Wavelet transform, denoising, deconvolution

Laure Blanc-Féraud

DR CNRS
Laboratoire I3S
Morpheme group, Sophia Antipolis

CONTENT

- Wavelet transform:
  - Introduction
  - Definitions and properties
  - Multiresolution decomposition

- Approximation/denoising
  - Best approximation on a basis
  - Image denoising by wavelet transform, thresholding
  - Invariance (translation, rotation), dual tree wavelets.

- Deconvolution in wavelet domain
  - Deconvolution wavelet packets
  - Algorithms (threshoding, filtering)

Wavelet transform

Decomposition on an orthogonal basis

Continuous signal/image $f \in \mathrm{L}^2([0,1]^d)$.

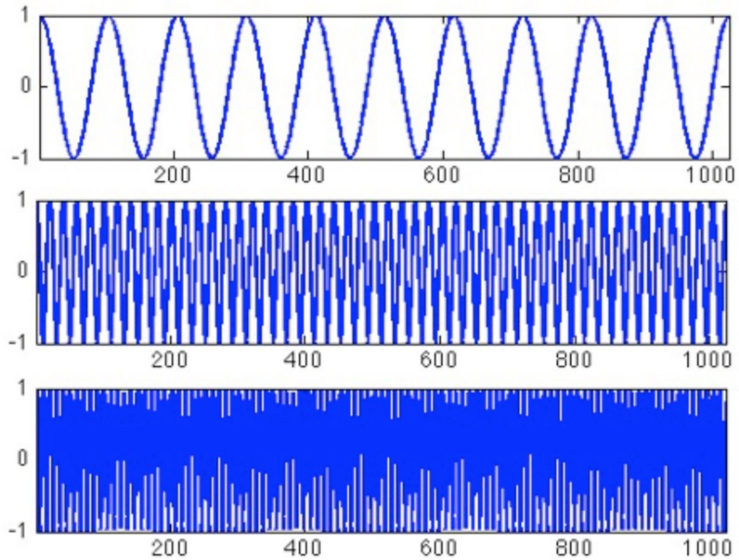
Orthogonal basis $\{\psi_m\}_m$ of $\mathrm{L}^2([0,1]^d)$ $\quad \langle f,\, g\rangle = \int f(x)\bar{g}(x)\mathrm{d}x$

Decomposition: $f = \sum_m \langle f,\, \psi_m\rangle \psi_m$

Energy conservation:

$$\|f\| = \int |f(x)|^2 \mathrm{d}x = \sum_m |\langle f,\, \psi_m\rangle|^2$$

ψ_m



Fourier: $\psi_m(x) = e_m(x) = e^{2i\pi m x}$

$$= \cos(2\pi m x) + i\sin(2\pi m x)$$

Frequency m.

Fourier $\rightarrow$ Wavelets ?

- The Fourier transform gives information on the frequencies present in the signal but we have no information on their location, because of the sum (continuous and discrete setting)

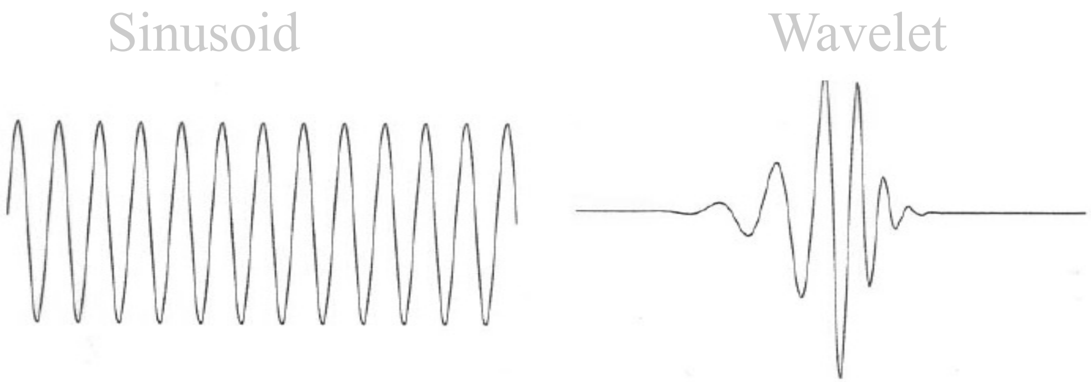
$$\hat{u}(\zeta) = \int_{R^d} u(x) \exp\left(-i\langle \zeta, x\rangle\right) dx$$

$$\hat{u}(k) = \sum_{i=0}^{N-1} u_n \exp\left(-\frac{2i\pi kn}{N}\right)$$

- Hence the development of time-frequency representations (Fourier with sliding window, Gabor, Wigner-Ville) and the work on wavelet transforms.

What is a wavelet?

- A function that “waves” above and below the x-axis with the following properties:
  - Varying frequency
  - Limited duration
  - Zero average value
- This is in contrast to sinusoids, used by FT, which have infinite duration and constant frequency.

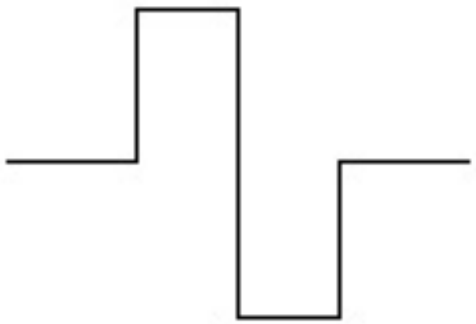


Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Types of Wavelets

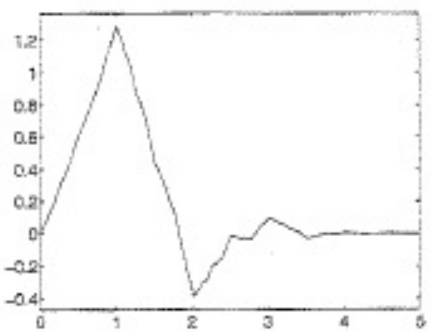
- There are many different wavelets, for example:

Haar



Morlet

Daubechies



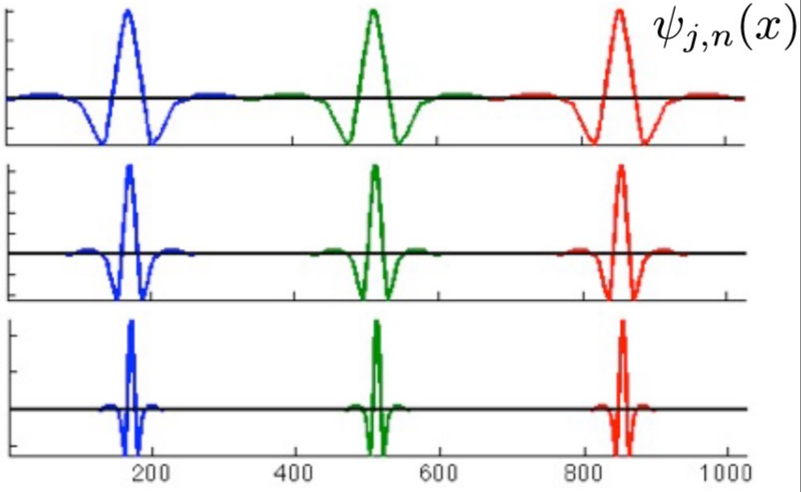
Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

