Multiresolution Analysis and Fast Wavelet Transform

Fondamenti di elaborazione del segnale multi-dimensionale
Multi-dimensional signal processing

Stefano Ferrari

Università degli Studi di Milano
stefano.ferrari@unimi.it

Elaborazione dei Segnali Multi-dimensionali e Applicazioni
2014–2015

Motivations

- CWT has valuable properties for signal processing.
- However, the use of CWT requires some approximations:
  - inner product computation;
  - scale and translation parameters sampling.
- A discrete version of wavelet transform (i.e., a wavelet transform that operates with only a dyadic set of wavelets and on a discrete set of samples of the signal) is possible: the Discrete Wavelet Transform (DWT).
- The theory that allows to obtain such a transform is better explained starting from the Multi-Resolution Analysis (MRA).
- A fundamental result of the MRA theory is that, under some conditions, the DWT can be obtained through a digital filtering operation.
- This transform is computationally very efficient and, for this reason, it is called Fast Wavelet Transform (FWT).

Multiresolution Analysis — Overview

- A Multiresolution Analysis (MRA) defines a sequence of nested spaces of functions, $\{V_j\}$:

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset V_j \subset V_{j+1} \subset \ldots$$

  such that lower the index, smoother the functions that belong to the space.
- This sequence will, at the end, cover the space of the finite energy functions, $L^2(\mathbb{R})$.
- For each function $f \in L^2(\mathbb{R})$, the best approximation, $P_j[f]$, in each space, V_j, can be defined by projecting the function onto this space.
- Hence, a sequence of approximating functions, $\{P_j[f]\}$, is obtained, such that:

$$\lim_{j\to\infty} P_j[f] = f$$

Multiresolution Analysis — Overview (2)

- The difference between two consecutive approximations represents the details that are added:

$$Q_j[f] = P_{j+1}[f] - P_j[f]$$

  and can be obtained as the projection of the function f onto an appropriate detail space, W_j.
- Hence, the function f can be represented by summing the sequence of the detail projections:

$$f = \sum_j Q_j[f]$$

- The basis of the W_j's spaces are the wavelets.

Scaling functions — Approximation spaces

A Multiresolution Analysis (MRA) of $L^2(\mathbb{R})$ is defined as the sequence of closed subspaces $V_j \in L^2(\mathbb{R})$, $j \in \mathbb{Z}$, which have the following properties:

1. $V_j \subset V_{j+1}$
2. $v(x) \in V_j \Leftrightarrow v(2x) \in V_{j+1}$
3. $v(x) \in V_0 \Leftrightarrow v(x+1) \in V_0$
4. $\bigcup_{j=-\infty}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$ and $\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$
5. There is a function $\varphi(x) \in V_0$, having non null integral, such that the set $\{(\varphi(x-k) | k \in \mathbb{Z}\}$ is a Riesz basis for V_0.

The function $\varphi(\cdot)$ is called *scaling function*.

Scaling functions — Approximation spaces (2)

There is a sequence $\{h_k\} \in l^2(\mathbb{Z})$ for which the scaling function satisfies:

$$\varphi(x) = 2 \sum_k h_k \, \varphi(2x - k)$$

- The relation is called the *refinement equation*;
  - aka *dilation equation* or *two-scale difference equation*
- Defining
  $$\varphi_{j,k}(x) = \sqrt{2^j} \, \varphi(2^j x - k)$$
  it can be shown that $\{\varphi_{j,k}(x) \,|\, k \in \mathbb{Z}\}$ is a Riesz basis for V_j
- Hence, $\{\varphi_{j,k}(x) \,|\, j, k \in \mathbb{Z}\}$ is a Riesz basis for $L^2(\mathbb{R})$

Scaling functions — Approximation spaces (3)

Hence there are at least three ways to build or identify a MRA:

- through the description of the V_js spaces;
- by means of the scaling function, φ;
- through the coefficients $\{h_k\}$ of the refinement equation.

As it will be shown, in order to obtain an approximation, the coefficients $\{h_k\}$ can be used directly.

- It is efficient.
- There is no need of using the scaling function.

However, a more detailed characterization of these coefficients is required.

Properties of the scaling functions

- It can be shown that:

$$\sum_k h_k = 1$$

- The normalization is a condition usually required:

$$\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1$$

  - In the frequency domain, this condition is equivalent to:

$$\hat{\varphi}(0) = 1$$

- From the refinement equation and the normalization condition, the scaling function is uniquely determined.

