P(x) = (1-x)^{-(\tilde{p}+p)} - x^{k}(1-x)^{-(\tilde{p}+p)} P(1-x). \tag{3.41}$$

Since $P(x)$ must be a polynomial of degree $\leq p+\tilde{p}-1$, the Taylor series of the r.h.s. of (3.41) gives that $P(x)$ is the Taylor polynomial of $(1-x)^{-(\tilde{p}+p)}$ of degree $\leq p+\tilde{p}-1$. This can be computed to obtain

$$P(x) = \sum_{n=0}^{p+\tilde{p}-1} \binom{k+n-1}{n} x^n \quad , k = p+\tilde{p} \qquad (3.42)$$

which is the polynomial of smallest degree satisfying (3.40).

EXAMPLE. For $p=\tilde{p}=1$, $P(x) = 1+2x$. Choose $p_0(\cos\omega) = 1$ and $\tilde{p}_0(\cos\omega) = 1 + 2\sin^2\frac{\omega}{2}$ (see Remark 3.15). Then, by (3.37)

$$h(\omega) = \cos^2\frac{\omega}{2} = \frac{1+\cos\omega}{2} = \frac{1}{4}e^{-i\omega} + \frac{1}{2} + \frac{1}{4}e^{i\omega}$$

so that $h[1] = h[-1] = \frac{1}{4}$ and $h[0] = \frac{1}{2}$. By (3.38)

$$\tilde{h}(\omega) = \left(\cos^2\frac{\omega}{2}\right)\left(1+2\sin^2\frac{\omega}{2}\right) = \frac{1+\cos\omega}{2}\left[1+1-\cos\omega\right]$$

$$= 1 - \frac{1}{2}\cos\omega + \frac{1}{2}\cos\omega - \cos^2\omega = 1 + \frac{1}{2}\cos\omega - \frac{1+\cos 2\omega}{2}$$

$$= \frac{3}{4} + \frac{1}{2}\cos\omega - \frac{1}{4}\cos 2\omega .$$

Thus,

$$\tilde{h}[-2] = \tilde{h}[2] = -\frac{1}{8}, \quad \tilde{h}[-1] = \tilde{h}[1] = \frac{1}{4} \quad \text{and} \quad \tilde{h}[0] = \frac{3}{4}.$$

REMARK 3.16. For wavelets $\tilde{h} = h$ and $\tilde{g} = g$; hence, we must find h such that h is a trigonometric polynomial satisfying

$$|h(\omega)|^2 + |h(\omega+\pi)|^2 = 1 \qquad (3.43)$$

and $h(0) = 1$. Since we want $h[n]$ real, $|h(\omega)|^2$ is even, so that as in (3.36)

$$|h(\omega)|^2 = p(\cos\omega). \qquad (3.44)$$

Since $h(\pi) = 0$, $p(-1) = |h(\pi)|^2 = 0$. Thus, we can write $p(x) = (1+x)^p q(x)$ for some $p = 1, 2, 3, \ldots$. Thus

$$|h(\omega)|^2 = p(\cos\omega) = (1+\cos\omega)^p P(\cos\omega) = \left(\cos^2\frac{\omega}{2}\right)^p P(\cos\omega) \qquad (3.45).$$

The condition (3.43) gives, as in (3.40)

$$(1-x)^p P(x) + x^p P(1-x) = 1 \qquad (3.46)$$

with $x = \sin^2\frac{\omega}{2} \in [0,1]$. As in (3.42), $P(x) = \sum_{n=0}^{p-1} \binom{p+n-1}{n} x^n \geq 0$, so that

$$|h(\omega)|^2 = \left(\cos^2\frac{\omega}{2}\right)^p \sum_{n=0}^{p-1} \binom{p+n-1}{n} \left(\sin^2\frac{\omega}{2}\right)^n$$

$$= \left|\frac{1+e^{-i\omega}}{2}\right|^{2p} \sum_{n=0}^{p-1} \binom{p+n-1}{n} \left(\sin^2\frac{\omega}{2}\right)^n \qquad (3.47)$$

A "square root" of the positive trigonometric polynomial $P\left(\sin^2\frac{\omega}{2}\right)$ can be found by a theorem of Fejer and Riesz. This allows us to compute $h(\omega)$.

For the simple case $p = 1$, $|h(\omega)|^2 = \left|\frac{1+e^{-i\omega}}{2}\right|^2$, so that $h(\omega) = \frac{1+e^{-i\omega}}{2} \Rightarrow h[0] = h[1] = \frac{1}{2}$ (Haar filter).

REMARK 3.17. For finite discrete signals $c_j[n]$, $0 \leq n \leq N-1$ the convolutions in the decomposition and reconstruction formulas can be interpreted in several ways

- $c_j[n] = 0$ for $n \notin [0, N-1]$
- $c_j[n]$ is periodize with period N, and thus, the convolutions are circular convolutions
- $c_j[n]$ is made even with respect to $-\frac{1}{2}$ and then periodize with period $2N$
- Find appropriate "border" filters.

3.10 THE CASE OF IMAGES

The definition of MRA given in section 3.1 for $L^2(\mathbb{R})$ extends to $L^2(\mathbb{R}^2)$. To indicate the dimension we shall write $\{V_j^{(2)} : j \in \mathbb{Z}\}$ for the family of subspaces, and $\varphi^{(2)}(x,y)$ the scaling function.

A 2D-MRA can be constructed from a 1D-MRA with scaling function φ. It is enough to consider

$$V_j^{(2)} \equiv V_j \otimes V_j = \{ f(x,y) = \sum_{n=-\infty}^{\infty} a[n] f_n(x) g_n(y) \,,\, f_n \in V_j \,,\, g_n \in V_j \}^{L^2(\mathbb{R})}$$

and $\varphi^{(2)}(x,y) = \varphi(x)\varphi(y)$.

EXERCISE: Show that $\{\varphi^{(2)}(x-k_1, y-k_2) : k_1, k_2 \in \mathbb{Z}\}$ is an orthonormal basis of $V_0^{(2)} = V_0 \otimes V_0$.

In order to find an o.n. basis of $L^2(\mathbb{R}^2)$ of the type consider in definition 3.1 it is enough to find a "translation" basis of $W_0^{(2)} = V_1^{(2)} \ominus V_0^{(2)}$. The following "formal" computation gives you a hint:

$$V_1^{(2)} \equiv V_1 \otimes V_1 = (V_0 \oplus W_0) \otimes (V_0 \oplus W_0) =$$

$$= (V_0 \otimes V_0) \oplus (V_0 \otimes W_0) \oplus (W_0 \otimes V_0) \oplus (W_0 \otimes W_0)$$

$$= V_0^{(2)} \oplus (V_0 \otimes W_0) \oplus (W_0 \otimes V_0) \oplus (W_0 \otimes W_0)$$

Hence,

$$W_0^{(2)} = (V_0 \otimes W_0) \oplus (W_0 \otimes V_0) \oplus (W_0 \otimes W_0) \qquad (3.48)$$

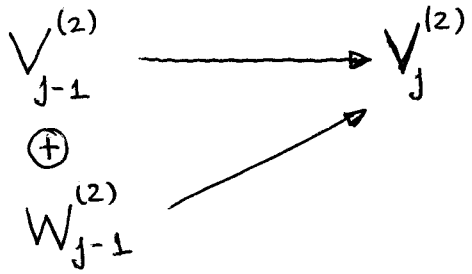
Define

$$\psi^1(x,y) = \varphi(x)\psi(y) \,,\, \psi^2(x,y) = \psi(x)\varphi(y) \,,\, \psi^3(x,y) = \psi(x)\psi(y)$$

so that the orthonormality of [illegible] of φ and ψ, and (3.48) shows that

$$\{\psi^l(x-k_1, y-k_2) : k_1, k_2 \in \mathbb{Z},\ l = 1,2,3\}$$

is an o.n. basis of $W_0^{(2)}$. Since $L^2(\mathbb{R}^2) = \bigoplus_{j=-\infty}^{\infty} W_j^{(2)}$, the collection $\{\psi^1, \psi^2, \psi^3\}$ is an o.n. wavelet system for $L^2(\mathbb{R}^2)$.