1D-Wavelet basis

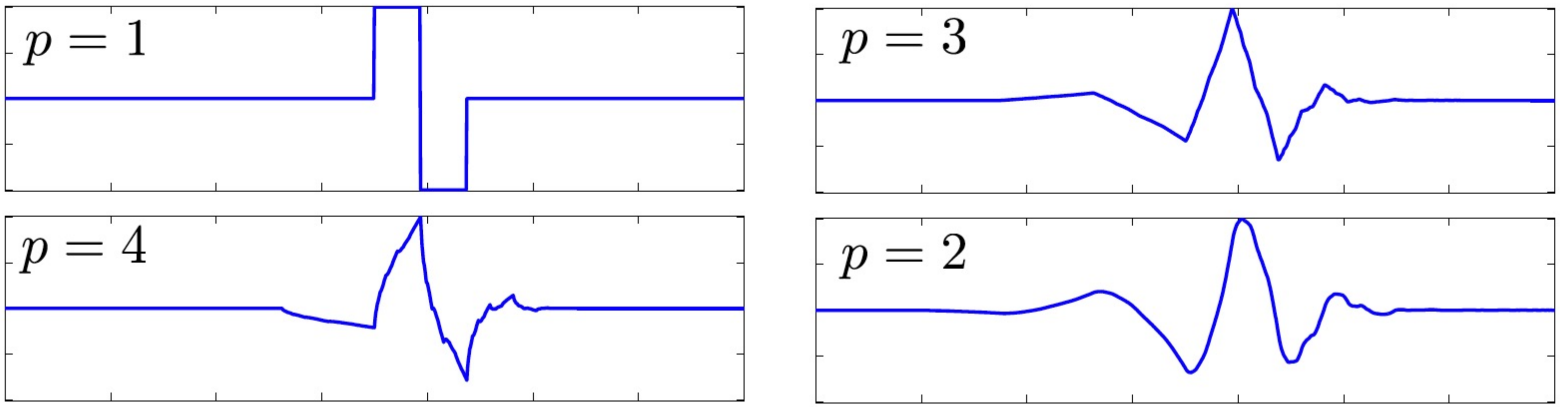
Wavelets:

$$\psi_{j,n}(x) = \frac{1}{2^{j/2}} \psi\left(\frac{x - 2^j n}{2^j}\right)$$

$\rightarrow$ Position n, scale 2^j, $m = (n, j)$.



$\rightarrow$ a large choice of "mother" wavelets.



Thanks to Gabriel Peyré https://mathematical-tours.github.io/slides/

Basis Functions Using Wavelets

- Like sin( ) and cos( ) functions in the Fourier Transform, wavelets can define a set of basis functions $\psi_k(t)$:

$$f(t) = \sum_k a_k \psi_k(t)$$

- Span of $\psi_k(t)$: vector space S containing all functions f(t) that can be represented by $\psi_k(t)$.

Continuous Wavelet Transform (CWT)

translation parameter
(measure of time)

scale parameter
(measure of frequency)

scale $=1/2^j$
(1/frequency)

Forward CWT:

$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t)\, \psi^*\left(\frac{t-\tau}{s}\right) dt$$

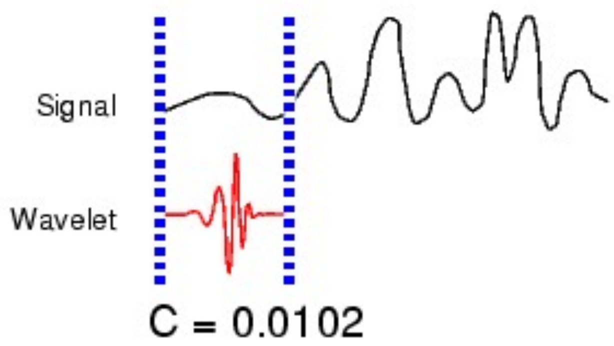
normalization constant

mother wavelet (i.e., window function)

Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Illustrating CWT

1. Take a wavelet and compare it to a section at the start of the original signal.
2. Calculate a number, C, that represents how closely correlated the wavelet is with this section of the signal. The higher C is, the more the similarity.

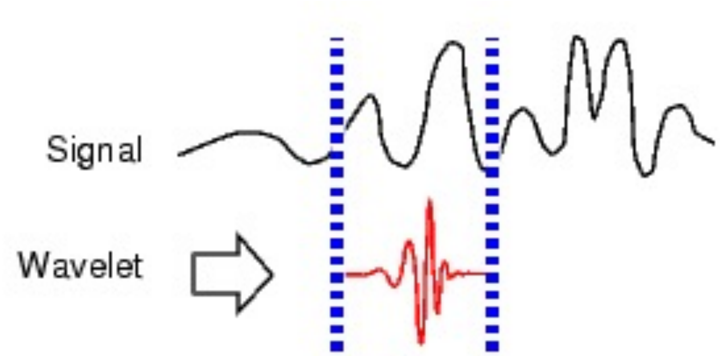


$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t) \psi^* \left(\frac{t - \tau}{s} \right) dt$$

Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Illustrating CWT (cont'd)

3. Shift the wavelet to the right and repeat step 2 until you've covered the whole signal.

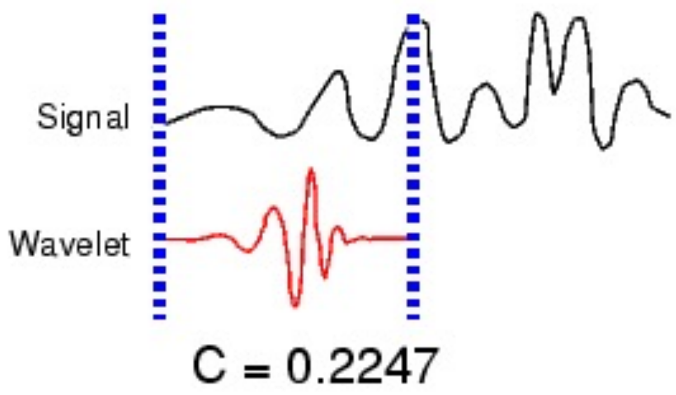


$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t) \psi^* \left(\frac{t - \tau}{s} \right) dt$$

Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Illustrating CWT (cont'd)

4. Scale the wavelet and go to step 1.



$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_t f(t) \psi^* \left(\frac{t - \tau}{s} \right) dt$$

5. Repeat steps 1 through 4 for all scales.

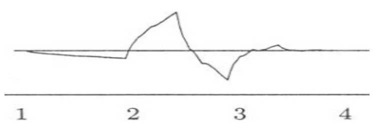
Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Basis Construction – "Mother" Wavelet

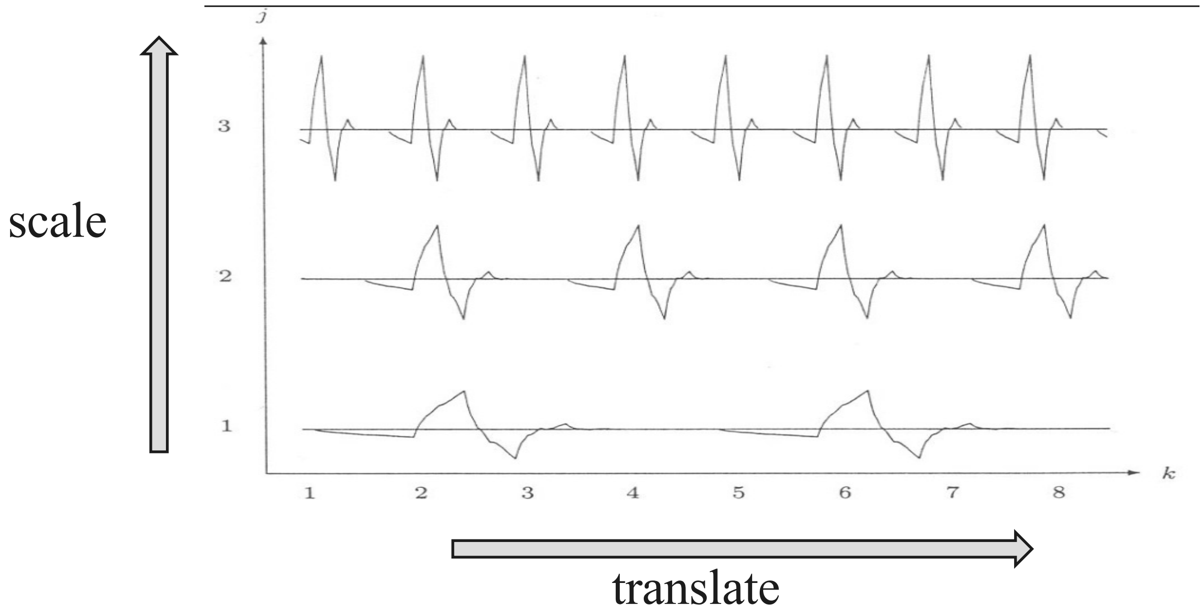
- The basis can be constructed by applying translations and scalings (stretch/compress) on the "mother" wavelet $\psi(t)$:

$$\psi(s,\tau,t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right)$$

Example:



- $\psi(t)$



Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Basis Construction - Mother Wavelet