Properties of the scaling functions (2)

- In order to be able to approximate simple function (e.g., constants), it is useful to assume that:

$$\forall x \in \mathbb{R}, \sum_k \varphi(x-k) = 1$$

  - the scaling function and its integer translates partition the unit.

- This condition is equivalent to:

$$\hat{\varphi}(2\pi k) = 0, \ k \in \mathbb{Z}, \ k \neq 0$$

  - or $\hat{\varphi}(2\pi k) = \delta, \ k \in \mathbb{Z}$, due to $\hat{\varphi}(0) = 1$.

Properties of the scaling functions (3)

- From the refinement equation, $\varphi(x) = 2\sum_k h_k \, \varphi(2x-k)$:

$$\hat{\varphi}(\nu) = H(\nu/2)\hat{\varphi}(\nu/2)$$

  where H is a 2π-periodic function defined as:

$$H(\nu) = \sum_k h_k \, e^{-\iota k \nu}$$

- Since $\hat{\varphi}(0) = 1$, the recursion on the above property produces:

$$\hat{\varphi}(\nu) = \prod_{j=1}^{\infty} H(2^{-j}\nu)$$

  This relation can be used for obtaining φ from $\{h_k\}$.

Properties of the scaling functions (4)

- It can be shown that $H(0) = 1$.
  - E.g., from $\hat{\varphi}(\nu) = H(\nu/2)\hat{\varphi}(\nu/2)$.
- It can also be shown that a condition for the partition of the unity is:

$$H(\pi) = 0 \quad \text{or} \quad \sum_k (-1)^k\, h_k = 0$$

Approximation at the j-th scale

For each function, its approximation can be obtained projecting it onto an approximation space:

$\forall f(\cdot) \in L^2(\mathbb{R})$, $\lim_{j\to\infty} P_j[f(\cdot)] = f(\cdot)$

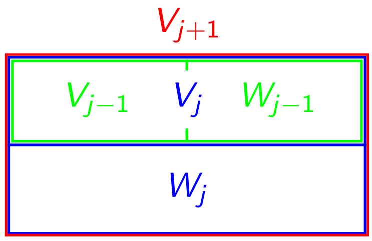
$$\boxed{P_j[f(x)] = \sum_k \lambda_{j,k}\, \varphi_{j,k}(x)}$$

for some $\{\lambda_{j,k}\}$.

Wavelets — Detail spaces

Let W_j be the complementary space of V_j in V_{j+1}, i.e., the space that satisfies:

$$\begin{aligned} V_{j+1} &= V_j \oplus W_j \\ &= \{v_j + w_j \,|\, v_j \in V_j,\, w_j \in W_j\} \end{aligned}$$



The space W_j contains the information about the "details" required for moving from a j-resolution approximation to the $j+1$-resolution one. As a consequence:

$$\bigoplus_j W_j = L^2(\mathbb{R})$$

Wavelets — Details space (2)

A function $\psi(\cdot)$ is a *wavelet* if the set of functions $\{\psi(x-k) \,|\, k \in \mathbb{Z}\}$ is a Riesz basis for the wavelet space W_0.

$$\{\psi_{j,k}(x) \,|\, j,k \in \mathbb{Z}\}, \text{ where } \psi_{j,k}(x) = \sqrt{2^j}\,\psi(2^j x - k),$$
is a Riesz bases for $L^2(\mathbb{R})$.

Since $\psi \in V_1$, there is a sequence $\{g_k\} \in l^2(\mathbb{Z})$ such that:

$$\psi_{0,0}(x) = \psi(x) = 2\sum_k g_k\,\varphi(2x-k)$$

The function $\psi(\cdot)$ is called *mother wavelet*.

.

Properties of the wavelets

- The Fourier transform of the wavelet is:

$$\hat{\psi}(\nu) = G(\nu/2)\,\hat{\psi}(\nu/2)$$

where G is a 2π-periodic function given by:

$$G(\nu) = \sum_k g_k e^{-\iota k \nu}$$

Detail at the j-th scale

As for the approximation, the detail at a given scale can be obtained by projecting the function onto a proper wavelet space:

$\forall f(x) \in L^2(\mathbb{R})$:

$$\boxed{f(x) = \sum_j Q_j[f(x)] = \sum_{j,k} \gamma_{j,k}\,\psi_{j,k}(x)}$$

Notes:

- The above equation is a “discrete” (in the scale and position parameters) inverse wavelet transform.
- The computational cost for computing the coefficients $\{\gamma_{j,k}\}$ depends by the properties of the wavelets and scaling functions.