The diagram on the left suggest how to find a coarse approximation

$$c_{j-1}[k_1,k_2] = \langle f, \varphi_{j-1,k_1}^{(1)} \varphi_{j-1,k_2}^{(1)} \rangle$$

from the original weffeicients

$$c_j[k_1,k_2] = \langle f, \varphi_{j,k_1}^{(1)} \varphi_{j,k_2}^{(1)} \rangle .$$

A computation similar to the one leading to (3.5) shows

$$c_{j-1}[p,q] = 2 \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} \overline{h[k-2p]}\, \overline{h[l-2q]}\, c_j[k,l]$$

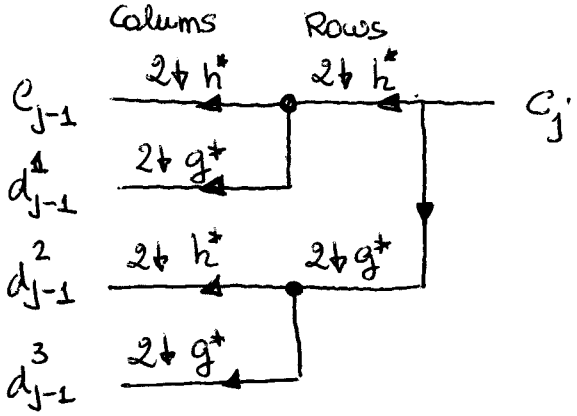
$$= 2\, c_j * \tilde{h}\,\tilde{h}\,[2p,2q] \qquad (3.49)$$

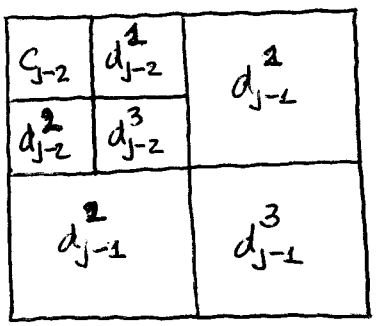
Three different type of details $d_{j-1}^1[p,q]$, $d_{j-1}^2[p,q]$ and $d_{j-1}^3[p,q]$ are now obtained corresponding to each one of the functions $\psi^1(x,y) = \varphi(x)\psi(y)$, $\psi^2(x,y) = \psi(x)\varphi(y)$ and $\psi^3(x,y) = \psi(x)\psi(y)$.

One obtains formulas similar to (3.49) with appropriate combinations of the 1D-filters h and g:

$$d_{j-1}^1[p,q] = 2\, c_j * \tilde{h}\,\tilde{g}\,[2p,2q] \quad , \quad d_{j-1}^2[p,q] = 2\, c_j * \tilde{g}\,\tilde{h}\,[2p,2q]$$

$$\text{and} \quad d_{j-1}^3[p,q] = 2\, c_j * \tilde{g}\,\tilde{g}\,[2p,2q]. \qquad (3.50)$$





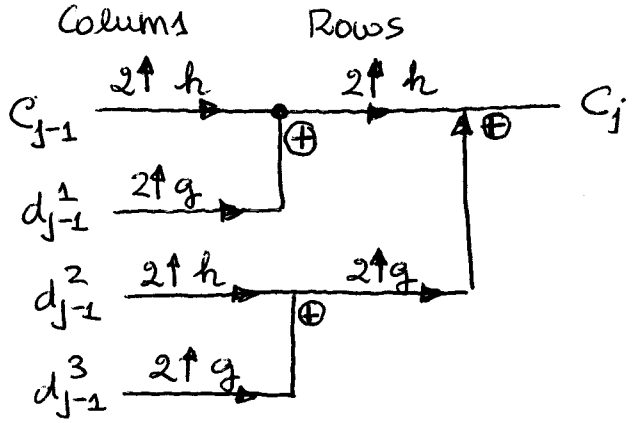
2D representation

Formulas (3.49), (3.50) require to do four 2D-convolutions, while the left diagram shows that six 1D-convolutions are enough if done by rows and then by columns.

As in Section 3.6 (see formula (3.19)) the reconstruction algorithm is

$$C_j[p,q] = 2\,\overset{v}{C}_{j-1} * hh[p,q] + 2\,\overset{v}{d}^{1}_{j-1} * hg[p,q] + 2\,\overset{v}{d}^{2}_{j-1} * gh[p,q] + 2\,\overset{v}{d}^{3}_{j-1} * gg[p,q] \qquad (3.51)$$

where $\overset{v}{x}[n] = \begin{cases} x[p] & \text{if } n=2p \\ 0 & \text{if } n=2p+1 \end{cases}$ is the upsampling of $x[n]$.



Again, six 1D-convolutions are enough to compute (3.51) with colums and then with rows as in the above diagram, instead of four 2D-convolutions needed to perform (3.51) directly.

The complexity of the reconstruction and decomposition algorithms can be shown to be CK^2N^2 for an inicial array of N^2 data, using finite filters of length K.

---

4. TREES OF DISCRETE BASES, WAVELET PACKETS AND BEST APPROXIMATION.

OBJECTIVE: To construct a library of bases using either the discrete cosine bases or wavelet basis and to explain how to select the best basis for a given signal or image.

4.1. DISCRETE BLOCK BASES

Let $\{(C_k[n])_{n=0}^{L-1}\}_{k=0}^{L-1}$ an o.n. basis of the space of discrete signals $(f[n])_{n=0}^{L-1}$ (for example DCI). Given $a<b$, $a,b\in\mathbb{Z}$ and $L=b-a$, the above basis can be adapted to the space of discrete signals $(g[n])_{n=a}^{a+L-1}$ by translations:

$$\{(g_k[n])_{n=a}^{a+L-1} = (C_k[n-a])_{n=a}^{a+L-1}\}_{k=0}^{L-1} \qquad (4.1)$$

Let $W\approx L^2[0,N)$ be the space of discrete signals of length N.

$$a_0=0 < a_1 < a_2 < \cdots < a_{M-2} < a_{M-1} < N=a_M \qquad a_j\in\mathbb{N}$$

Let $L_p=a_{p+1}-a_p$, $p=0,1,\ldots,M-1$. For each discrete interval $[a_p,a_{p+1})\cap\mathbb{N}$ we consider the basis $\{(g_{psk}[n])_{n=a_p}^{a_p+L_p-1}\}_{k=0}^{L_p-1}$. Writing W^p as the space of signals of length L_p defined on $[a_p,a_{p+1})\cap\mathbb{N}$ we have

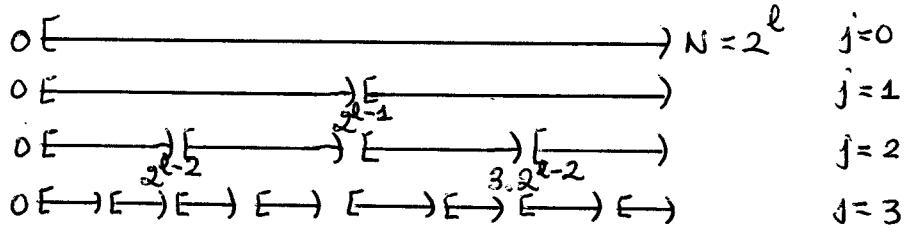
$$W = W^0\oplus W^1\oplus\cdots\oplus W^{M-1}$$

In this way we obtain the discrete block basis

$$\{(g_{pk}[n])_{n=a_p}^{a_p+L_p-1}\}_{k=0}^{L_p-1},\ p=0,\ldots,M-1 \qquad (4.2)$$

4.2. DYADIC TREE OF DISCRETE BASES

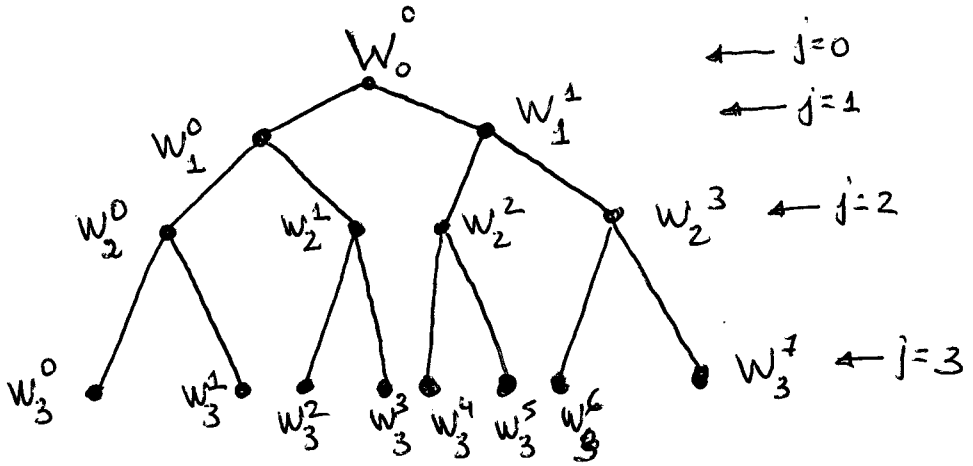
Dyadic division of a discrete interval $[0,N)$, $N=2^l$, reduces programing tasks and produces a dyadic tree of bases.



For each $j \geq 0$, the discrete interval $[0, 2^l)$ is divided in 2^j discrete intervals each one of length 2^{l-j}. Write $a_{p,j} = p\,\frac{2^l}{2^j}$, $0 \leq p \leq 2^j - 1$. Each discrete interval $[a_{p,j}, a_{p+1,j})$ has a discrete basis as in (5.2):

$$\mathcal{B}_j^p = \left\{ \left(g_{j,k}^p[n] \right)_{n=a_{p,j}}^{a_{p,j}+2^{l-j}} \right\}_{k=0}^{2^{l-j}} \qquad (4.3)$$

that generates a space W_j^p of signals of length 2^{l-j} for each $p=0,1,\ldots,2^j-1$. Thus, for each level j $(0 \leq j \leq 2^l)$ $\left\{ \mathcal{B}_j^p \right\}_{p=0}^{2^j-1}$ is an orthonormal bases of $L^2[0,N)$.