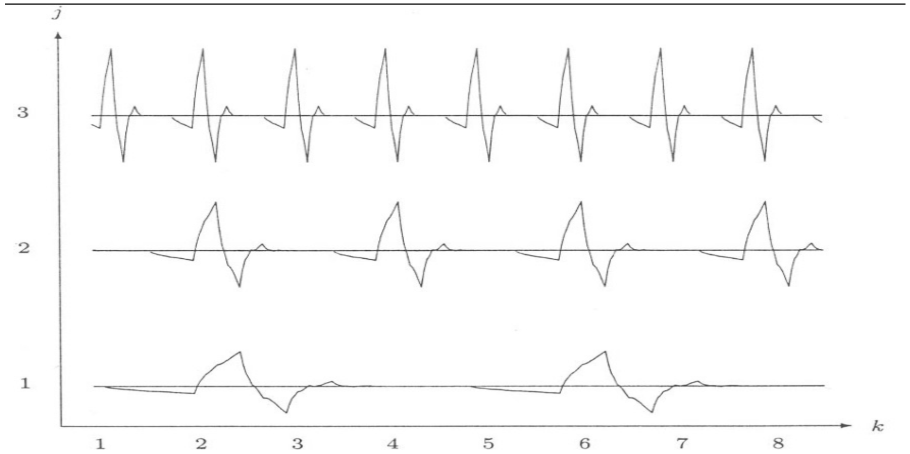
- It is convenient to take special values for s and τ in defining the wavelet basis: $\mathbf{s = 2^{-j}}$ and $\boldsymbol{\tau} \mathbf{= k \, . \, 2^{-j}}$

(dyadic/octave grid)

$$\psi(s,\tau,t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) = \frac{1}{\sqrt{2^{-j}}}\psi\left(\frac{t-k.2^{-j}}{2^{-j}}\right) = 2^{\frac{j}{2}}\psi(2^{j}t-\mathrm{k}) = \psi_{jk}(t)$$

scale $=1/2^{j}$
(1/frequency)

$\uparrow$ j



k $\longrightarrow$

Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

Continuous Wavelet Transform (cont'd)

Forward CWT:
$$C(\tau,s)=\frac{1}{\sqrt{s}}\int_t f(t)\psi^*\left(\frac{t-\tau}{s}\right)dt$$

Inverse CWT:
$$f(t)=\frac{1}{\sqrt{s}}\int_\tau\int_s C(\tau,s)\psi(\frac{t-\tau}{s})d\tau ds$$

Note the double integral

Thanks to Pr Bebis https://slideplayer.com/slide/4897156/

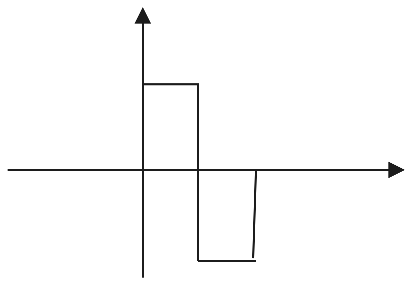
- The set of functions $\{\psi_{j,k}\}_{j,k\in Z^2}$ is an orthogonal basis of $L^2(\mathrm{R})$. Then

$$\forall u \in L^2(\mathrm{R}) \quad u = \sum_{j,k\in \mathrm{Z}^2} c_{j,k}(u)\psi_{j,k} \quad \text{où} \quad c_{j,k}(u) = \langle u, \psi_{j,k}\rangle = \int_{\mathrm{R}} u(x)\psi_{j,k}(x)dx$$

$c_{j,k}(u)$ are the wavelet coefficients of u.

- The oldest known wavelet basis is the Haar basis.

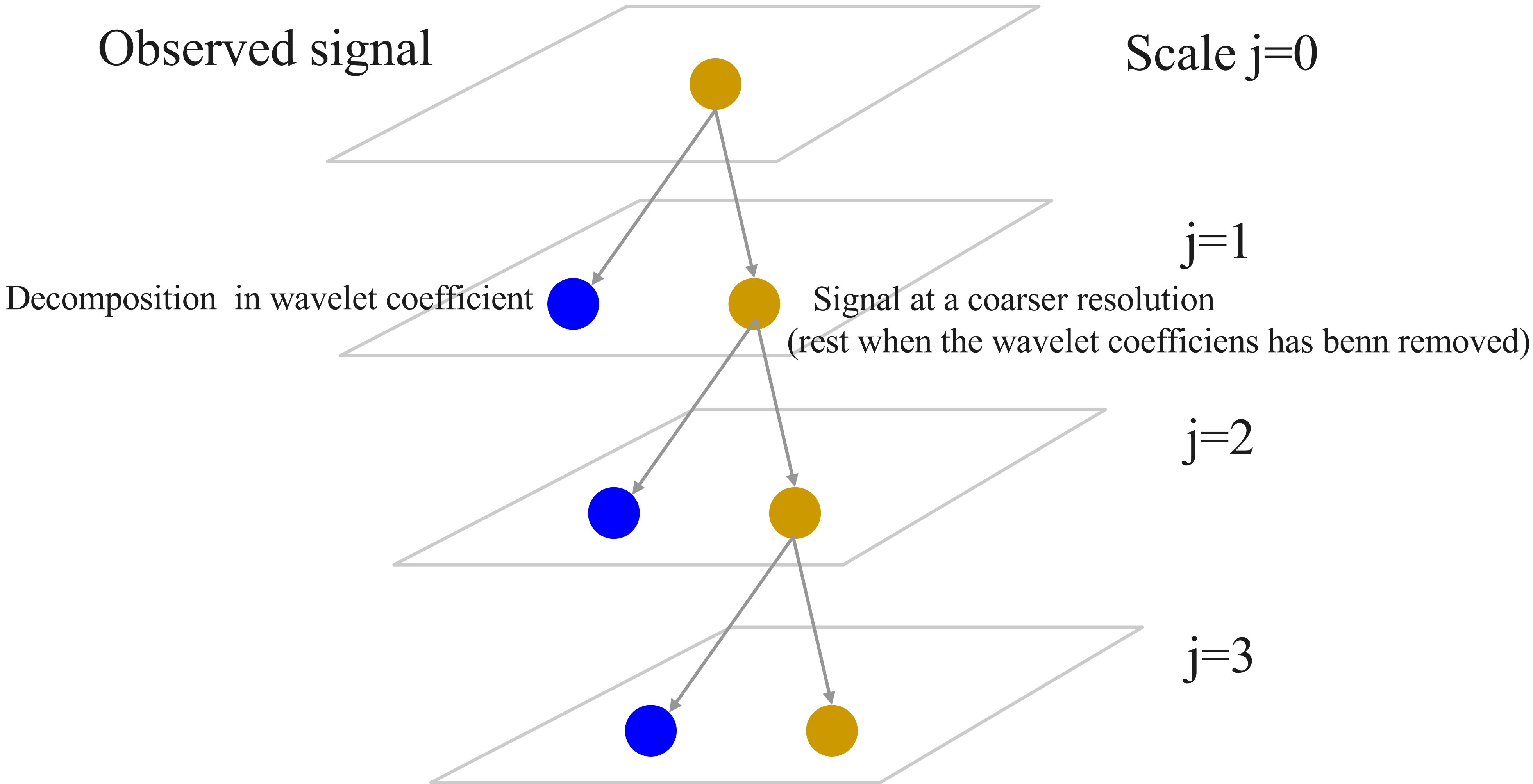
$$\psi(x) = \begin{cases} 1 & \text{si } 0 \le x < \frac{1}{2} \\ -1 & \text{si } \frac{1}{2} \le x < 1 \\ 0 & \text{ailleurs} \end{cases}$$



- Then construction of ψ more regular by [Meyer 86, Lemarié 88 Battle 87, Daubechies 88]