Orthogonal wavelets

- The use of an orthogonal basis is particularly interesting as it allows to decompose a function in uncorrelated elements.
- In this case, the coefficient $\lambda_{j,k}$ are obtained by the orthogonal projection of the function f onto the basis element $\varphi_{j,k}$:

$$P_j[f(x)] = \sum_k \langle f, \varphi_{j,k}\rangle \, \varphi_{j,k}(x)$$

- $P_j[f(\cdot)]$ is the best representation of $f(\cdot)$ in V_j, as:

$$\forall g \in V_j, \; ||g - f|| \geq ||P_j[f] - f||$$

- Similarly, if the wavelets $\{\psi_{j,k}\}$ form an orthogonal basis for W_j, the projection Q_j is an orthogonal projection and the coefficient $\gamma_{j,k}$ can be obtained by orthogonally projecting f onto $\psi_{j,k}$:

$$Q_j[f(x)] = \sum_k \langle f, \psi_{j,k}\rangle \, \psi_{j,k}(x)$$

Orthogonal wavelets (2)

- A MRA where the wavelet spaces W_j are defined as the orthogonal complement of V_j in V_{j+1}.
  - As a consequence, the wavelet spaces, $\{W_j\}$, are mutually orthogonal,
  - the above defined projections P_j and Q_j are orthogonal, and
  - the expansion

    $$f(x) = \sum_j Q_j[f(x)]$$

    is an expansion of orthogonal functions.
- If the above mentioned conditions on the scaling function are satisfied, a sufficient condition for the orthogonality of a MRA is:

$$W_0 \perp V_0$$

or

$$\langle \psi(x), \varphi(x - k)\rangle = 0$$

.

Orthogonal wavelets (3)

- Under mild conditions, $\langle \psi(x), \varphi(x-k) \rangle = 0$ is equivalent to:

$$\forall \nu \in \mathbb{R}, \ \sum_k \hat{\psi}(\nu + 2k\pi)\bar{\hat{\varphi}}(\nu + 2k\pi) = 0$$

- In order to investigate on the properties of the orthogonal wavelets and scaling functions, the following 2π-periodic function is introduced:

$$F(\nu) = \sum_k |\hat{\varphi}(\nu + 2k\pi)|^2$$

- Since $\{\varphi(x-k) \,|\, k \in \mathbb{Z}\}$ is a Riesz basis, there are two constants A and B such that:

$$0 < A \leq F(\nu) \leq B < \infty$$

  i.e., $F(\cdot)$ is bounded (and the bounds do not depend on ν).

Orthogonal wavelets (4)

- Since $\hat{\varphi}(\nu) = H(\nu/2)\,\hat{\varphi}(\nu/2)$, it derives:

$$F(2\nu) = |H(\nu)|^2\, F(\nu) + |H(\nu + \pi)|^2\, F(\nu + \pi)$$

  which shows that F is actually π-periodic.
- The scaling function is orthogonal when

$$\langle \varphi(x),\, \varphi(x-k) \rangle = \delta_k, \ k \in \mathbb{Z}$$

  In this case, $\{\varphi_{j,k} \,|\, k \in \mathbb{Z}\}$ is an orthonormal basis for V_j.
- Under mild conditions, the above relation is equivalent to:

$$\forall \nu \in \mathbb{R}, \ \sum_k |\hat{\varphi}(\nu + 2k\pi)|^2 = F(\nu) = 1$$

- Hence

$$\forall \nu \in \mathbb{R}, \ |H(\nu)|^2 + |H(\nu + \pi)|^2 = 1$$

  which is equivalent to

$$\forall k \in \mathbb{Z}, \ \sum_j h_j h_{j-2k} = \frac{\delta_k}{2}$$

Orthogonal wavelets (5)

- $\sum_j h_j h_{j-2k} = \frac{\delta_k}{2}$ and $\langle \varphi(x), \varphi(x-k) \rangle = \delta_k$ describe the orthogonality necessary conditions in the time domain;
- $\sum_k |\hat{\varphi}(x + 2k\pi)|^2 = 1$ and $|H(\nu)|^2 + |H(\nu + \pi)|^2 = 1$ describes the orthogonality necessary conditions in the frequency domain.
- These conditions can be used to build orthogonal scaling functions.
- Similarly, the basis $\{\psi_{j,k} \mid k \in \mathbb{Z}\}$ is an orthonormal basis for W_0 if

$$\langle \psi(x), \psi(x-k) \rangle = \delta_k$$

or, equivalently

$$\sum_k |\hat{\psi}(\nu + 2k\pi)|^2 = 1$$

from which results the necessary condition:

$$|G(\nu)|^2 + |G(\nu + \pi)|^2 = 1$$

Orthogonal wavelets (6)

