Wavelets

Marialuce Graziadei

1. A brief summary

2. Vanishing moments

3. 2D-wavelets

4. Compression

5. De-noising

1. A brief summary

- $\phi(t)$: scaling function.
  For ϕ the 2-scale relation hold

$$\phi(t) = \sum_{k=-\infty}^{\infty} p_k \phi(2t-k) \qquad (t \in \mathbb{R})$$

- $\psi(t)$: mother wavelet. For ψ the 2-scale relation hold

$$\psi(t) = \sum_{k=-\infty}^{\infty} q_k \phi(2t-k) \qquad (t \in \mathbb{R})$$

- The decomposition for ϕ reads

$$\phi(2t-k) = \sum_{m=-\infty}^{\infty} h_{2m-k}\phi(t-m) + g_{2m-k}\psi(t-m) \qquad (t \in \mathbb{R})$$

2. Vanishing moments

Moments of a mother-wavelet

$$m_\mu = \int_{-\infty}^{\infty} t^\mu \psi(t)\, dt.$$

Goal: to show that, if the mother-wavelet has successive moments equal to zero (for a fixed b) the wavelet coefficients decrease 'quickly' when a decreases.

Assumptions:

- the function $f(t)$ is $(k-1)$ times continuously differentiable;
- $f^{(k)}(t)$ has jumps at most in a finite number of points.

Then $f(t)$ can be expanded as

$$f(t+b) = f(b) + f'(b)t + ... + \frac{f^{(k-1)}(b)}{(k-1)!} t^{k-1} + t^k r(t),$$

where $r(t)$ is piecewise continuous and bounded, with at most a finite number of jumps.

Wavelet coefficients

$$
\begin{aligned}
< f, \psi_{a,b} > &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi((t-b)/a)\,dt = \sqrt{a}\int_{-\infty}^{\infty} f(b+at)\psi(t)\,dt \\
&= \sqrt{a}\int_{-\infty}^{\infty} (f(b) + f'(b)(at)) + \ldots + \frac{f^{(k-1)}(b)}{(k-1)!}(at)^{k-1})\psi(t)\,dt \\
&+ a^{k+\frac{1}{2}} \int_{-\infty}^{\infty} t^k r(at)\psi(t)\,dt
\end{aligned}
$$

If the first $(k-1)$ moments of the mother-wavelet are equal to zero

$$
< f, \psi_{a,b} > = a^{k+\frac{1}{2}} \int_{-\infty}^{\infty} t^k r(at)\psi(t)\,dt
$$

This means that

$$
< f, \psi_{a,b} >= O(a^{k+\frac{1}{2}}) \qquad\qquad (a \to 0)
$$

Example

The function

$$f(t) = \sin(\pi |t - \frac{1}{2}|),$$

is not-differentiable in $t = \frac{1}{2}$.

The following table shows the wavelet coefficients, computed for different values of a and for $b = 0.5$ and $b = 0.6$, using the Mexican hat. The convergence factors ($\log_2$) are also reported.

a	$b = 0.5$	$k + \frac{1}{2}$	$b = 0.6$	$k + \frac{1}{2}$
0.128	−6.6277	1.329	−4.0958	2.7389
0.064	−2.6678	1.4559	−0.6136	4.3382
0.032	−0.9724	1.4888	0.0303	2.1796
0.016	−0.3465	1.4987	0.0067	2.4986
0.008	−0.1226	1.5050	0.0012	2.4997
0.004	−0.0432		0.0002	

For $b = 0.5$ the wavelet coefficients go to zero more slowly than for $b = 0.6$!!!

3. 2D-wavelets

Notation:

$$f(x-k, y-l) = f_{k,l}(x, l)$$

are the translations of f.

$$f(x, y) \to f(2^n x, 2^n y)$$

are the dilatations of f.

The definition of a 2D Multi-Resolution Analysis (MRA) is similar to a 1D-MRA.