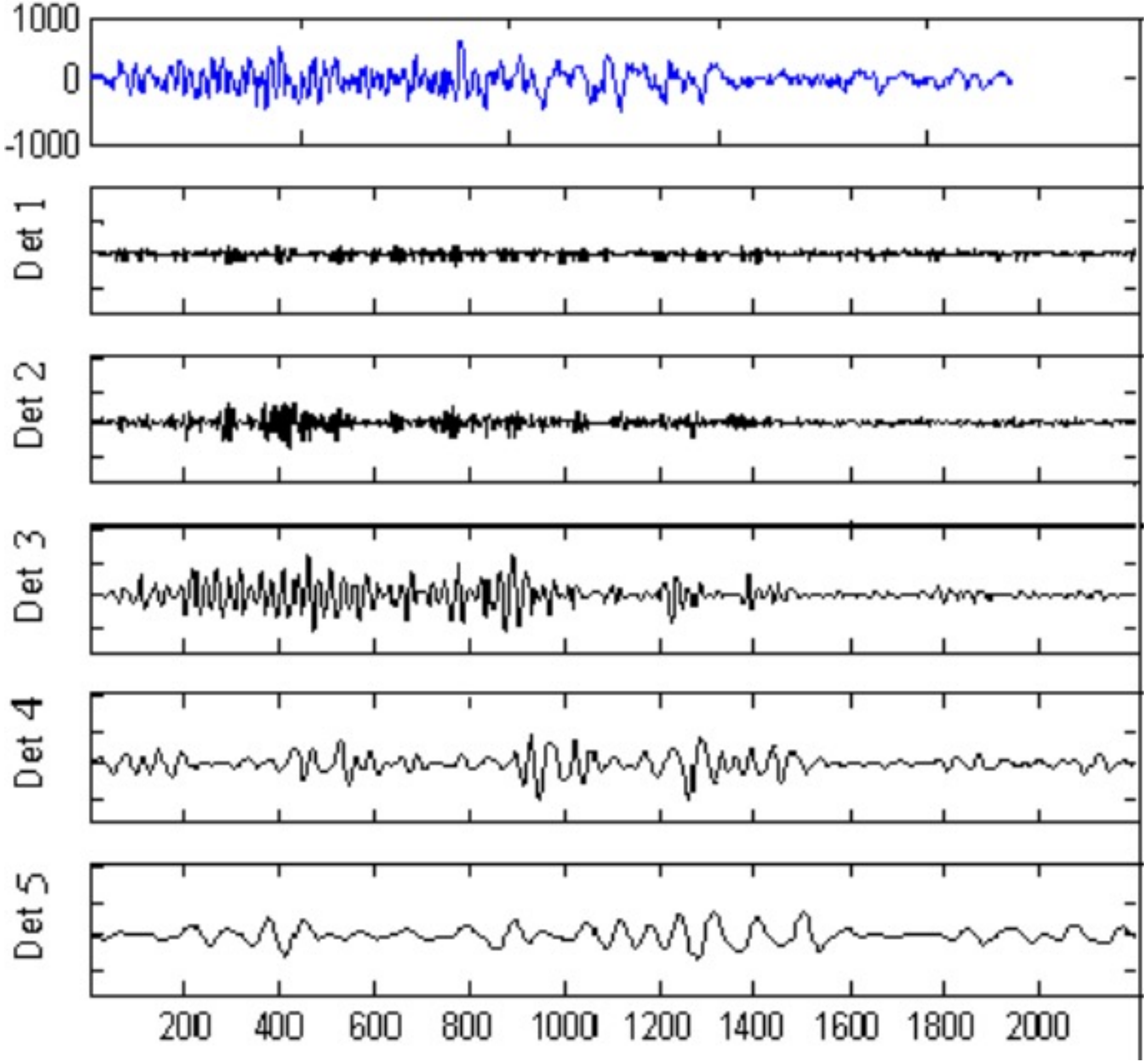
[I. Daubechies « Ten lectures on wavelet » *Number 61 in CBMS-NSF Series in Applied Mathematics*, SIAM Publications, Philadelphia, 1992]

1D wavelet transform



1D wavelet transform

Details are localized (≠ Fourier)



Scale j=0
Original signal

Wavelet coeff. at scale j=1
fine details

Wavelet coeff. at scale j=2

Wavelet coeff. at scale j=5

1D wavelet basis

In practice: **discrete** wavelet transform

The samples of the observed (original) signal are considered as the approximation coefficients of a signal at index j= 0 resolution (2^0). This signal is noted u_0. It is shown that the coefficients of the signal sampled at a lower resolution (resp. the wavelet coefficients) can be calculated by low-pass filtering h (resp. high-pass filtering g).
ϕ and ψ the continuous scaling function and wavelet function are completely defined by the discrete filters h and g

u_{j+1} is calculated by convolution and subsampling of u_j by h.

By successive convolutions, we compute all the approximations of u.

1D wavelet transform

low-pass filtering h
high-pass filtering g

h and g are linked by the relation:

$$g(n) = (-1)^n h(-n+1), \quad \sum_n h(n) = \sqrt{2}, \quad \sum_n g(n) = 0$$

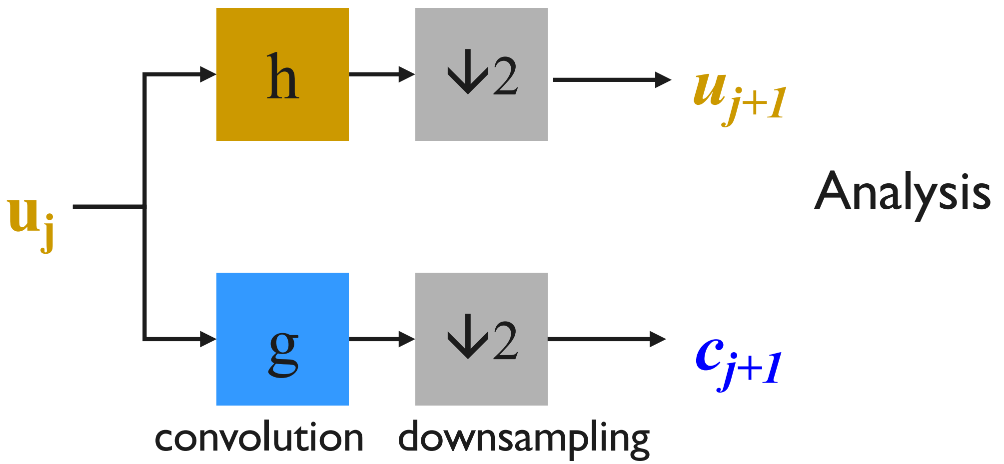
Decomposition/Analysis

low-pass filter h corresponds to the scaling function (signal at coarser resolution)
high-pass filter g corresponds to the wavelet function (wavelet coefficients = details)

Let u a discrete signal. Its decomposition at successive levels j is obtained by recursive filtering

$$u_j(n) = \sum_k h(2n-k)\,u_{j-1}(k)$$

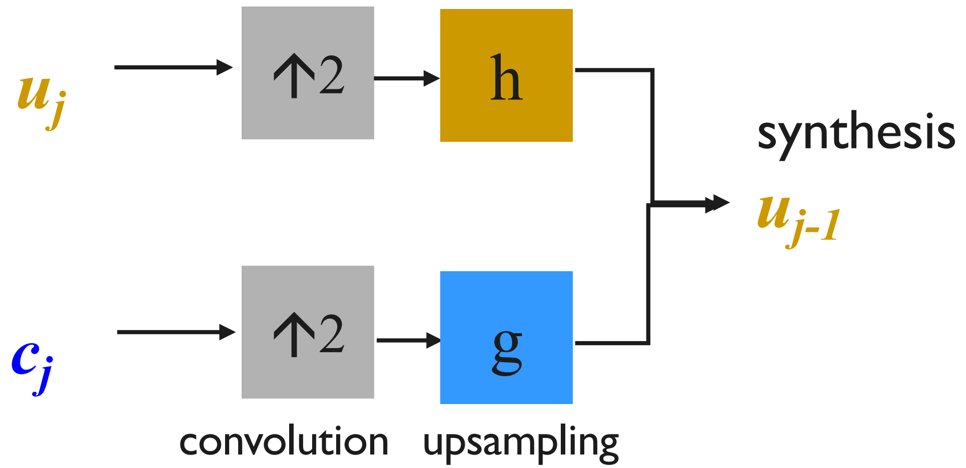
$$c_j^u(n) = \sum_k g(2n-k)\,u_{j-1}(k)$$



Downsampling by factor 2: remove one sample over two.