Let B_j be the number of orthonormal bases in a tree of depth j. Then, $B_0 = 1$, $B_1 = 2$, $B_2 = 5$. In general $B_j = 1 + B_{j-1}^2 \geq B_{j-1}^2 \geq B_{j-2}^4 \geq \ldots \geq (B_1)^{2^{j-1}} = 2^{2^{j-1}}$. It can be shown that $B_j \leq 2^{\frac{5}{4}\cdot 2^{j-1}}$.

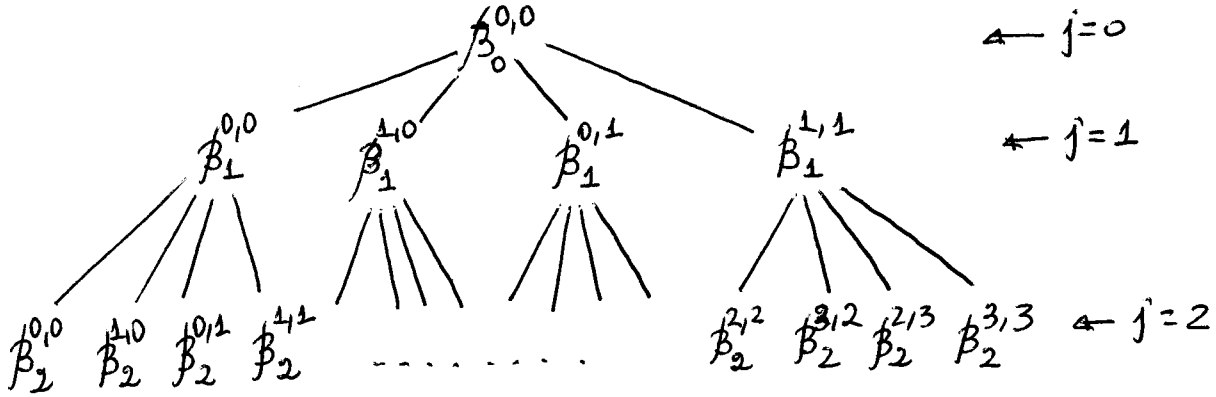
4.3. TREES OF DISCRETE BASES FOR IMAGES

Consider an image given by $N \times N$ samples, defined on $[0,N) \times [0,N)$. With $N = 2^l$, the original square $[0,N)^2$ is divided in four squares of side length $N/2$. The process could be continued. Each one of the new squares can carry a discrete basis.

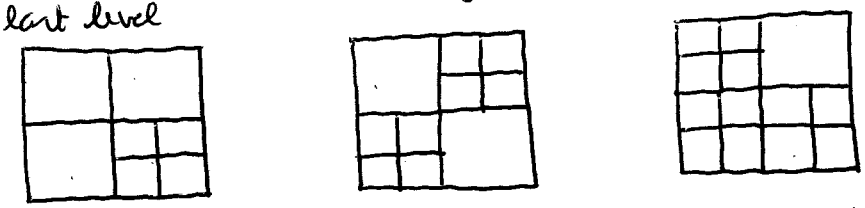
Let $a_{p,j} = p\,\frac{2^l}{2^j}$, $0 \leq p \leq 2^j - 1$. In each square $[a_{p,j}, a_{p+1,j}) \times [a_{q,j}, a_{q+1,j})$ we put an o.n. discrete basis (for example, DCI) obtaining

$$\mathcal{B}_j^{p,q} = \left\{ \left(g^p_{j,k}[n]\, g^q_{j,\ell}[m] \right)_{n=a_p,\, m=a_q}^{a_p+L_p-1,\, a_q+L_q-1} \right\}_{k=0,\, \ell=0}^{2^{l-j}-1,\, 2^{l-j}-1}$$

where $L_p = a_{p+1} - a_p = 2^{l-j}$. Varying p, q with $0 \leq p, q \leq 2^j - 1$ we obtain a block basis of $L^2[0,N)^2$ at level j.

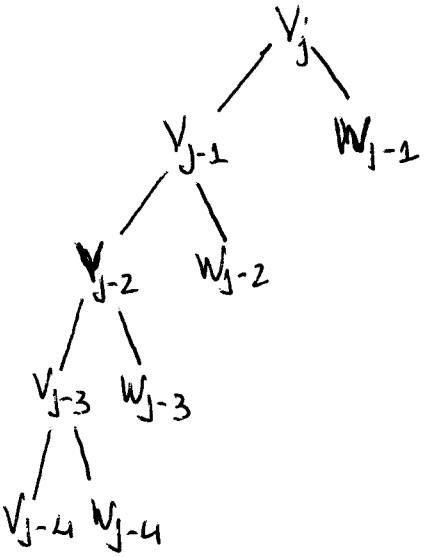


Other o.n. bases of $L^2[0,N)^2$ can be obtained choosing bases from different nodes of the quad-tree, taking into consideration that the four bases of a node $\mathcal{B}_j^{p,q}$ have to be chosen of one of them as, at the last level



Let $B_j^{(2)}$ be the number of o.n. bases in a quad-tree of depth j. Then, $B_0^{(2)} = 1$, $B_1^{(2)} = 2$, $B_2^{(2)} = 1 + 2^4 = 17$. In general, $B_j^{(2)} = 1 + \left[B_{j-1}^{(2)}\right]^4 \geq \left[B_{j-1}^{(2)}\right]^4 \geq \ldots \geq \left[B_1^{(2)}\right]^{4^{j-1}} = 2^{4^{j-1}}$. It can be shown that $B_j^{(2)} \leq e^{\frac{49}{48} \cdot 4^{j-1}}$.

4.4. WAVELET PACKETS



In the wavelet transform algorithm only the spaces V_j are decompose, while the "detail" spaces W_j are not. R. Coifman, Y. Meyer and M. V. Wickerhauser (1992) that wave packets could be obtained by decomposing the detail spaces W_j.

There are many ways of decomposing and o.n. bases in a Hilbert space $\mathbb{H}$ into two o.n. bases.

EXAMPLE 1. Let $\{e_k : k \in \mathbb{Z}\}$ be an o.n. basis of a Hilbert space $\mathbb{H}$. With

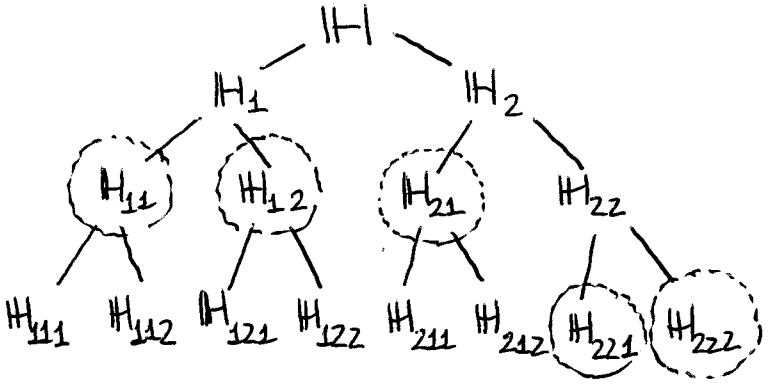
$$\mathbb{H}_1 = \overline{\{e_{2k}\}} \quad \text{and} \quad \mathbb{H}_2 = \overline{\{e_{2k+1}\}}$$

one has $\mathbb{H} = \mathbb{H}_1 \oplus \mathbb{H}_2$

EXAMPLE 2. Let $\{e_k : k \in \mathbb{Z}\}$ be an o.n. basis of a Hilbert space $\mathbb{H}$. Set

$$u_k = \frac{1}{\sqrt{2}}[e_{2k-1} + e_{2k}] \quad \text{and} \quad v_k = \frac{1}{\sqrt{2}}[e_{2k-1} - e_{2k}], \; k \in \mathbb{Z}$$

and $\mathbb{H}_1 = \overline{\{u_k\}}$, $\mathbb{H}_2 = \overline{\{v_k\}}$. Since $u_k + v_k = \sqrt{2}\, e_{2k-1}$ and $u_k - v_k = \sqrt{2}\, e_{2k}$ we have $\mathbb{H} = \mathbb{H}_1 + \mathbb{H}_2$. It can be proved (exercise) that $\langle u_k, v_\ell \rangle = 0$ all $k, \ell \in \mathbb{Z}$, so that $\mathbb{H} = \mathbb{H}_1 \oplus \mathbb{H}_2$. Moreover, $\{u_k\}$ and $\{v_k\}$ are o.n. systems.



This wave packet tree is done with the Haar filters. The tree gives many o.n bases for $\mathbb{H}$.