If a sequence of subspaces (V_n) satisfies the following properties

a) $V_n \subset V_{n+1} \quad (n \in \mathbb{Z})$,

b) $\overline{\bigcup_{n\in\mathbf{Z}} V_n} = L^2(\mathbb{R}^2)$,

c) $\bigcap_{n\in\mathbf{Z}} V_n = \{0\}$,

d) $f(x, y) \in V_n \Leftrightarrow f(2x, 2y) \in V_{n+1}$,

e) $f(x, y) \in V_0 \Rightarrow f_{k,l}(x, y) \in V_0$.

then it is called a MRA of $L^2(\mathbb{R}^2)$.

ϕ is a scaling function for the MRA

- it is continuous, with a compact support in $\mathbb{R}^2$, with possible jumps on the boundary of the support;
- the integer translations of ϕ, $\phi_{k,l}$, form a Riesz-basis for V_0.

Scaling functions

Sufficient conditions for a compactly supported function ϕ to be a scaling function for an MRA

1. There exists a sequence of numbers $p_{k,l}$ (only a finite number differs from zero) such that

$$\phi(x, y) = \sum_{k,l} p_{k,l}\phi(2x - k, 2y - l) \qquad ((x, y) \in \mathbb{R}^2).$$

(2-scale relation).

2. Introduce the *2D-autocorrelation function* of ϕ as

$$\rho(k, l) := \iint_{-\infty}^{\infty} \phi(x, y)\phi(x + k, y + l)\, dx\, dy$$

and the 2D-Riesz-function as

$$R_\phi(z_1, z_2) = \sum_{k,l} \rho(k, l)\, z_1^k\, z_2^k \qquad ((z_1, z_2) \in \mathbb{C}^2)$$

The Riesz function $R_\phi(z_1, z_2)$ is positive for $|z_1| = |z_2| = 1$.

3. The translation of the function ϕ are such that

$$\sum_k \phi(t - k) \equiv 1. \qquad \text{(Partition of the unity).}$$

If ϕ has these properties, the following hold

- The reference space V_0 consists of functions $f(x, y)$ that can be expressed as

$$f(x, y) = \sum_{k,l} a_{k,l}\phi_{k,l}(x, y) \qquad ((x, y) \in \mathbb{R}^2)$$

- Whatever is n, the space V_n consists of functions f such that

$$f(x, y) = \sum_{k=-\infty}^{\infty} a_{k,l}\phi_{k,l}(2^n x, 2^n y) \qquad ((x, y) \in \mathbb{R}^2)$$

As in 1D-MRA, the goal is to build the detail space (W_n) such that

$$V_{n+1} = V_n \oplus W_n$$

The space W_n is built from a mother-wavelet ψ.

The functions $\psi(2^n x - k, 2^n y - l)$ form a Riesz-basis for W_n.

In 2-D more than one wavelet is necessary to span W_0.

Example: 2-D Haar wavelets

$\phi_1(x)$: 1-D Haar scaling function;

$\psi_1(x)$: 1-D Haar wavelet.

The 2-D Haar scaling function is defined from the 1-D Haar scaling function as

$\phi(x, y)$=$\phi_1(x)\phi_1(y)$

and, if one want to fill V_0 to obtain V_1:

$$\psi^{(1)}(x, y) = \psi_1(x)\phi_1(y) = \phi(2x, y) - \phi(2x - 1, y),$$

$$\psi^{(2)}(x, y) = \phi_1(x)\psi_1(y) = \phi(x, 2y) - \phi(x, 2y - 1),$$

$$\begin{aligned}\psi^{(3)}(x, y) =&\psi_1(x)\psi_1(y) = \phi(2x, 2y) - \phi(2x - 1, 2y)\\ &+ \phi(2x - 1, 2y - 1) - \phi(2x, 2y - 1).\end{aligned}$$

These three functions belong to V_1 and are orthogonal to V_0, so they belong to W_0.