Reconstruction: synthesis

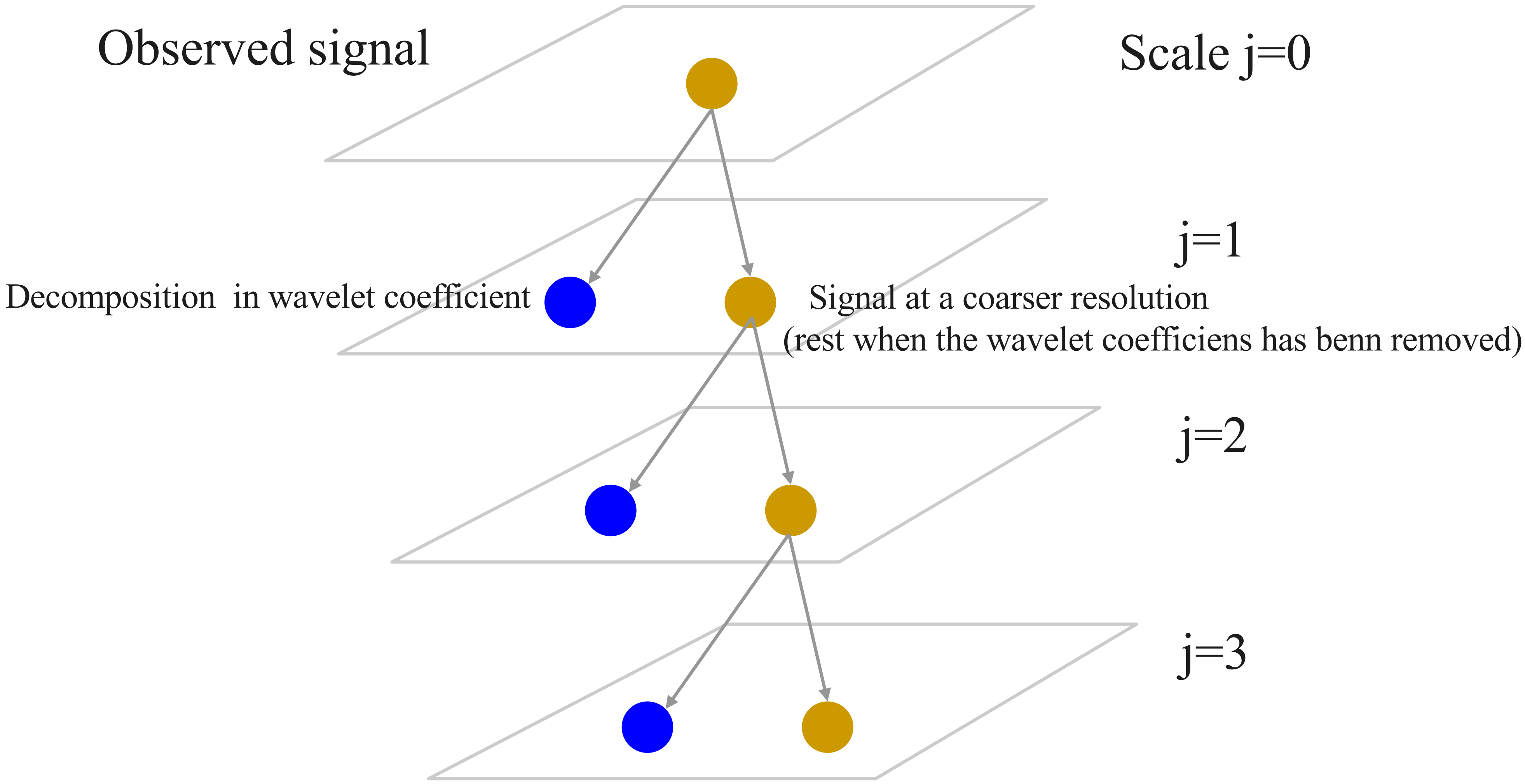
Reconstruction is also obtained by recursive filtering



$$u_{j-1}(k) = \sum_n h(2n-k)\,u_j(n) + \sum_n g(2n-k)\,c_j^u(n)$$

Upsampling by factor: insert one zero between two samples.

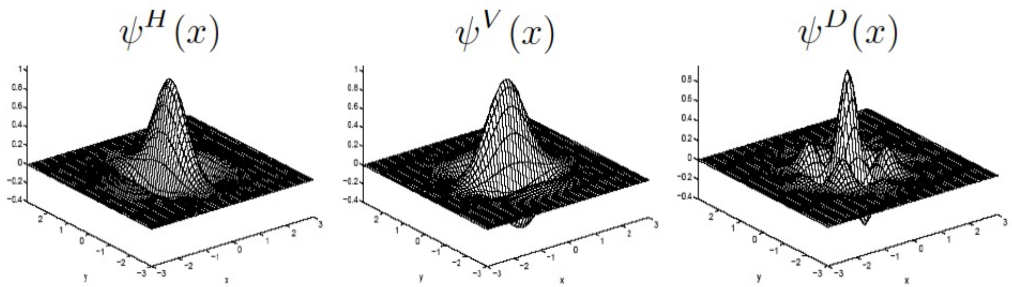
1D wavelet transform



- Separable case : $\phi(x,y) = \phi(x)\ \phi(y)$ where $\phi(x)$ is the $1D$ scaling function. From the $1D$ wavelet ψ, three $2D$ wavelets are defined

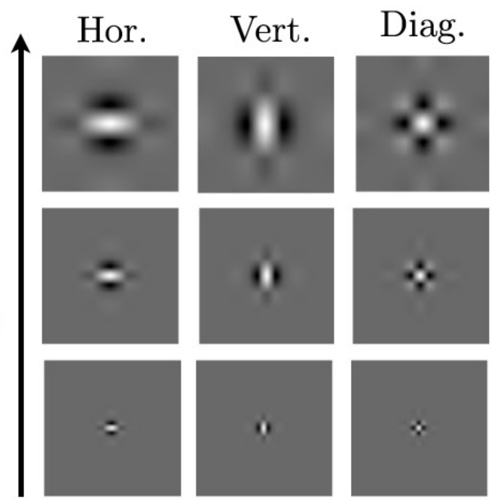
$$\psi^{H}(x,y)=\phi(x)\psi(y), \quad \psi^{V}(x,y)=\psi(x)\phi(y), \quad \psi^{D}(x,y)=\psi(x)\psi(y)$$

3 elementary wavelets $\{\psi^H, \psi^V, \psi^D\}$.



Orthogonal basis of $\mathrm{L}^2([0,1]^2)$:

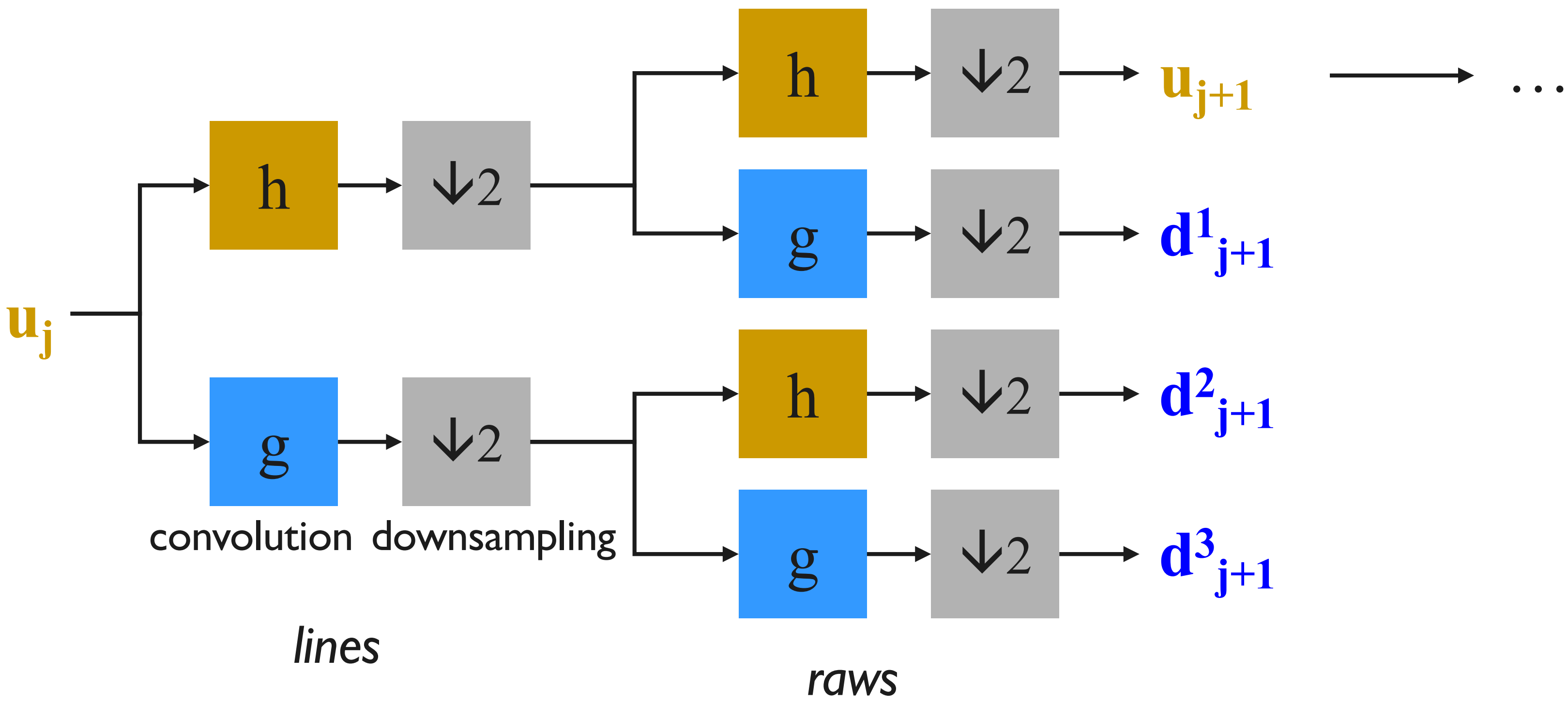
$$\left\{\psi_{j,n}^{k}(x) = 2^{-j}\psi(2^{-j}x - n)\right\}_{j<0,2^j n\in[0,1]^2}^{k=H,V,D}$$