THEOREM 4.1 (Coifman, Meyer, Wickerhauser ~1992)

Let $\{\theta_{j,k} : k \in \mathbb{Z}\}$ be an o.n. basis of a space U_j. Let h and g be a pair of filters that satisfy

$$\left.\begin{aligned} &a)\ |h(\omega)|^2 + |h(\omega+\pi)|^2 = 1 \\ &b)\ |g(\omega)|^2 + |g(\omega+\pi)|^2 = 1 \\ &c)\ h(\omega)\overline{g(\omega)} + h(\omega+\pi)\overline{g(\omega+\pi)} = 0 \end{aligned}\right\} \quad (4.4)$$

Define, for $k \in \mathbb{Z}$

$$\theta^1_{j-1,k}(t) = \sqrt{2}\sum_{n=-\infty}^{\infty} h[n-2k]\,\theta_{j,n}(t) \quad \text{and} \quad \theta^2_{j-1,k}(t) = \sqrt{2}\sum_{n=-\infty}^{\infty} g[n-2k]\,\theta_{j,n}(t). \quad (4.5)$$

Let

$$U^1_{j-1} = \overline{\{\theta^1_{j-1,k} : k \in \mathbb{Z}\}} \quad \text{and} \quad U^2_{j-1} = \overline{\{\theta^2_{j-1,k} : k \in \mathbb{Z}\}}.$$

Then

$$U_j = U^1_{j-1} \oplus U^2_{j-2}$$

and the families given in (4.5) are o.n. bases of U^1_{j-1} and U^2_{j-1} respectively.

REMARK 4.2. Conditions (4.4) are equivalent to the matrix

$$M(\omega) = \begin{pmatrix} h(\omega) & h(\omega+\pi) \\ g(\omega) & g(\omega+\pi) \end{pmatrix}$$ been unitary.

LEMMA 4.3.

Part a) of (4.4) is equivalent to

$$2\sum_{l \in \mathbb{Z}} h[l-2k]\,\overline{h[l]} = \delta_{k,0}, \quad k \in \mathbb{Z}$$

PROOF. $1 = \left(\sum_k h[k] e^{-ik\omega}\right)\left(\sum_l \overline{h[l]}\, e^{il\omega}\right) + \left(\sum_k h[k](-1)^k e^{-ik\omega}\right)$.

$$\left(\sum_l \overline{h[l]}(-1)^l e^{il\omega}\right) =$$

$$= \sum_k \left\{ h[k]\left(\sum_l \overline{h[l]}\, e^{i(l-k)\omega} + \sum_l \overline{h[l]}(-1)^{l+k} e^{i(l-k)\omega}\right)\right\}$$

$$\overset{l-k=m}{=} \sum_{m\in\mathbb{Z}} \Big\{ \sum_{l\in\mathbb{Z}} h[l-m]\,\overline{h[l]}\,\big[1+(-1)^m\big] \Big\} e^{im\omega} =$$

$$= \sum_{k\in\mathbb{Z}} \Big\{ 2 \sum_{l\in\mathbb{Z}} h[l-2k]\,\overline{h[l]} \Big\} e^{i2k\omega} \quad \therefore \quad 2\sum_{l\in\mathbb{Z}} h[l-2k]\,\overline{h[l]} = \delta_{k,0}$$

■

LEMMA 4.4.

Part b) of (4.4) is equivalent to

$$2 \sum_{l\in\mathbb{Z}} g[l-2k]\,\overline{g[l]} = \delta_{k,0}, \quad k\in\mathbb{Z}$$

LEMMA 4.5

Part c) of (4.4) is equivalent to $\sum_{l\in\mathbb{Z}} h[l-2k]\,\overline{g[l]} = 0$.

PROOF (OF THEOREM 4.1).

$$\langle \theta^1_{j-1,k}, \theta^1_{j-1,l} \rangle = \Big\langle \sqrt{2} \sum_{n=-\infty}^{\infty} h[n-2k]\,\theta_{j,m}, \sqrt{2} \sum_{n=-\infty}^{\infty} h[n-2l]\,\theta_{j,m} \Big\rangle$$

$$= 2 \sum_{n=-\infty}^{\infty} h[n-2k]\,\overline{h[n-2l]} \overset{n-2l=s}{=\!=} 2 \sum_{s=-\infty}^{\infty} h[s+2l-2k]\,\overline{h[s]}$$

$$= \delta_{l-k,0} = \delta_{l,k} \quad \text{by lemma 4.3.}$$

Similarly, due to lemma 4.4, one shows

$$\langle \theta^2_{j-1,k}, \theta^2_{j-1,l} \rangle = \delta_{l,k}.$$

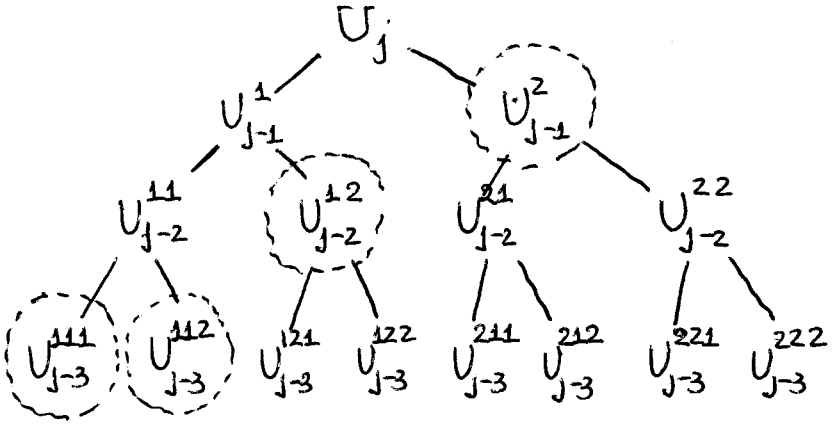
Using lemma 4.5 one can show

$$\langle \theta^1_{j-1,k}, \theta^2_{j-1,l} \rangle = 0 \quad , \; k,l\in\mathbb{Z}.$$

Thus shows $U^1_{j-1} \oplus U^2_{j-1} \subset U_j$. Equality follows from

$$\theta_{j,l} = \sum_{k\in\mathbb{Z}} \sqrt{2}\,\overline{h[l-2k]}\,\theta^1_{j-1,k} + \sum_{k\in\mathbb{Z}} \sqrt{2}\,\overline{g[l-2k]}\,\theta^2_{j-1,k} \qquad (4.6)$$

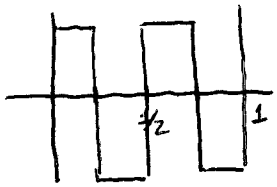
■

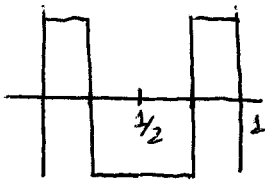


Orthonormal baxs of U_j can be obtained by choosing o.n. baxs of different nodes of the tree.

EXAMPLE. Wavelet packets corresponding to the Haar wavelet give the Walsh series. With $\psi(t) = \chi_{[0,1/2)}(t) - \chi_{[1/2,1)}(t)$ we have

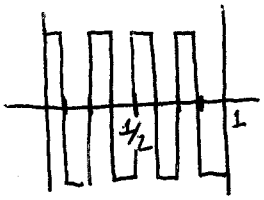
$$\psi^1(t) = \psi(2t) + \psi(2t-1) \quad ; \quad \psi^2(t) = \psi(2t) - \psi(2t-1)$$

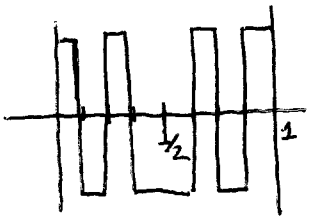




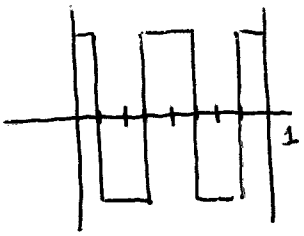
At the second level

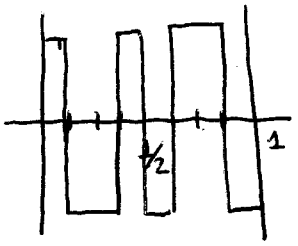
$$\psi^{11}(t) = \psi^1(2t) + \psi^1(2t-1) \quad ; \quad \psi^{12}(t) = \psi^1(2t) - \psi^1(2t-1)$$





$$\psi^{21}(t) = \psi^2(2t) + \psi^2(2t-1) \quad ; \quad \psi^{22}(t) = \psi^2(2t) - \psi^2(2t-1)$$





4.5. BEST BASIS AND CONCAVITY

Consider a dictionary $\mathcal{D}$ that is made of orthonormal bases in a signal space of finite dimension N: $\mathcal{D} = \bigcup_{\lambda \in \Lambda} \mathcal{B}^\lambda$. Each orthonormal basis is a family of N vectors

$$\mathcal{B}^\lambda = \{ g_m^\lambda \}_{m=1}^N$$

Wavelet packets and cosine trees are examples of dictionaries.