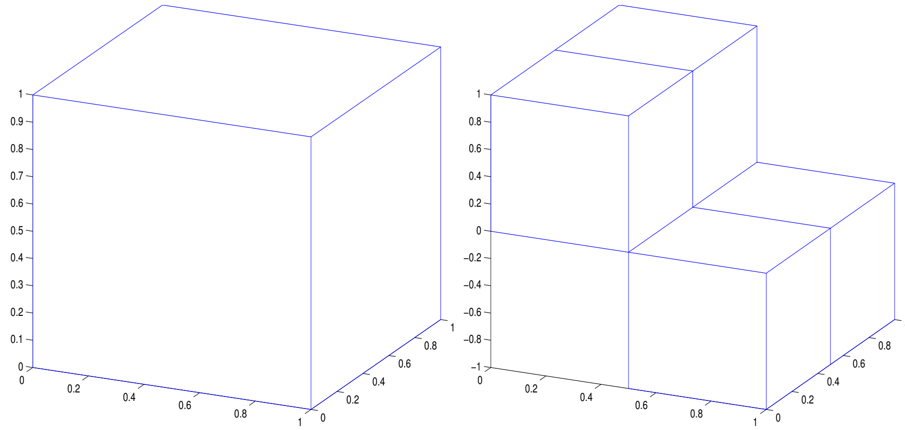


Figure 1: *2-D Haar scaling function* ϕ *and Haar wavelet* $\psi^{(1)}$

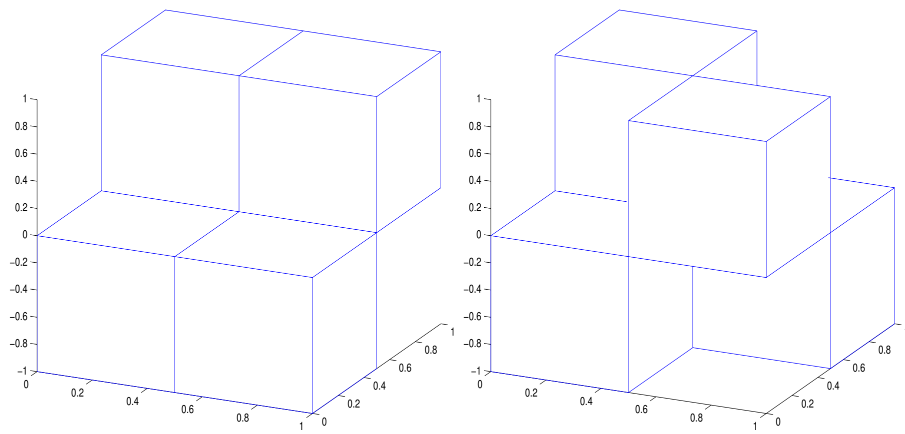


Figure 2: *Haar wavelet* $\psi^{(2)} and Haar wavelet \psi^{(3)}$

The function $\psi^{(1)}$, $\psi^{(2)}$ and $\psi^{(3)}$ must satisfy the following

- $$W_0 = W_0^1 \oplus W_0^2 \oplus W_0^3$$
- the translation of $\psi^{(j)}$ are a Riesz basis for $W_0^{(j)}$

Then

$$\phi(x, y) = \sum_{k,l} p_{k,l}\phi(2x - k, 2y - l) \qquad ((x, y) \in \mathbb{R}^2).$$

$$\psi^{(j)}(x, y) = \sum_{k,l} q_{k,l}^{(j)}\phi(2x - k, 2y - l) \qquad ((x, y) \in \mathbb{R}^2).$$

These relations, taking into account that

$$V_1 = V_0 \oplus W_0^1 \oplus W_0^2 \oplus W_0^3$$

allows to find the coefficients h_k^j, with $(j = 0, 1, 2, 3)$ such that

$$\phi_{m,n}(2x, 2y) = \sum_{k,l} h_{2k-m,2l-n}^0 \phi_{k,l}(x, y) + \sum_{j=1}^{3} \sum_{k,l} h_{2k-m,2l-n}^{(j)} \psi_{k,l}^{(j)}(x, y).$$

Furthermore, for all functions $f \in L^2(\mathbb{R}^2)$ there are three series $(a^{(j)}_{r,k,l})$, $(j = 1, 2, 3)$ such that

$$f(x, y) = \sum_{m=1}^{3} \sum_{r=-\infty}^{\infty} \sum_{k,l} a^{(m)}_{r,k,l} \psi^{(m)}(2^r x - k, 2^r y - l).$$