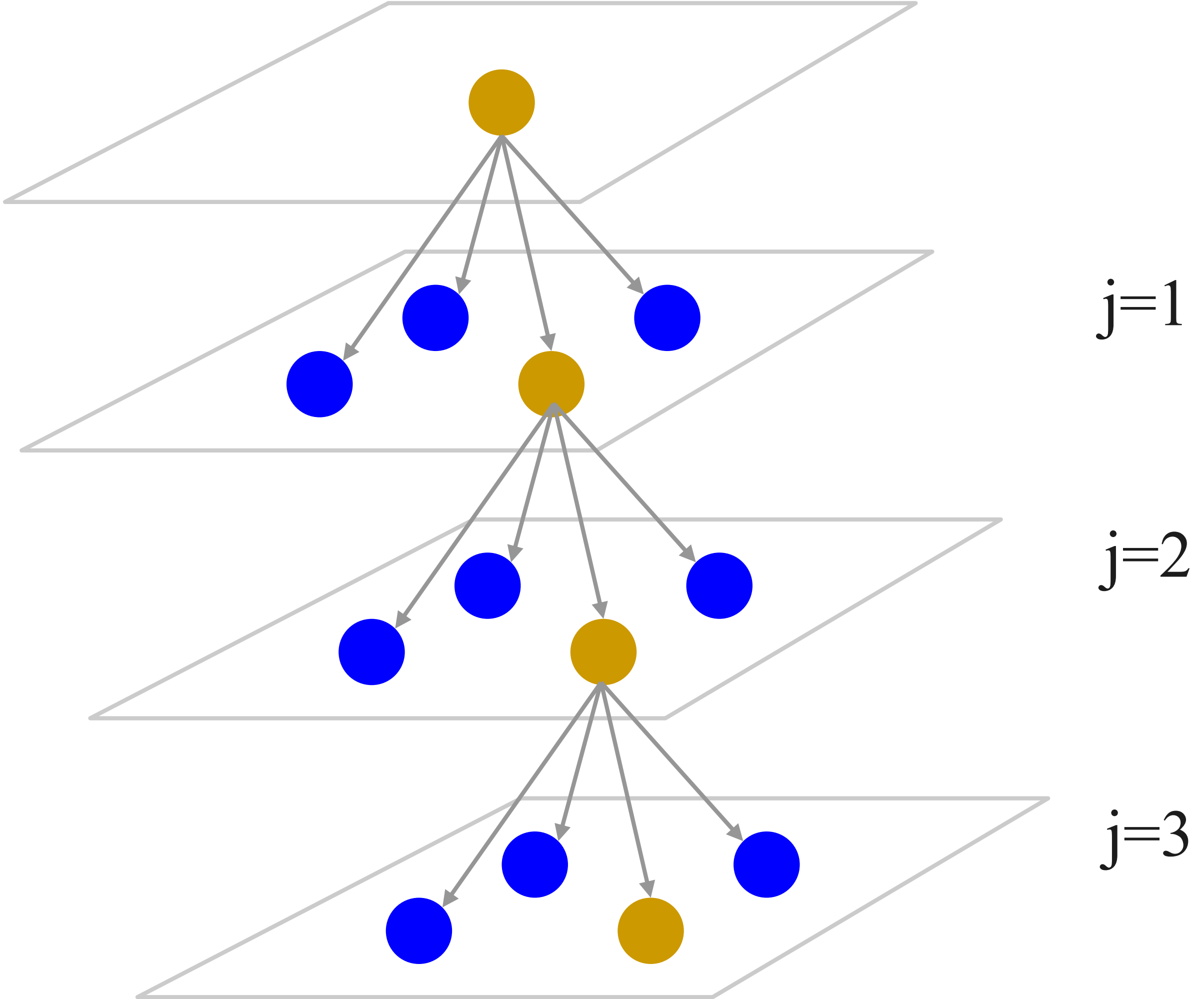
Thanks to Gabriel Peyré https://mathematical-tours.github.io/slides/

2D wavelet transform

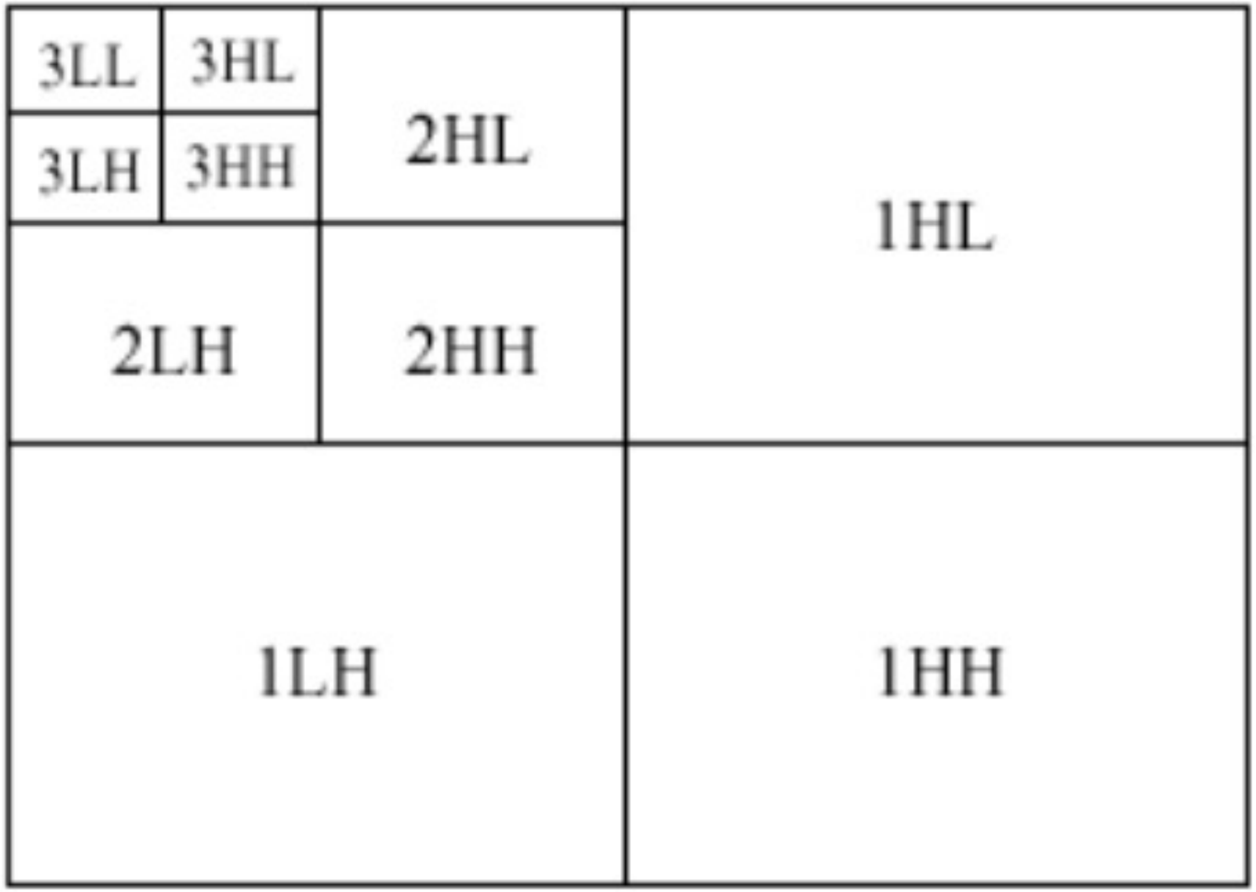
- The scale factor is 2. At each scale, a coarser image is computed u_{j+1}, and 3 images of details d^1_{j+1}, d^2_{j+1}, d^3_{j+1}