- The G function (and, hence the g coefficients), can be better characterized.
- It can be shown that

$$\forall \nu \in \mathbb{R}, \; G(\nu)\, \bar{H}(\nu) + G(\nu + \pi)\, \bar{H}(\nu + \pi) = 0$$

- An important result [Mallat, 1989] show that

$$G(\nu) = A(\nu)\bar{H}(\nu + \pi)$$

where A is a 2π periodic function such that:

$$A(\nu + \pi) = -A(\nu)$$

- With the above conditions,

$$|A(\nu)| = 1$$

- Hence, the above relations allow to build an orthogonal wavelet given the orthogonal scaling function, for a chosen A.

Orthogonal wavelets (7)

- For practical uses, the compactness of the wavelet and scaling function is very important.
- It can be shown that this can be obtained for

$$A(\nu) = Ce^{-(2k+1)\nu}, \quad \text{for } |C| = 1 \text{ and } k \in \mathbb{Z}$$

- The standard choice is

$$A(\nu) = e^{-\iota\nu}$$

  for which G and H are the transfer functions of a pair of quadrature mirror filters:

$$g_k = (-1)^k \bar{h}_k$$

- This choice has also the advantage of yielding real coefficients g_ks, provided that also h_ks are reals.

Biorthogonal wavelets

- The orthogonality puts strong limitation on the construction of the wavelets (e.g., on compactness of the wavelets).
- More flexibility can be achieved by using biorthogonal wavelets.
- The definition on a compact domain allows for an accurate implementation of the transform.
- In this case, the wavelet and the scaling function are represented by FIR filters,
  - h_k and g_k have a finite number of non-null coefficients.

Biorthogonal wavelets — Dual spaces

- The biorthogonal MRA requires the existence of a *dual scaling function*, $\tilde{\varphi}$, and a *dual wavelet*, $\tilde{\psi}$.
- They generate a dual multiresolution analysis with subspaces $\tilde{V}_j$ and $\tilde{W}_j$ such that:

$$\tilde{V}_j \perp W_j \text{ and } V_j \perp \tilde{W}_j$$

- Hence

$$\tilde{W}_j \perp W_{j'} \quad \text{for } j' \neq j$$

- The above orthogonality relations imply:

$$\langle \tilde{\varphi}(x), \psi(x-k) \rangle = \langle \tilde{\psi}(x), \varphi(x-k) \rangle = 0$$

Biorthogonal wavelets — Dual spaces (2)

- Moreover:

$$\langle \tilde{\varphi}_{j,l}, \varphi_{j,k} \rangle = \delta_{l-k} \quad j,k,l \in \mathbb{Z}$$

$$\langle \tilde{\psi}_{j,l}, \psi_{i,k} \rangle = \delta_{j-i}\delta_{l-k} \quad j,k,l \in \mathbb{Z}$$

- In particular

$$\langle \tilde{\varphi}(x), \varphi(x-k) \rangle = \delta_k \quad k \in \mathbb{Z}$$

$$\langle \tilde{\psi}(x), \psi(x-k) \rangle = \delta_k \quad k \in \mathbb{Z}$$

- The properties of the dual wavelet and scaling function, are similar to those of the wavelet and scaling function, respectively.

Biorthogonal wavelets — Dual spaces (3)

- The role of primal and dual MRA is interchangeable:
  - both can have the role of the primal or the dual MRA;
  - the effects on the transform and the inverse will depend on the characteristics of the primal and the dual.
- However, the biorthogonal MRA maintains the main advantage of the orthogonal MRA:
  - the coefficients can be computed by means of orthogonal projections;
  - the dual MRA is used to compute the transform (analysis MRA);
  - the primal MRA is used to reconstruct the signal from the transform coefficients (synthesis).
- The projection operator P_j and Q_j are here defined as:

$$P_j[f(x)] = \sum_k \langle f, \tilde{\varphi}_{j,k} \rangle \, \varphi_{j,k}(x)$$

and

$$Q_j[f(x)] = \sum_k \langle f, \tilde{\psi}_{j,k} \rangle \, \psi_{j,k}(x)$$

Biorthogonal wavelets — Dual spaces (4)