Filterbanks

As in 1-D, one can decompose $f \in V_0$ using a filterbank. f can be written as

$$f(x, y) = \sum_{k,l} a^0_{k,l} \phi_{k,l}(x, y).$$

But $f = f_{-1} + g_{-1} + g_{-2} + g_{-3}$ with $f_{-1} \in V_{-1}$ and $g_{-1}, g_{-2}, g_{-3} \in W_{-1}$.

These functions can be represented as

$$f_{-1}(x, y) = \sum_{k,l} a^{-1}_{k,l} \phi_{k,l}(x/2, y/2);$$

$$g_{-1}(x, y) = \sum_{k,l} d_{-1,k,l} \psi^{(1)}_{k,l}(x/2, y/2);$$

$$g_{-2}(x, y) = \sum_{k,l} d_{-2,k,l} \psi^{(2)}_{k,l}(x/2, y/2);$$

$$g_{-3}(x, y) = \sum_{k,l} d_{-3,k,l} \psi^{(3)}_{k,l}(x/2, y/2);$$

Then

$$a_{k,l}^{-1} = \sum_{u,v} h^0_{2k-u,2l-v} a^0_{u,v},$$

$$d_{-m,k,l} = \sum_{u,v} h^m_{2k-u,2l-v} a^0_{u,v} \qquad (m = 1, 2, 3).$$

The reconstruction coefficients can be computed in the same way as in the 1-D case.