2D wavelet transform

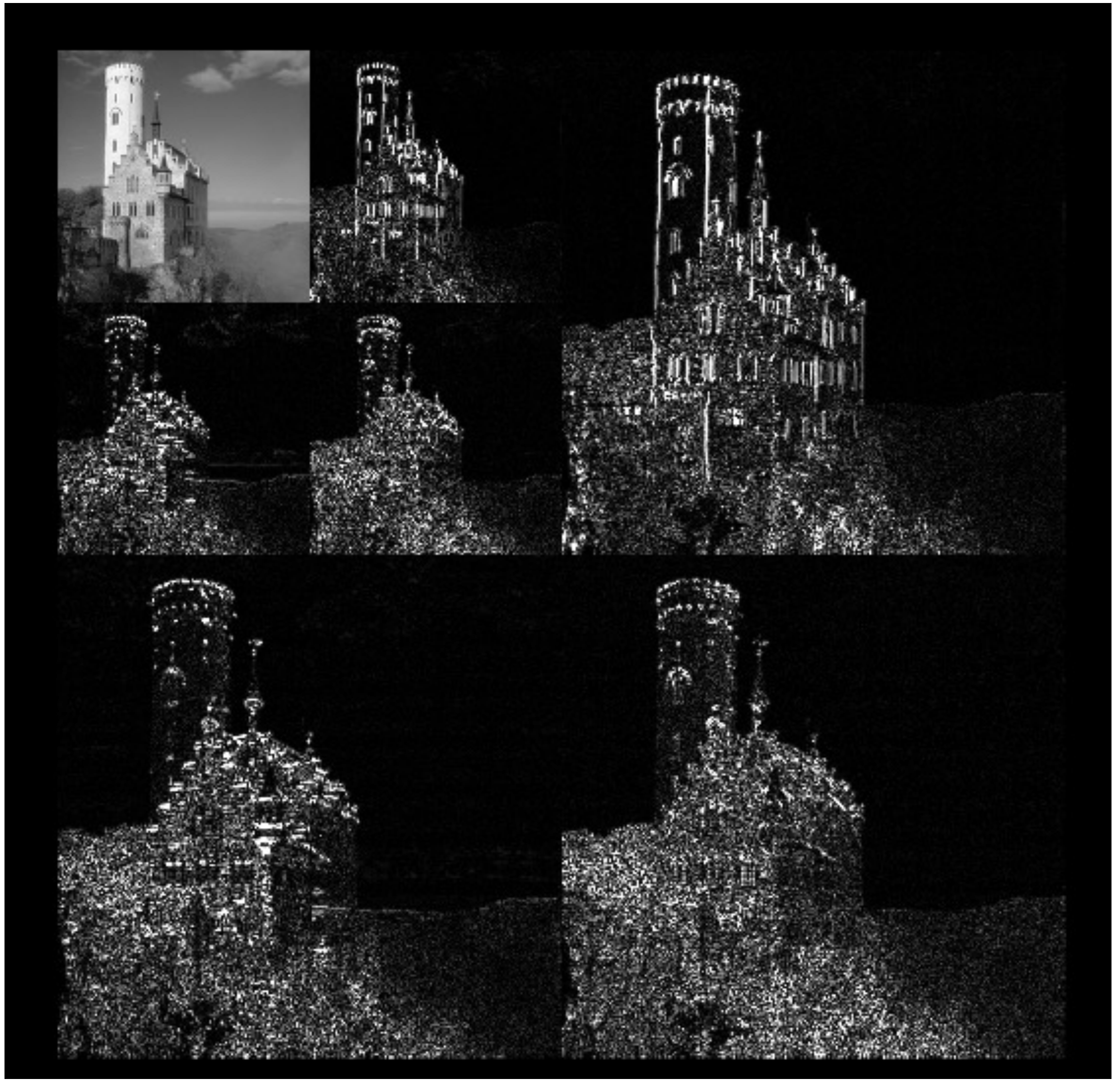


2D wavelet transform



Representation of the 2D wavelet transform of an image

2D wavelet transform



Wikipedia https://en.wikipedia.org/wiki/Wavelet_transform

Why WT for denoising?

- If the WT gives a sparse representation of the image, then the image is represented by few strong coefficients, and many almost zero coefficients. Thus if the image is noisy, the few strong coefficients in the WT domain will be above the noise level, and the rest of the coefficients can be set to 0, because they represent only noise.

- In order to study the optimality of the WT for denoising, we study the problem of approximating any signal (image) u by a restricted number of its coefficients on a base.

Linear approximation

- Let $B=\{e_m\}_{m\in\mathbb{N}}$ an orthonormal basis of an Hilbert space H. For example $e_m = \psi_m$

All $u \in H$ can be decomposed on this basis as $u = \sum_{m=0}^{+\infty} \langle u, e_m \rangle e_m$

It means that from u we can compute its coefficients $\{<u, e_m>\}_m$ in the new basis and from the set of coefficients $\{<u, e_m>\}_m$, u can be exactly reconstructed

Rather than taking all coefficients, we reconstruct u with only M coefficients, we have an approximation of u

$$u_M = \sum_{m=0}^{M} \langle u, e_m \rangle e_m$$

The error of approximation is

$$\varepsilon[M] = \|u - u_M\|^2 = \sum_{m=M}^{+\infty} \left|\langle u, e_m \rangle\right|^2$$

The faster the coefficients $|< u, e_m >|$ when m increases, the better the approximation.

Linear approximation

Image u

Approximated image u_M

u_M is approximated with the first M coefficients of the decomposition

Non linear approximation

- The linear approximation projects a signal onto M orthogonal vectors selected a priori. This approximation is improved by choosing the M vectors according to each signal: this is the non-linear approximation.
- Let $u \in H$. u is approximated with M vectors adaptively selected in the basis $B=\{e_m\}_{m \in \mathrm{N}}$. Let I_M the set of indices of the approximation vectors of u. u_M is the projection of u on $\{e_m\}_{m \in I_M}$.

$$u_M = \sum_{m \in I_M} \langle u, e_m \rangle e_m$$

The approximation error is $\varepsilon[M] = \|u - u_M\|^2 = \sum_{m \notin I_M} |\langle u, e_m \rangle|^2$

Optimal choice: error is minimal. Then e_m for $m \in I_M$ must be such that

|< u, e_m >| are the highest.. These are the vectors best correlated to u, the main structures of u. We must therefore choose a decomposition basis such that it best represents the image in the sense that it contains the same elementary forms.

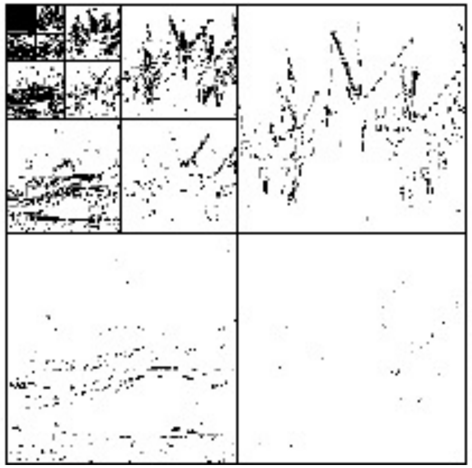
Non linear approximation

Optimal choice:

$$u_M = \sum_{m \in I_M} \langle u, e_m \rangle e_m$$ and I_M = {M largest coefficients $|\langle u, e_m \rangle|$ }.