- Hence, the discrete wavelet transform is:

$$f(x) = \sum_{j.k} \langle f, \tilde{\psi}_{j,k} \rangle \, \psi_{j,k}(x)$$

- Properties and conditions similar to those obtained to the orthogonal MRA can be obtained. In particular:

$$\tilde{\varphi}(x) = 2 \sum_k \tilde{h}_k \tilde{\varphi}(2x-k) \text{ and } \tilde{\psi}(x) = 2 \sum_k \tilde{g}_k \tilde{\psi}(2x-k)$$

from which can be obtained

$$\tilde{h}_{k-2l} = \langle \tilde{\varphi}(x-l), \varphi(2x-k) \rangle \text{ and } \tilde{g}_{k-2l} = \langle \tilde{\psi}(x-l), \varphi(2x-k) \rangle$$

Biorthogonal wavelets — Dual spaces (5)

- In particular, by writing $\varphi(2x-k) \in V_1$ as element of V_0 and W_0:

$$\varphi(2x-k) = \sum_l \tilde{h}_{k-2l}\varphi(x-l) + \sum_l \tilde{g}_{k-2l}\psi(x-l)$$

- By imposing that h_k, g_k, $\tilde{h}_k$, $\tilde{g}_k$ have finite components, it can be shown that, under mild conditions:

$$\tilde{G}(\nu) = e^{-\iota\nu}\bar{H}(\nu+\pi) \text{ and } G(\nu) = e^{-\iota\nu}\bar{\tilde{H}}(\nu+\pi)$$

The properties of the orthogonal and biorthogonal MRA can be used to formulate an efficient algorithm for computing the wavelet transform and its inverse.

Fast Wavelet Transform

- As $V_j = V_{j-1} \oplus W_{j-1}$, $v_j \in V_j$ can be uniquely write as sum of a function $v_{j-1} \in V_{j-1}$ and a function $w_{j-1} \in W_{j-1}$:

$$\begin{aligned} v_j(x) &= \sum_k \lambda_{j,k}\varphi_{j,k}(x) = v_{j-1}(x) + w_{j-1}(x) \\ &= \sum_k \lambda_{j-1,k}\varphi_{j-1,k}(x) + \sum_k \gamma_{j-1,k}\psi_{j-1,k}(t) \end{aligned}$$

  for proper coefficients $\{\lambda_{j,k}\}$, $\{\lambda_{j-1,k}\}$, $\{\gamma_{j-1,k}\}$.
- Hence the same function v_j can be represented either by means the sequence $\{\lambda_{j,k}\}$, and by the sequences $\{\lambda_{j-1,k}\}$ $\{\gamma_{j-1,k}\}$.
- This is a key relation for obtaining an efficient algorithm for the analysis and synthesis.

Fast Wavelet Transform (2)

- In fact:

$$\begin{aligned}\lambda_{j-1,l} &= \langle v_j, \tilde{\varphi}_{j-1,l}\rangle = \sqrt{2}\langle v_j, \textstyle\sum_k \tilde{h}_{k-2l}\tilde{\varphi}_{j,k}\rangle \\ &= \sqrt{2}\,\textstyle\sum_k \tilde{h}_{k-2l}\,\lambda_{j,k}\end{aligned}$$

  and, similarly,

$$\gamma_{j-1,l} = \sqrt{2}\sum_k \tilde{g}_{k-2l}\,\lambda_{j,l}$$

  The refinement equations allow to obtain the inverse transform:

$$\lambda_{j,k} = \sqrt{2}\sum_l h_{k-2l}\,\lambda_{j-1,l} + g_{k-2l}\,\gamma_{j-1,l}$$

- The recursive application of these formulas provide the Fast Wavelet Transform (FWT) or *cascade algorithm*.