Filterbanks

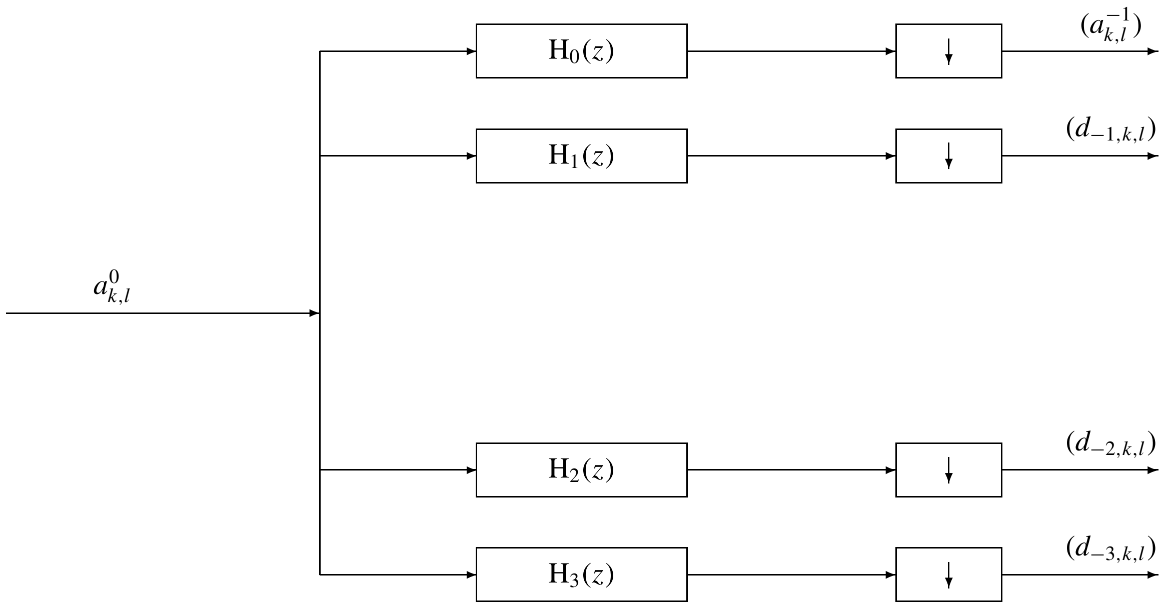


Figure 3: Decomposition

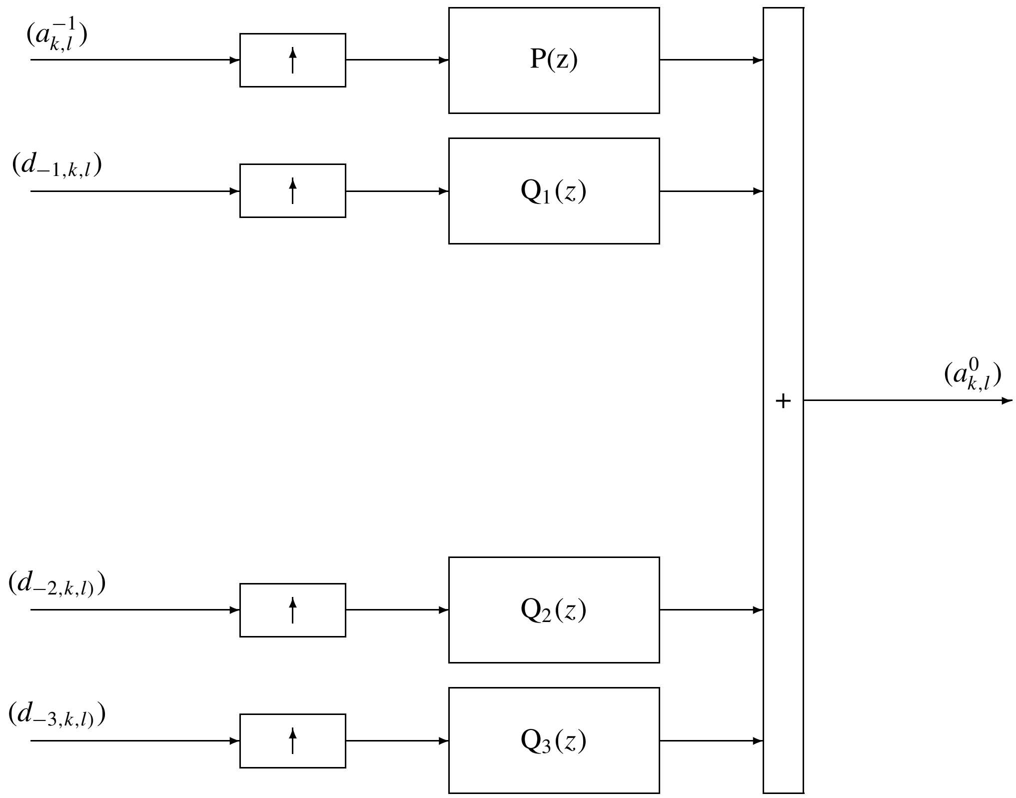


Figure 4: Reconstruction

Analysis of 2-D images

Discrete images: arrays f of M rows and N columns

$$\mathbf{f} = \begin{pmatrix} f_{1,M} & f_{2,M} & \cdots & f_{N,M} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ f_{1,2} & f_{2,2} & \cdots & f_{N,2} \\ f_{1,1} & f_{2,1} & \cdots & f_{N,1} \end{pmatrix}$$

The 2-D wavelet transform of such an image can be performed in two steps.

Step 1. Perform a 1-D wavelet transform on each row of $\mathbf{f}$, thereby producing a new image.

Step 2. Perform a 1-D wavelet transform on each column of of the matrix obtained with the **Step 1.**

A 1-level wavelet transform can be therefore symbolized as follows:

$$\mathbf{f} \to \begin{pmatrix} \mathbf{a}^1 & | & \mathbf{h}^1 \\ - & & - \\ \mathbf{v}^1 & | & \mathbf{d}^1 \end{pmatrix}$$