Given a signal or image f, and a basis $\mathcal{B}^\lambda$, the non-linear approximation of f from $\mathcal{B}^\lambda$ is defined in the following way: let I_M^λ be the index set of the M vectors of $\mathcal{B}^\lambda$ that maximize $|\langle f, g_m^\lambda \rangle|$; then best non-linear approximation of f by M terms is

$$f_M^\lambda = \sum_{m \in I_M^\lambda} \langle f, g_m^\lambda \rangle \, g_m^\lambda \qquad (4.7)$$

The approximation error is

$$E^\lambda[M] = \sum_{m \notin I_M^\lambda} |\langle f, g_m^\lambda \rangle|^2 = \|f\|_2^2 - \sum_{m \in I_M^\lambda} |\langle f, g_m^\lambda \rangle|^2. \qquad (4.8)$$

DEFINITION 4.6.

We say that $\mathcal{B}^\alpha = \{ g_m^\alpha \}_{1 \leq m \leq N}$ is a better basis than $\mathcal{B}^\gamma = \{ g_m^\gamma \}_{1 \leq m \leq N}$ for approximating f if for all $M \geq 1$

$$E^\alpha[M] \leq E^\gamma[M]. \qquad (4.9)$$

Inserting (4.8) proves that the better basis condition (4.9) is equivalent to

$$\sum_{m \in I_M^\alpha} |\langle f, g_m^\alpha \rangle|^2 \geq \sum_{m \in I_M^\gamma} |\langle f, g_m^\gamma \rangle|^2 \quad , \forall M \geq 1. \qquad (4.10)$$

The following theorem derives a criteria to choose a best basis if it exists. Notice that (4.9) is a partial order relation, so that best basis may not exists.

THEOREM 4.7.

A basis $\mathcal{B}^\alpha$ is a better basis than $\mathcal{B}^\gamma$ for approximating f if and only if for all concave functions $\Phi(u)$

$$\sum_{m=1}^{N} \Phi\left(\frac{|\langle f, g_m^\alpha\rangle|^2}{\|f\|_2^2}\right) \leq \sum_{m=1}^{N} \Phi\left(\frac{|\langle f, g_m^\gamma\rangle|^2}{\|f\|_2^2}\right) \qquad (4.11)$$

The proof is based on a classical result due to Hardy, Littlewood, and Polya.

LEMMA 4.8.

Let $X[m] \geq 0$ and $y[m] \geq 0$ be such that

$$X[m] \geq X[m+1] \quad \text{and} \quad y[m] \geq y[m+1] \qquad \forall m \in [1, N] \qquad (4.12)$$

and satisfying $\sum_{m=1}^{N} X[m] = \sum_{m=1}^{N} y[m]$. For all $M \leq N$ these sequences satisfy

$$\sum_{m=1}^{M} X[m] \geq \sum_{m=1}^{M} y[m] \qquad (4.13)$$

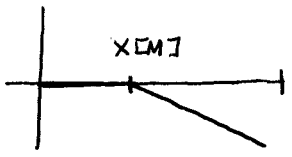
if and only if for all concave functions $\Phi(u)$

$$\sum_{m=1}^{N} \Phi(X[m]) \leq \sum_{m=1}^{N} \Phi(y[m]) \qquad (4.14)$$

PROOF OF (4.14) $\Rightarrow$ (4.13).

(4.14) is valid for all concave Φ. In particular, is valid for the following family of concave functions

$$\Phi_M(u) = \begin{cases} X[M] - u & \text{if } u \leq X[M] \\ 0 & \text{otherwise} \end{cases}$$



We have

$$\sum_{m=1}^{N} \Phi_M(X[m]) = \sum_{m=1}^{M} X[M] - X[m] = M\,X[M] - \sum_{m=1}^{M} X[m] \qquad (4.15)$$

The hypothesis (4.14) implies

$$\sum_{m=1}^{N} \Phi_M(X[m]) \leq \sum_{m=1}^{N} \Phi_M(y[m]) \qquad \forall M. \qquad (4.15)$$

Moreover, $\Phi_M(u) \leq 0$ and $\Phi_M(u) \leq X[M] - u$, so that

$$M x[M] - \sum_{m=1}^{M} x[m] = \sum_{m=1}^{N} \Phi_M(x[m]) \overset{(4.16)}{\leqq} \sum_{m=1}^{N} \Phi_M(y[m]) \leqq$$

$$\leqq \sum_{m=1}^{M} \Phi_M(y[m]) \leqq \sum_{m=1}^{M} x[m] - y[m] = M x[M] - \sum_{m=1}^{M} y[m] .$$

$\uparrow$ $\Phi_M(u) \leqq x[M] - u$

This gives (4.13). ∎

PROOF OF THEOREM 4.7.

For any basis $\mathcal{B}^\lambda$, sort the inner products $|\langle f, g^\lambda_m \rangle|$ so that

$$|\langle f, g^\lambda_{m_1} \rangle| \geqq |\langle f, g^\lambda_{m_2} \rangle| \geqq |\langle f, g^\lambda_{m_3} \rangle| \geqq \cdots .$$

and denote

$$x^\lambda[k] = \frac{|\langle f, g^\lambda_{m_k} \rangle|^2}{\|f\|_2^2} \geqq \frac{|\langle f, g^\lambda_{m_{k+1}} \rangle|^2}{\|f\|_2^2} = x^\lambda[k+1]$$

By Plancherel, $\sum_{k=1}^{N} x^\lambda[k] = 1$. Condition (4.10) shows that $\mathcal{B}^\alpha$ is better than $\mathcal{B}^\gamma$ if and only if for all $M \geqq 1$

$$\sum_{k=1}^{M} x^\alpha[k] \geqq \sum_{k=1}^{M} y^\alpha[k]$$

Lemma 4.8 shows that this is equivalent to have

$$\sum_{k=1}^{N} \Phi(x^\alpha[k]) \leqq \sum_{k=1}^{N} \Phi(x^\gamma[k])$$

for all concave functions Φ. Thus condition coincides with (4.11). ∎

4.6. EXAMPLES OF COST FUNCTIONS

In practice, a single concave function $\Phi(u)$ is used to compare two bases. The <u>COST</u> of approximating f in a basis $\mathcal{B}^\lambda$ is defined by

$$C(f, \mathcal{B}^\lambda) = \sum_{m=1}^{N} \Phi\left(\frac{|\langle f, g_m^\lambda \rangle|^2}{\|f\|^2} \right).$$

Theorem 4.7 shows that if $\mathcal{B}^\alpha$ is better than $\mathcal{B}^\gamma$ for approximating f, then $C(f, \mathcal{B}^\alpha) \leq C(f, \mathcal{B}^\gamma)$. But this condition is not sufficient to prove that $\mathcal{B}^\alpha$ is better than $\mathcal{B}^\gamma$ since a single concave function has been used.

<u>IDEAL BASIS</u>. Given f, an ideal basis $\mathcal{B}^I$ for approximating f has one of its vectors proportional to f, that is $g_m = \eta f$ with $|\eta| = 1/\|f\|$. Thus, we need only one vector to recover f. If $\Phi(0) = 0$, we have

$$C(f, \mathcal{B}^I) = \Phi(1).$$

An ideal basis is clearly better than any other basis since it produces a zero error for $M \geq 1$, and hence (4.9) holds for all M. Thus, for any o.n. basis $\mathcal{B}$,

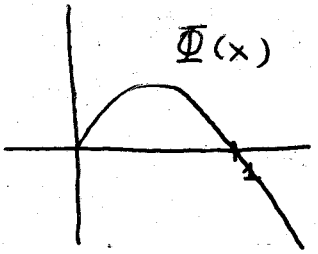
$$C(f, \mathcal{B}) \geq C(f, \mathcal{B}^I) = \Phi(1).$$

<u>ENTROPY COST</u>. The function

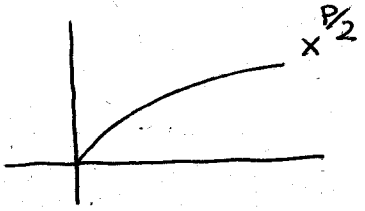
$$\Phi(x) = -x \ln x$$

is concave for $x \geq 0$. The cost function is called entropy of the energy distribution

$$C(f, \mathcal{B}) = -\sum_{m=1}^{N} \frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \ln \frac{|\langle f, g_m \rangle|^2}{\|f\|^2}$$



ℓ^p COST. For $p<2$, $\Phi(x)=x^{p/2}$ is concave for $x\geq 0$.