Original u

Linear approximation



Coefficients

Non Linear approximation

Thanks to Gabriel Peyré https://mathematical-tours.github.io/slides/

Non linear approximation

- Let's rank the coefficients $(|<u, e_m>|)_{m\in\mathbb{N}}$ by descending order. We note the ordered coefficients u^r. We have

$$\left|u^r(k)\right| \geq \left|u^r(k+1)\right|$$

- The optimal nonlinear approximation is then

$$u_M = \sum_{k10}^{M} u^r(k) e_{m_k}$$

- This approximation can also be obtained by applying the hard thresholding function

$$\theta_T^H(t) = \begin{cases} t & \text{si } |t| \geq T \\ 0 & \text{si } |t| < T \end{cases}$$ for threshold T s.t. $u^r(M+1) < T \leq u^r(M)$

The approximation is then

$$u_M = \sum_{m=0}^{+\infty} \theta_T^H\left(\langle u, e_m \rangle\right) e_m$$

Image denoising

Observation equation is $g=u+n$, written in the basis B : $\{e_m\}_{m\in\{0,1,..N-1\}}$.

$$g_B = u_B + n_B \quad \text{où} \quad q_B = \langle q, e_m \rangle$$

n is a white Gaussian noise with zero mean and variance σ^2 ($\mathcal{N}(0,\sigma^2 I)$), this is the same in basis B. Effectively if G is the basis change matrix, it is orthonormal ($G^TG=GG^T=I$ and we have:

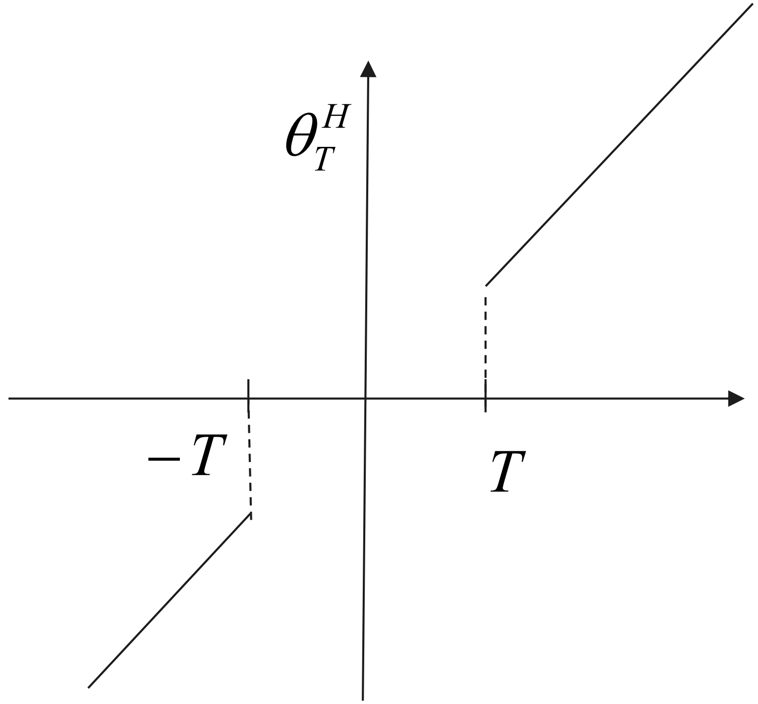
$$E\left[Gn.n^T G^T\right] = GE\left[n.n^T\right]G^T = \sigma_2 I$$

- Denoising by non linear approximation: threshoding

Hard threshold

- The hard threshold function is defined by

$$\theta_T^H(t) = \begin{cases} t & \text{si } |t| > T \\ 0 & \text{si } |t| \le T \end{cases}$$



When B is a wavelet basis, the hard thresholding preserves the strong coefficients that are at the fine scales at the contours. Thus there is no smoothing of the contours. The small coefficients (less than T) are set to 0: in the areas where the image transitions are weak, the noisy coefficients are averaged locally.

Denoising by Hard-Thresholding

Original clean image u

Noisy image $g = u + n$

Denoised image by HT

Thanks to Gabriel Peyré https://mathematical-tours.github.io/slides/

Choice of the threshold

- T must be large enough so that the amplitude of the noise $|n_B(m)|$ in the transform domain has a high probability to be such that $|n_B(m)|<T$.
- T should not be too large so as not to set too many coefficients of the signal to zero.
- For a Gaussian white noise $\mathcal{N}(0,\sigma^2 I)$ it can be shown that the maximum amplitude of the noise has a high probability of being below $T=\sigma\sqrt{2\log N}$
- Theorem [Donoho Johnstone] $T=\sigma\sqrt{2\log N}$

This threshold has good properties (not reported here)

In practice:

- This value is often too high
- This value depends on the size N of the signal which seems not logical.
- We can choose the threshold $T=3\sigma$

[D. Donoho & I.Johnstone « Ideal spatial adaptation via wavelet shrinkage » *Biometrika,* 81, p. 425-455,1994]

Estimation of the noise variance

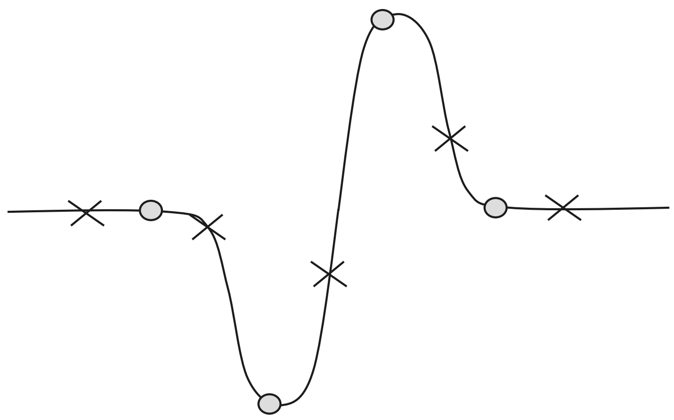
- The estimation of the variance of the noise from noisy observations can be done in the following way: we consider that u is piecewise regular (which is the case for images). Then a robust estimator is computed from the median of the coefficients at the finest scale.

Indeed, if the signal is piecewise regular it will have few large coefficients, and many at 0. So the noisy signal will have few large coefficients and many of the wavelet coefficients are due to noise. At the original resolution (first resolution), the signal component is therefore only on a small number of coefficients. The whole set of these coefficients therefore largely follow a Gaussian distribution of mean zero and variance σ^2. A robust estimator consists in computing the median of the observed coefficients at first scale. We know that if we have P independent Gaussian random variables, of mean zero and variance σ^2 we can show that E(median)=0,6745 σ. So we use

$$\sigma_{est} = \frac{\text{Médiane}\left\{\langle u_0, \psi_{1,m}\rangle, m = 0,..\frac{N}{2}\right\}}{0,6745}$$

Translation invariance

- A wavelet transform (with decimation) is not translation-invariant: the coefficients will not be the same for a signal $u(t)$ and for a shifted version $u(t+\delta)$ as long as δ is not a multiple of the decimation period
- Let $u(t)$ be a pulse. The wavelet coefficients at the first level are exactly the response of the high pass filter:



- The thresholding of the crosses and circles will be different and the reconstructed signal after thresholding will not be an offset version of each other.
- The translation invariant algorithm of [Coifman et Donoho 95] : calculation of the thresholding result for all shifted versions and averaging of the result.

- [R. Coifman & D. Donoho « Translation-Invariant de-noising » rapport de recherche 475, Standford University, 1995]

Invariance par translation

- The previous solution is to use a wavelet transform without decimation and to average the coefficients at the reconstruction. This solution has a great redundancy and the number of samples at each level of resolution increase of factor 2^d where d is the dimension of the space (d=2 for an image).

- Transforms have been introduced to approximate translation (and rotation) invariance with limited redundancy

.

An example is the complex transform on dual tree proposed by N. Kingsbury which has a redundancy of 2^d:1 where d is the dimension of the space.

[N. Kingsbury 1998 : *Proc. EUSIPCO* p. 310-322,Rhodes, Greece
8th IEEE DSP Workshop, Bryce Canyon UT, USA],