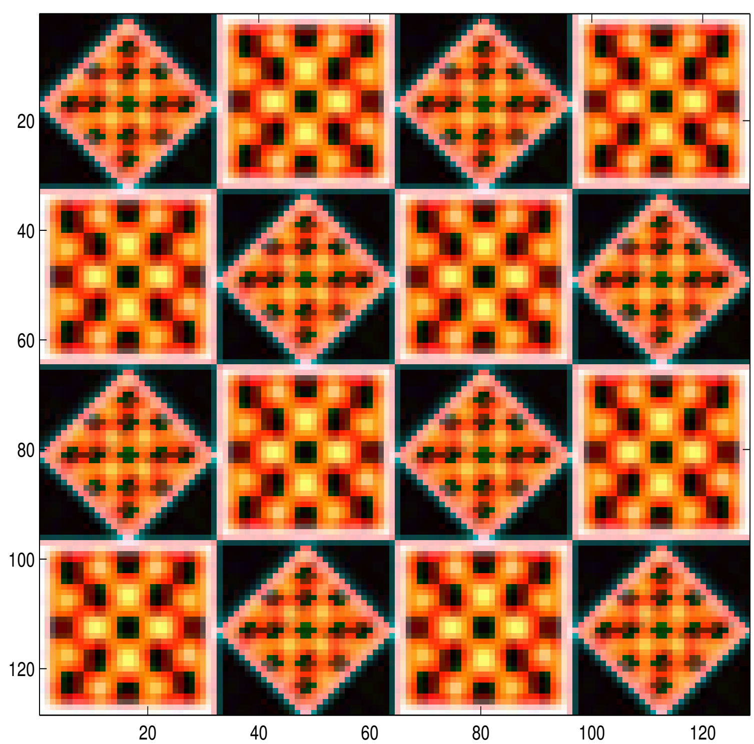
where the sub-images $\mathbf{h}^1$, $\mathbf{d}^1$, $\mathbf{a}^1$ and $\mathbf{v}^1$ each have $M/2$ rows and $N/2$ columns.

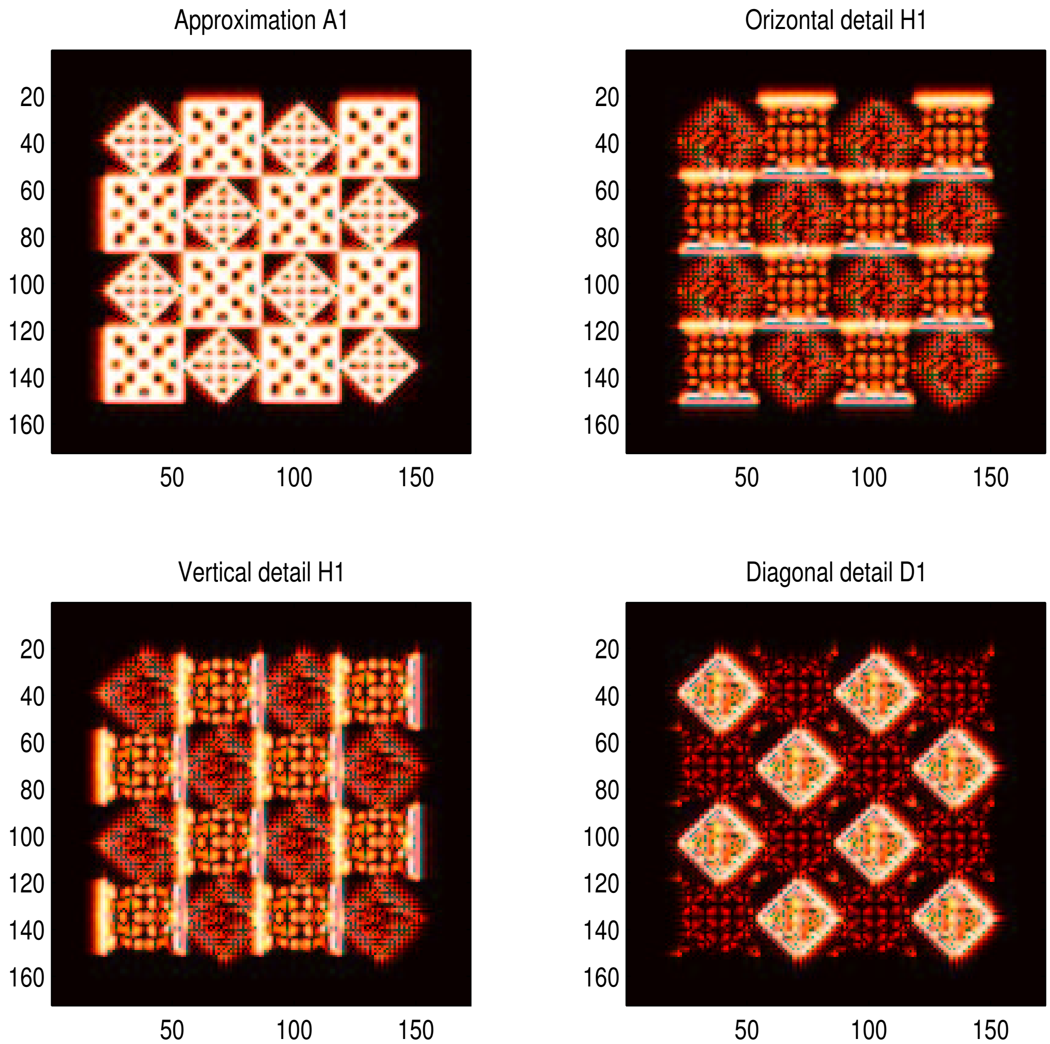
Sub-image $\mathbf{a}^1$: it is created computing trends along rows of $\mathbf{f}$, following by computing trends along columns, so it is an an averaged, lower resolution version of $\mathbf{f}$.