The cost function is

$$C(f,\mathcal{B})=\sum_{m=1}^{N}\frac{|\langle f,g_m\rangle|^p}{\|f\|^p}=\frac{1}{\|f\|^p}\,\|\{\langle f,g_m\rangle\}\|_{\ell^p}^p$$

In this case

$$E[M]\leq\frac{\|f\|^2\,C(f,\mathcal{B})^{\frac{2}{p}}}{\frac{2}{p}-1}\,\frac{1}{M^{\frac{2}{p}-1}}\qquad(4.17)$$

We prove (4.17). Write

$$|\langle f,g_{m_1}\rangle|\geq|\langle f,g_{m_2}\rangle|\geq|\langle f,g_{m_3}\rangle|\geq\cdots$$

Then, for any $n\in\mathbb{N}$,

$$\|\{\langle f,g_m\rangle\}\|_{\ell^p}^p=\sum_{k=1}^{N}|\langle f,g_{m_k}\rangle|^p\geq\sum_{k=1}^{n}|\langle f,g_{m_k}\rangle|^p\geq n\,|\langle f,g_{m_n}\rangle|^p,$$

which shows

$$|\langle f,g_{m_n}\rangle|\leq\frac{\|\{\langle f,g_m\rangle\}\|_{\ell^p}}{n^{1/p}},\quad n\in\mathbb{N}\qquad(4.18).$$

Therefore,

$$E[M]=\sum_{k=M+1}^{N}|\langle f,g_{m_k}\rangle|^2\leq C(f,\mathcal{B})^{2/p}\|f\|^2\sum_{k=M+1}^{N}1/k^{2/p}$$

$$\leq C(f,\mathcal{B})^{2/p}\|f\|^2\sum_{k=M+1}^{\infty}\frac{1}{k^{2/p}}\leq\frac{C(f,\mathcal{B})^{2/p}\|f\|^2}{(\frac{2}{p}-1)}\,\frac{1}{M^{\frac{2}{p}-1}}$$

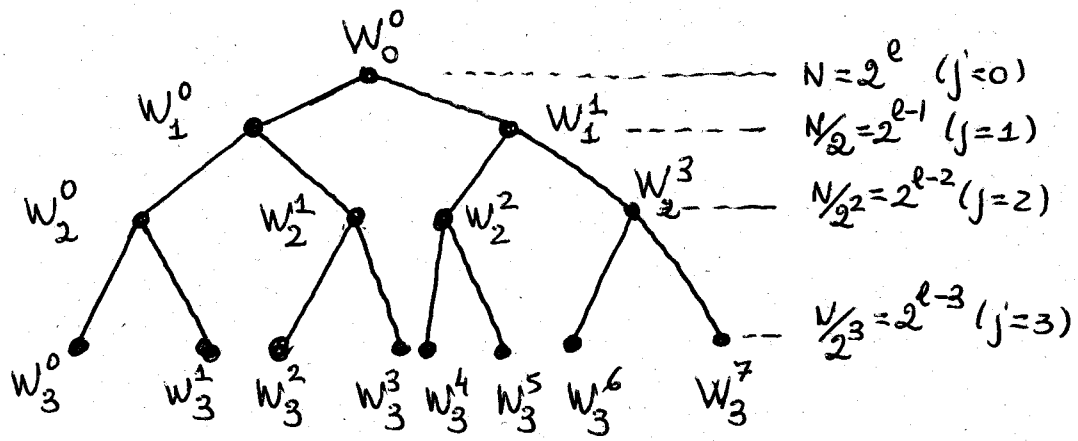
as we wanted to prove. ■

4.7. BEST BASIS SEARCH IN TREES

Trees of discrete bases and wavelet packets were described in sections 4.2, 4.3 and 4.4. For dyadic trees and dyadic wave packets it was proved that for signals of size N, these dictionaries include more than $2^{N/2} = 2^{2^{l-1}}$ $(N=2^l)$ bases.

Given a signal f of size $N=2^l$, minimizing the cost $C(f, \mathcal{B}^\lambda)$ for all bases $\mathcal{B}^\lambda$ requires more than $N 2^{N/2}$ operations.

The dynamic programming algorithm of Coifman and Wickerhauser finds the best basis with $O(N \log_2 N)$ operations, by taking advantage of the tree structure.



In a tree each node at depth j is indexed by its position p from left to right. At this node, the space W_j^p has dimension $N/2^j$ and is generated by an orthonormal basis $\mathcal{B}_j^p$. This space is divided in two orthogonal subspaces

$$W_j^p = W_{j+1}^{2p} \oplus W_{j+1}^{2p+1} .$$

In addition to $\mathcal{B}_j^p$, the space W_j^p has also the orthonormal bases $\mathcal{B}_{j+1}^{2p} \cup \mathcal{B}_{j+1}^{2p+1}$ made with bases from each one of the children nodes. Many other bases can be

found in W_j^p using the tree structure.

Recall that for a basis $\mathcal{B} = \{g_m : m = 0 \ldots, 2^j N - 1\}$ of W_j^p and a fixed concave function Φ, the cost is

$$C(f, \mathcal{B}) = \sum_{m=0}^{2^j N - 1} \Phi\left(\frac{|\langle f, g_m \rangle|^2}{\|f\|^2}\right). \qquad (4.19)$$

This cost is additive, since for any two orthonormal bases $\mathcal{B}^0$ and $\mathcal{B}^1$ of two different orthogonal subspaces

$$C(f, \mathcal{B}^0 \cup \mathcal{B}^1) = C(f, \mathcal{B}^0) + C(f, \mathcal{B}^1) \qquad (4.20)$$

The <u>best basis</u> $\mathcal{O}_j^p$ of W_j^p is the basis that minimizes (4.19) among all the bases of W_j^p that can be constructed from the tree.

PROPOSITION 4.9 (Coifman, Wickerhauser ~ 1992)

If C is an additive cost function, then

$$\mathcal{O}_j^p = \begin{cases} \mathcal{O}_{j+1}^{2p} \cup \mathcal{O}_{j+2}^{2p+1} & \text{if } C(f, \mathcal{O}_{j+1}^{2p}) + C(f, \mathcal{O}_{j+1}^{2p+1}) < C(f, \mathcal{B}_j^p) \\ \mathcal{B}_j^p & \text{if } C(f, \mathcal{O}_{j+1}^{2p}) + C(f, \mathcal{O}_{j+1}^{2p+1}) \geq C(f, \mathcal{B}_j^p) \end{cases}$$

The best basis of W_0^0 is obtained by finding the best bases of all spaces W_j^p in the tree, starting with the subspaces W_J^p at the leaves of the tree. At this level, the best basis of W_J^p is $\mathcal{B}_J^p$. In W_{J-1}^p the best basis is found by the algorithm of proposition 4.9. Repeating this argument for $J > j \geq 0$ we obtain the best basis of f in W_0^0 that minimizes the cost for a given concave function Φ.

COMPLEXITY.

In W_j^P the wavelet packets coefficients $\langle f, g_m \rangle$ for $\mathcal{B}_j^P$ are calculated in $O(2^{-j}N)$ operations. Thus, all the coefficients at level j are calculated with $O(N)$ operations. Since there are at most $l = \log_2 N$ levels, the computation of all the inner products of f with all vectors in the tree require $O(N \log_2 N)$ operations.

At each node, the cost, calculated with (4.19), requires $O(2^{-j}N)$ operations. Since there are 2^j o.n. bases at level j, the best basis of all $\{W_j^P : j = 0, \dots, 2^j - 1\}$, computed with proposition 4.9, using the bases of the spaces $\{W_{j+1}^P : j = 0, \dots, 2^{j+1} - 1\}$, requires $O(N)$ operations. Since J is at most $l = \log_2 N$, the best basis of W_0^0 is selected with $O(N \log_2 N)$ operations in the case of wavelet packets.

For the discrete cosine tree, the inner products for $\mathcal{B}_j^P$ requires $O(2^{-j}N \log_2 N)$ operations. An argument as in the first paragraph shows that the inner products of f with all vectors of a discrete cosine trees require $O(N (\log_2 N)^2)$ operations.

The complexity of the calculation of the best basis according to proposition 4.9 is the same as before: $O(N \log_2 N)$. Thus, a best basis of discrete cosines of W_0^0 is selected with $O(N (\log_2 N)^2)$ operations.

Session 5: Wavelet compression and non linear approximation

1.- Introduction

Fourier compression processes can be schematically described in 3 steps:

$$\text{Image} \longrightarrow \begin{matrix}\text{Basis}\\ \text{coefficients}\end{matrix} \longrightarrow \text{Quantization} \longrightarrow \text{Encoding}$$

• The main feature of wavelets for image compression applications is:

> The largest wavelet coefficients are located precisely where the signal has singularities or sharp transitions (e.g. edges).

Thus, at least formally, quantization to 0 of many small wavelet coefficients should keep the basic geometry of the image.

The purpose of the talk is to justify this fact mathematically, by studying the decay of the approximation error

$$E_M(f) = \|f - \tilde{f}\|_2^2,$$

when only M significant coefficients are retained. Using Harmonic Analysis, we will describe the classes of signals for which the approximation errors decay like $E_M \approx M^{-\alpha}$. In practice, this can be translated into average rates of bits/pixel, once the distortion error has been prescribed.

Example: For images $f \in BV$, the error decays like M^{-1}, and good quality images may be obtained with 0.2 bits/pixel.

Wavelet compression vs JPEG:

JPEG'2000 is the best known wavelet-based image compression format. In addition to Discrete Wavelet Transforms, it incorporates new and more sophisticated entropy encoding schemes (embedding coding, zero-tree algorithms,...).

Some of the claimed **advantages** with respect to standard JPEG are:

• Superior compression performance at high bit rates, with average savings of roughly 20%.

• Much better performance at low-bit rates: elimination of blocky artifacts of DCT; good quality images at 0.2 bits/pixel and decent at 0.05 bits/pixel.