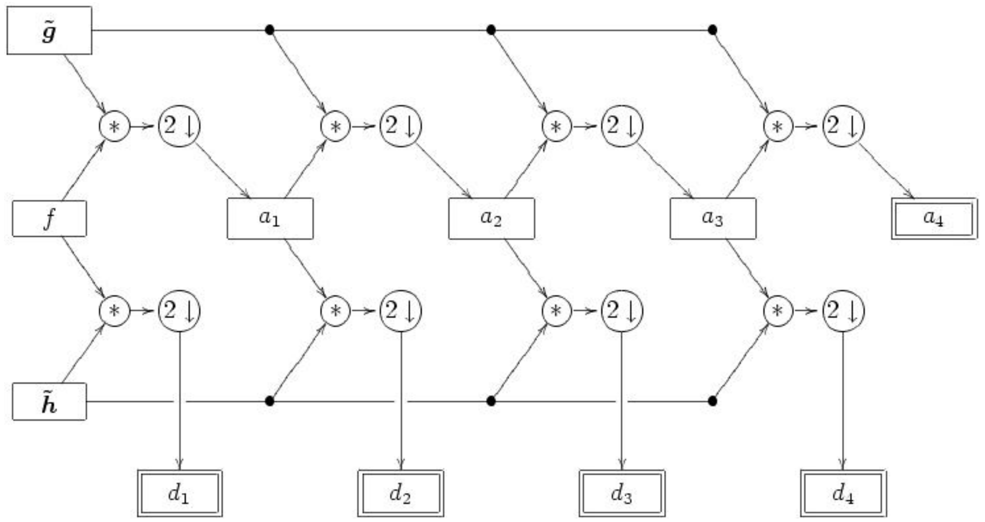
Fast Wavelet Transform (3)

- It should be noticed that the filters $\tilde{h}$ and $\tilde{g}$ are translated by two positions.
- Hence the $\lambda_{j-1,l} = \sqrt{2}\sum_k \tilde{h}_{k-2l}\,\lambda_{j,k}$ do not describe a convolution.
- However, they can be computed as a convolution followed by a subsampling.
- If the signal is defined over an interval, the number of $\lambda_{j,k}$ coefficients will be the double of that of $\lambda_{j-1,k}$ and $\gamma_{j-1,k}$.
- The number of coefficients to represent the signal does not change.
- The inverse transform can be obtained by upsampling the coefficients $\lambda_{j-1,k}$ and $\gamma_{j-1,k}$, putting zeros between the coefficients.

.

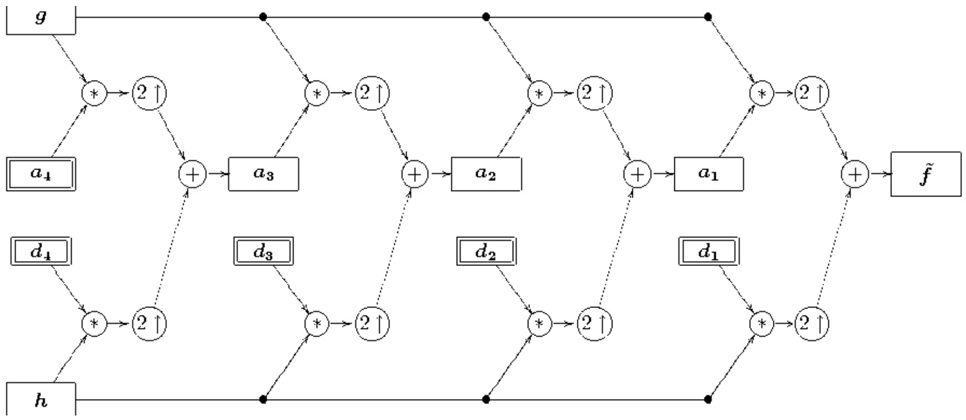
Fast Wavelet Transform (FWT) (4)

Transform scheme



Fast Wavelet Transform (FWT) (5)

Inverse transform scheme



Fast Wavelet Transform (6)

- A problem is the estimate of the initial coefficients λ_0.
- They should be the inner product of the (mother) scaling function and the signal itself.
- A simple choice is using a sampling of the signal for the starting level, n:

$$\lambda_{n,l} = f\left(\frac{l}{2^n}\right)$$

- It is equivalent to suppose that the initial scaling function is an approximation of the Dirac's δ.
- It is important to notice that the FWT allows to obtain an *exact* inner product of the signal with the basis functions of the successive levels, by using only the $\lambda_{n,l}$ coefficients.

Plotting the basis functions

- The basis functions (wavelet and scaling function) sometimes cannot be expressed analytically.
- In this case, the cascade algorithm can be used to obtain an approximation of them.
- In fact, $f \in L^2(\mathbb{R})$ can be represented as:

$$f(x) = \sum_k \lambda_{j,k}\varphi_{j,k}(x) + \sum_{l \geq j}\sum_k \gamma_{l,k}\psi_{l,k}(x)$$

- From proper coefficients $\{\lambda\}$ and $\{\gamma\}$, the function f can be reconstructed.
- The scaling function $\varphi_{j,k}$ is characterized by having only the coefficient $\lambda_{j,k}$ set to 1; all the others are null.
- Hence, starting from such a sequence, after few iterations of the cascade algorithm a good sampling of the scaling function is obtained.
  - The number of coefficients doubles at each iteration.