Sub-image $\mathbf{h}^1$: it is created computing trends along rows of $\mathbf{f}$, following by computing fluctuation along columns. Consequently it detects horizontal edges.

Sub-image $\mathbf{v}^1$: it is created like $\mathbf{h}^1$, but inverting the role of rows and columns: it detects vertical edges.

Sub-image $\mathbf{d}^1$: it tends to emphasize diagonal features, because it is created from fluctuation along both rows and columns.




Approximation A1
Orizontal detail H1
Vertical detail H1
Diagonal detail D1
20
40
60
80
100
120
140
160
50
100
150

4. De-noising

- The transmission of a signal over some distance often implies the contamination of the signal itself by noise.
- The term noise refers to any undesirable change that has altered the value of the original signal.
- The simplest model for acquisition of noise by a signal is the *additive noise* , which has the form

  $$\mathbf{f} = \mathbf{s} + \mathbf{n},$$

  where **f** is the contaminated signal, **s** is the original signal and **n** is the noise.

The most common types of noise are the following:

1.Random noise. The noise signal is highly oscillatory above and below an average mean value.

2.Pop noise. The noise is perceived as randomly occurring, isolated 'pops'. As a model for this type of noise we add a few non-zero values to the original signal at isolated locations.

3.Localized random noise. It appears as in type 1, but only over a (some) short segment(s) of the signal. This can occur when there is a short-lived disturbance in the environment during transmission of the signal.

De-noising procedure principles

1. Decompose. Choose a wavelet and a level N. Compute the wavelet decomposition of the signal at level N.

2. Threshold detail coefficients. For each level from 1 to N, select a threshold and apply soft or hard thresholding to the detail coefficients.

3. Reconstruct.

How to choose a threshold?

The most frequently encountered noise in transmission is the gaussian noise. It can be characterised by a the **mean** μ and by the **standard deviation** σ.

- Assume that $\mu = 0$.
- The gaussian nature of the noise is preserved during the transformation $\Rightarrow$ the wavelet coefficients are distributed according to a Gaussian curve having $\mu = 0$ and standard deviation σ.
- From the theory, if one chooses

  $$T = 4.5\sigma,$$

  the 99.99% of the wavelet coefficients will be eliminated.

Usually the finest detail consist almost entirely of noise. Its standard deviation can be assumed as a good estimate for σ.

Soft or hard thresholding?

Let T denote the threshold, x the wavelet transform values and $H(x)$ the transform value after the thresholding.