• *Scalability:* possibility of progressive encoding (and transmission) by successively refining the resolution accuracy (embedded codings).

• Possibility of performing lossless compression, by using integer-valued biorthogonal wavelets, which avoid rounding and hence quantization noise.

Disadvantages

• At the moment, it is not yet widely supported in web browsers.

• It is not completely license free: there exist legal issues with patents...

2.- An abstract setting for non-linear approximation

Let $\big(\mathcal{S}, \|\cdot\|\big)$ be a Banach space with a basis $\{e_j\}$.

Note: Typically, vectors in $\mathcal{S}$ model a large class of signals, and $\|\cdot\|_{\mathcal{S}}$ quantifies their magnitude or "energy". We may set $\mathcal{S} = L^2$, but models based on other Banach spaces such as L^p can be allowed.

DEF:

• The error of linear approximation with M basis coefficients is:

$$E_M^{\text{lin}}(f) = \inf\big\{\|f - \sum_{j\le M} c_j e_j\|_{\mathcal{S}} \ : \ c_j \in \mathbb{C}\big\}.$$

• The error of non-linear approximation with M basis coefficients is:

$$E_M^{\text{N-L}}(f) = \inf\big\{\|f - \sum_{\lambda\in\Lambda} c_\lambda e_\lambda\|_{\mathcal{S}} \ : \ \ \text{Card}\,\Lambda \le M\big\}.$$

Three main questions in Approximation Theory are:

(i) Decide, depending on applications, what norms $\|\cdot\|_{\mathcal{S}}$ and bases $\{e_j\}$ should be used to approximate signals in $\mathcal{S}$.

(ii) Find computable algorithms leading to minimizers of $E_M(f)$.

(iii) Identify the approximation classes

$$\mathcal{A}^\alpha \ = \ \big\{f \in \mathcal{S} \ : \ E_M(f) \le C\, M^{-\alpha}\big\}, \quad \text{for } \alpha > 0.$$

We will consider these questions for Fourier and wavelet bases.

3.- The Hilbert space setting

When $\mathcal{S} = L^2$ and $\{e_j\}$ an ONB, the minimizers are explicitly given by:

(i) Orthogonal projection: $\mathbb{P}_M(f) = \sum_{1 \le j \le M} c[j] e_j$ with $c[j] = \langle f, e_j \rangle$.

(ii) Greedy algorithm: $\mathbb{G}_M(f) = \sum_{1 \le k \le M} c^*[k] e_{j_k}$ with $c^*[k] = \langle f, e_{j_k} \rangle$ and

$$\big|c^*[1]\big| \;\ge\; \big|c^*[2]\big| \;\ge\; \big|c^*[3]\big| \;\ge\; \dots$$

The errors of approximation are explicitly given by

$$E_M^{\text{lin}}(f) \;=\; \Big(\sum_{k>M} |c[k]|^2\Big)^{\frac{1}{2}} \quad \text{and} \quad E_M^{\text{n-l}}(f) \;=\; \Big(\sum_{k>M} |c^*[k]|^2\Big)^{\frac{1}{2}}.$$

Thus signals will have a prescribed error decay, say $E_M(f) \le M^{-\alpha}$, at least when the coefficients $c[k]$ or $c^*[k]$ decay like $c\, k^{-\alpha-\frac{1}{2}}$. This fact almost characterizes the classes $\mathcal{A}^\alpha$...

Remark: When $\mathcal{S}$ is only a Banach space, finding minimizers is a more subtle question, of current theoretical interest to mathematicians in approximation theory. In the case $\mathcal{S} = L^p$ and for Fourier and wavelet bases, the approximations $\mathbb{P}_M(f)$ and $\mathbb{G}_M(f)$ are "almost minimizers", that is

$$\|f - \mathbb{P}_M(f)\|_{\mathcal{S}} \;\le\; c\, E_M^{\text{lin}}(f) \quad \text{and} \quad \|f - \mathbb{G}_M(f)\|_{\mathcal{S}} \;\le\; c\, E_M^{\text{n-l}}(f)$$

for some constant $c > 0$.

4.- Linear approximation in Hilbert spaces

Set

$$\mathfrak{s}^\alpha \;=\; \Big\{\mathfrak{c}=(c[k]) \;:\; \|\mathfrak{c}\|^2_{\mathfrak{s}^\alpha} := \sum_k |k^\alpha c[k]|^2 < \infty\Big\}.$$

Observe that sequences decaying like $k^{-\alpha-\frac{1}{2}-\epsilon}$ belong to this class.

Proposition.

- If $f \in L^2$ and $c[k] \in \mathfrak{s}^\alpha$, then $E_M(f) \lesssim \|\mathfrak{c}\|_{\mathfrak{s}^\alpha}\, M^{-\alpha}$.
- Conversely, if $E_M(f) \le cM^{-\alpha}$, then $c[k] \in \mathfrak{s}^{\alpha-\varepsilon}$, for all $\varepsilon > 0$.
- Moreover, the following characterization holds:

$$c[k] \in \mathfrak{s}^\alpha \quad\Longleftrightarrow\quad \sum_{M\ge 1} \Big[M^\alpha E_M(f)\Big]^2 \frac{1}{M} \;<\; \infty. \tag{0.1}$$

Remark: As suggested by the proposition, a natural class of approximation spaces in L^2 is defined by

$$\mathcal{A}^\alpha_2 \;=\; \Big\{f \in L^2 \;:\; \sum_{M\ge 1} \big[M^\alpha E_M(f)\big]^2 \frac{1}{M} \;<\; \infty\Big\}$$

Exercise: Show that $\mathcal{A}^{\alpha+\varepsilon} \subset \mathcal{A}^\alpha_2 \subset \mathcal{A}^\alpha$.

Proof of the proposition.

The proof is based in the following elementary lemma:

Lemma. *For any sequence* $s = \{s_k\}_{k=1}^{\infty}$ *we have*

(i) $\left(\sum_{k=M}^{\infty} |s_k|^2\right)^{\frac{1}{2}} \leq c\,\|s\|_{\mathfrak{s}^\alpha}\, M^{-\alpha}$, *whenever* $M \geq 1$.

(ii) For all $\varepsilon > 0$, *there exists* $c_\varepsilon > 0$ *so that*

$$\|s\|_{\mathfrak{s}^{\alpha-\varepsilon}} \leq c_\varepsilon \sup_{M\geq 1} M^\alpha \Big(\sum_{k=M}^{\infty} |s_k|^2\Big)^{\frac{1}{2}}.$$

(iii) There exist constants $c_1, c_2 > 0$ *so that*

$$c_1\,\|s\|_{\mathfrak{s}^\alpha} \leq \Big[\sum_{M=1}^{\infty} M^{2\alpha} \sum_{k>M} |s_k|^2\,\frac{1}{M}\Big]^{\frac{1}{2}} \leq c_2\,\|s\|_{\mathfrak{s}^\alpha}.$$

5.- Fourier Linear Approximation.

Consider the Fourier basis in $L^2[0,1]$, given by $e_k(t) = e^{2\pi ikt}$, $k \in \mathbb{Z}$. Then the minimizer with M-linear approximation is the partial sum

$$\tilde{f}_M(t) = \sum_{|k|\leq M} \hat{f}[k]e_k(t).$$

By the proposition, the approximation class is characterized by

$$\mathcal{A}_2^\alpha = \Big\{ f \in L^2 \ : \ \sum_{|k|\geq 1} \left|k^\alpha \hat{f}[k]\right|^2 < \infty \Big\}.$$

Can we recognize this space of signals?

Theorem *For every $\alpha > 0$, the approximation space $\mathcal{A}_2^\alpha$ is equal to the Sobolev space $H^\alpha[0,1]$.*

Conclusion: Sobolev norms measure global smoothness of functions. Audio signals have a reasonably high global smoothness, hence Fourier Linear Approximation works reasonably well in this case. However, images have many local singularities, which turn into low global smoothness. Thus Fourier Linear Approximation gives poor decays for the error.

Exercise: Let R be a rectangle in $\mathbb{R}^2$. Show that $f = \chi_R \in H^\alpha(\mathbb{R}^2)$ if and only if $\alpha < \frac{1}{2}$. Conclude that the Fourier approximation error with frequencies $\leq M$ cannot decay faster than $M^{-1/4}$.

6.- Wavelet Linear Approximation.

Consider a wavelet basis in $L^2([0,1]^2)$, given by

$$\left\{\varphi_{j_0,\mathbf{k}},\ \psi^\ell_{j,\mathbf{k}}\ :\ 0 \le |\mathbf{k}| < 2^j, \quad j \ge j_0,\ \ell = 1,2,3\right\}$$

(or a variant version modified at the boundaries). Set

$$\mathbb{P}_J(f) = \sum_{|\mathbf{k}|<2^{j_0}} \langle f, \varphi_{j_0,\mathbf{k}}\rangle \varphi_{j_0,\mathbf{k}} + \sum_{\ell=1}^{3}\sum_{j=j_0}^{J}\sum_{|\mathbf{k}|<2^j} \langle f, \psi^\ell_{j,\mathbf{k}}\rangle \psi^\ell_{j,\mathbf{k}}$$

which corresponds to linear approximation at scales of sidelength $\ge 2^{-J}$. Strictly speaking, this is the minimizer for linear approximation with $M = 2^{2J}$ basis elements.