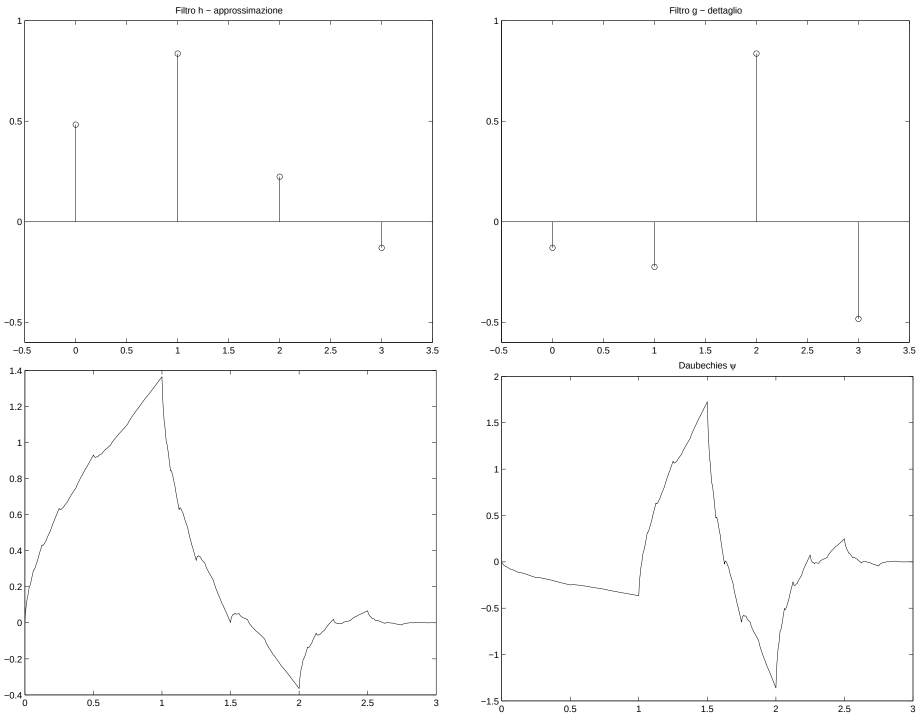
.

Plotting the basis functions (2)

- Similarly, the wavelet can be obtained.
  - All the λ and γ are set to 0, but one of γ is set to 1.
- The Fourier transforms of wavelet and scaling functions can also be obtained.