Hard thresholding means

$$H(x) = \begin{cases} x & \text{if } |x| \geq T \\ 0 & \text{if } |x| \leq T \end{cases}$$

Soft thresholding means

$$H(x) = \begin{cases} sign(x)(|x| - T) & \text{if } |x| \geq T \\ 0 & \text{if } |x| \leq T \end{cases}$$

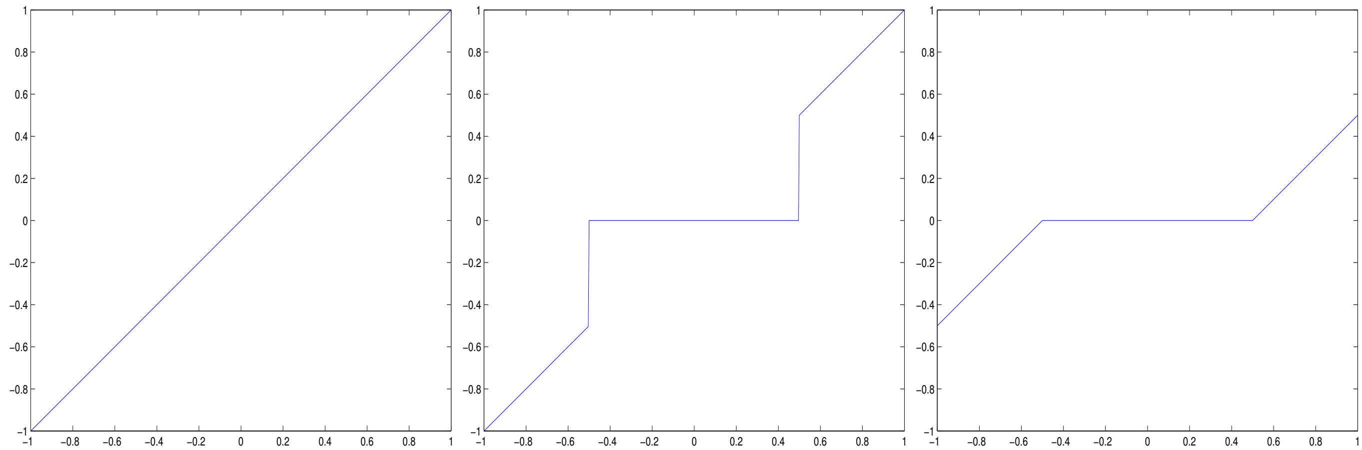


Figure 5: Original signal, hard thresholding and soft thresholding

1. Hard thresholding exaggerates small differences in transform values that have magnitude near the threshold value T

2. Soft thresholding has risk of oversmoothing

5. Compression

Compression means converting the signal(s) data into a new format that requires less bit to be transmitted.

Lossless compression

It is completely error free (ex: techniques that produce .zip files). The maximum compression ratio are $2 : 1$.

Lossy compression

It is used when inaccurancies can be accepted because quite imperceptible. The compression ratio vary from $10 : 1$ to $100 : 1$ when more complex techniques are used. Wavelets are applied in this field.

Compression procedure

- Perform wavelet transform of the signal up to level N;
- Set equal to zero all values of the wavelet coefficients which are insignificant, i.e. which are below some threshold value;
- Transmit only the significant, non-zero values of the transform obtained from **Step 2**;
- Reconstruct the signal.

How to choose a threshold?

Suppose that L_j are the transform coefficients. One can put them in decreasing order

$$L_1 \geq L_2 \geq L_3 \geq ... \geq L_M$$

The energy of a signal is

$$E_f = \sum_{j=1}^{M} L_j^2$$

Compute the cumulative energy profile

$$\frac{L_1^2}{E_f}, \frac{L_1^2 + L_2^2}{E_f}, ..., \frac{L_1^2 + L_2^2 + L_3^2 + ... + L_N^2}{E_f}, ..., 1.$$

The threshold T is chosen according with the amount of energy that we want to retain. If the term

$$\frac{L_1^2 + L_2^2 + L_3^2 + ... + L_N^2}{E_f}$$

is such that a sufficient amount of energy is kept, then all the coefficients smaller than L_N can be put equal to zero.