Reasoning as in the proposition above, one can establish that

$$\begin{aligned}\mathcal{A}^\alpha_2 &:= \left\{f \in L^2\ :\ \sum_{J=j_0}^{\infty}\Big(2^{2J\alpha}\|f - \mathbb{P}_J(f)\|\Big)^2 < \infty\right\} \\ &= \left\{f \in L^2\ :\ \sum_{\ell=1}^{3}\sum_{j=j_0}^{\infty}\sum_{|k|<2^j} \left|2^{2\alpha j}\,\langle f, \psi^\ell_{j,\mathbf{k}}\rangle\right|^2 < \infty\right\}.\end{aligned}$$

Can we recognize this space of signals?

Theorem. *For every $\alpha > 0$, the wavelet approximation space $\mathcal{A}^\alpha_2$ is equal to the Sobolev space $H^{2\alpha}([0,1]^2)$.*

Conclusion: Wavelet Linear Approximation performs similarly to Fourier Linear Approximation. For images, singularities along edges force global Sobolev smoothness of order $H^{\frac{1}{2}}$, and hence poor decays for the approximation error: for $f = \chi_R$

$$E_J(f) \approx 2^{-J/2} \,=\, M^{-\frac{1}{4}}.$$

7.- Abstract Non-Linear Approximation in Hilbert spaces.

For non-linear approximation, we must study the decay of the decreasingly arranged coefficients $c^*[k]$. Recall that an error decay $E_M(f) \leq M^{-\alpha}$ can be obtained if $c^*[k] \approx k^{-\alpha-\frac{1}{2}}$.

Set

$$\ell^\tau \;=\; \Big\{\mathfrak{c} = (c[k]) \;:\; \sum_k \big|\, c[k]\,\big|^\tau < \infty\Big\}.$$

Observe that $\{k^{-\alpha-\frac{1}{2}}\}_{k=1}^{\infty}$ belongs to this class for all $\tau > (\alpha+\frac{1}{2})^{-1}$.

Proposition. Let $\alpha > 0$ and $\frac{1}{\tau} = \alpha + \frac{1}{2}$.

• If $f \in L^2$ and $c[k] \in \ell^\tau$, then $E_M^{\text{n-l}}(f) \lesssim \|\mathfrak{c}\|_{\ell^\tau}\, M^{-\alpha}$.

• Conversely, if $E_M^{\text{n-l}}(f) \leq cM^{-\alpha}$, then $c[k] \in \ell^{\tau-\varepsilon}$, for all $\varepsilon > 0$.

• Moreover, the following characterization holds:

$$c[k] \in \ell^\tau \quad\Longleftrightarrow\quad \sum_{M\geq 1} \Big[M^\alpha E_M^{\text{n-l}}(f)\Big]^\tau \,\frac{1}{M} \;<\; \infty. \tag{0.2}$$

Note: For non-linear approximation in L^2, the natural approximation classes are defined by

$$\mathcal{A}_\tau^\alpha \;=\; \Big\{f \in L^2 \;:\; \sum_{M\geq 1} \big[M^\alpha E_M^{\text{n-l}}(f)\big]^\tau \,\frac{1}{M} \;<\; \infty\Big\}$$

By the proposition these are characterized by $c[k] \in \ell^\tau$.

8.- Wavelet Non-Linear Approximation.

Consider again a wavelet basis in $L^2([0,1]^2)$:

$$\left\{\varphi_{j_0,\mathbf{k}},\ \psi^\ell_{j,\mathbf{k}}\ :\ 0 \le |\mathbf{k}| < 2^j,\quad j \ge j_0,\ \ell = 1,2,3\right\}$$

The minimizer of $E^{\text{n-l}}_M(f)$ with M-wavelet coefficients is

$$\tilde{f}_M = \mathbb{P}_{j_0}(f) + \sum_{j=1}^{M} \langle f, \psi_{\lambda_j}\rangle \psi_{\lambda_j}$$

where the wavelet coefficients have been arranged by decreasing sizes

$$\left|\langle f, \psi_{\lambda_1}\rangle\right| \ge \left|\langle f, \psi_{\lambda_2}\rangle\right| \ge \left|\langle f, \psi_{\lambda_3}\rangle\right| \ge \dots$$

The last proposition shows that, for $\frac{1}{\tau} = \alpha + \frac{1}{2}$

$$\mathcal{A}^\alpha_\tau = \left\{ f \in L^2\ :\ \sum_{\ell=1}^{3}\sum_{j=j_0}^{\infty}\sum_{|k|<2^j} |\langle f, \psi^\ell_{j,\mathbf{k}}\rangle|^\tau < \infty \right\}.$$

Can we recognize this space of signals?

Theorem. *Let $\alpha > 0$ and $\tau = (\alpha + \frac{1}{2})^{-1}$. Then the wavelet approximation space $\mathcal{A}^\alpha_\tau$ is equal to the Besov space $B^{2\alpha}_\tau([0,1]^2)$.*

Remark: Besov spaces measure global smoothness, except perhaps for local singularities at isolated points or along rectifiable curves. One can show experimentally that most images belong to $B^{2\alpha}_\tau$ with $\alpha \in [0.2, 0.6]$.

For example, $f = \chi_R$ with R a rectangle of $\mathbb{R}^d$ belongs to $B^{d\alpha}_\tau$ (with $\frac{1}{\tau} = \alpha + \frac{1}{2}$) whenever $\alpha < \frac{1}{2(d-1)}$. When $d = 1$, this improves dramatically the $M^{-1/2}$ decay obtained with linear wavelet approximation. When $d = 2$ one can show that

$$E^{\text{n-l}}_M(f) \approx M^{-\frac{1}{2}},$$

which also improves considerably the $M^{-\frac{1}{4}}$ bound of linear approximation.

9.- Functions of bounded variation.

DEF:

$$BV(\mathbb{R}^2) = \{ f \in L^2 \ : \ \nabla f \text{ is a Radon measure } \}.$$

• This space is considered a good model for images ('$u + v$' models). It may be equivalently defined as functions with "finite graph surfaces", and in particular contains characteristic functions of sets with finite perimeter. It can be shown that

$$B_1^1(\mathbb{R}^2) \subset W^{1,1}(\mathbb{R}^2) \subset BV(\mathbb{R}^2).$$

Theorem (Cohen).

• *If f has wavelet coefficients in ℓ^1, then $f \in BV(\mathbb{R}^2)$.*

• *If $f \in BV(\mathbb{R}^2)$, then the wavelet coefficients belong to the Lorentz space $\ell^{1,\infty}$, meaning that*

$$\left|\langle f, \psi_{\lambda_k} \rangle\right| \lesssim k^{-1}, \quad k = 1, 2, \ldots,$$

where $\left|\langle f, \psi_{\lambda_1} \rangle\right| \geq \left|\langle f, \psi_{\lambda_2} \rangle\right| \geq \ldots$

• *All functions $f \in BV(\mathbb{R}^2)$ satisfy $E_M^{n\text{-}l}(f) \lesssim M^{-\frac{1}{2}}$.*

10.- Final remarks: the L^p theory.

Wavelet non-linear approximation for $\mathcal{S} = L^p$, $1 < p < \infty$, was first studied by DeVore-Jawerth-Popov in 1992, and later by Lucier, Temlyakov, Cohen and others... Sometimes this leads to interesting Harmonic Analysis problems. E.g., higher dimensional models based on $BV(\mathbb{R}^d)$ require $p = d/(d-1)$. Some of the results can be summarized as follows:

• The greedy approximant $\mathbb{G}_M(f)$ is an almost minimizer for $E_M^{\text{n-l}}(f)$ (provided we normalize $\|\psi_\lambda\|_{L^p} = 1$).

• $\mathcal{A}_\tau^\alpha(L^p(\mathbb{R}^d)) = B_\tau^{d\alpha}(\mathbb{R}^d)$, when $\frac{1}{\tau} = \alpha + \frac{1}{p}$.

• For $f \in BV(\mathbb{R}^d)$, then $E_M^{\text{n-l}}(f) \lesssim M^{-1/d} \, |f|_{BV}$.

• $\mathcal{A}_\tau^\alpha(W^{p,s}(\mathbb{R}^d)) = \mathcal{A}_\tau^\alpha(B_p^s(\mathbb{R}^d)) = B_\tau^{d(\alpha+s)}(\mathbb{R}^d)$, when $\frac{1}{\tau} = \alpha + \frac{1}{p}$.

• The cases L^1 or L^∞ remain open.

• For other types of bases (such as the hyperbolic Haar basis) or other Banach spaces (such as $L^p(LogL)$)), the greedy approximant $\mathbb{G}_M(f)$ minimizes $E_M^{\text{n-l}}(f)$ with an $(\log M)^\gamma$-loss, and thus leads only to partial characterizations of the approximation spaces.