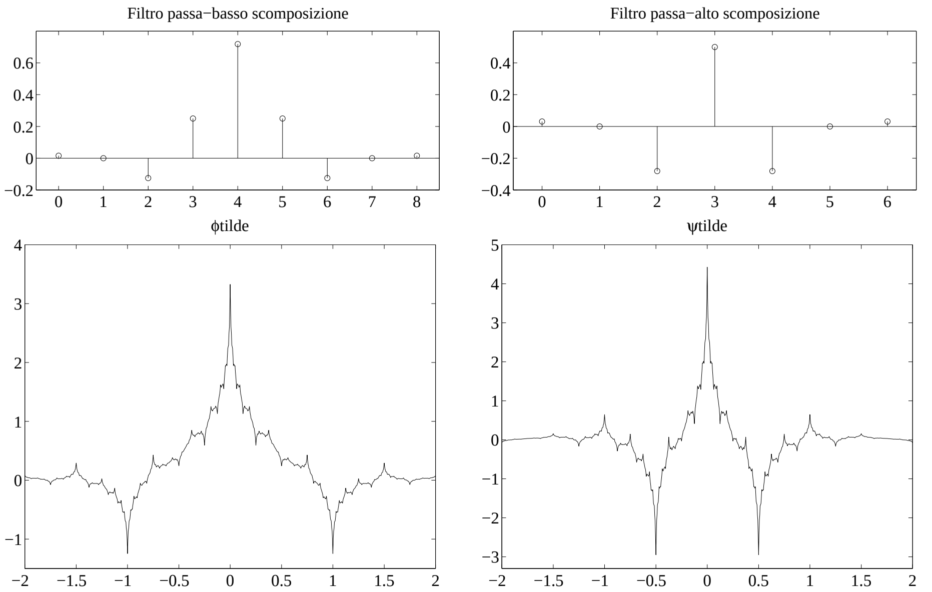
Orthogonal basis

Daubechies

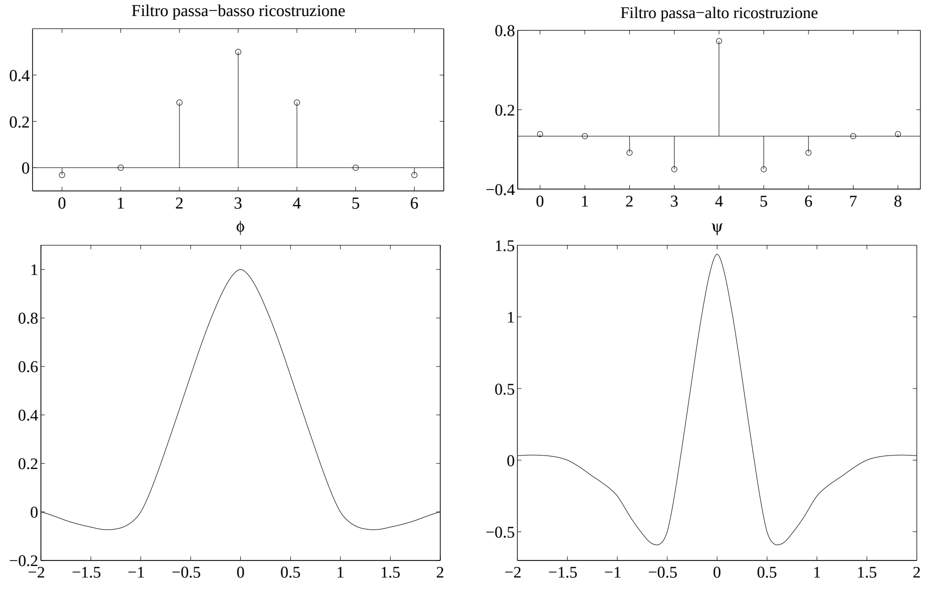


.

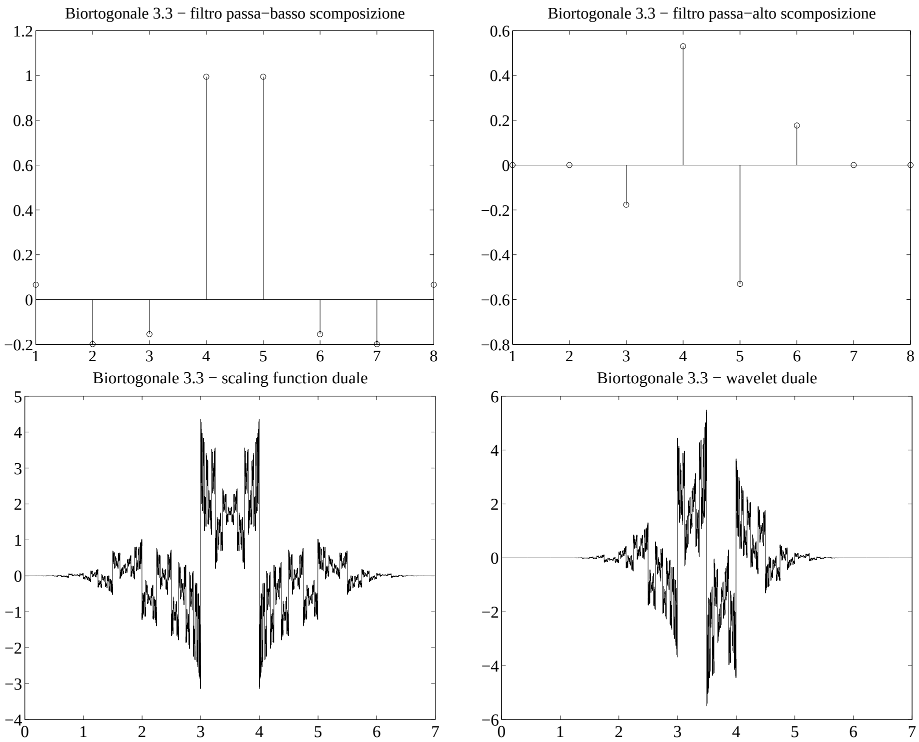
Biorthogonal basis (dual)



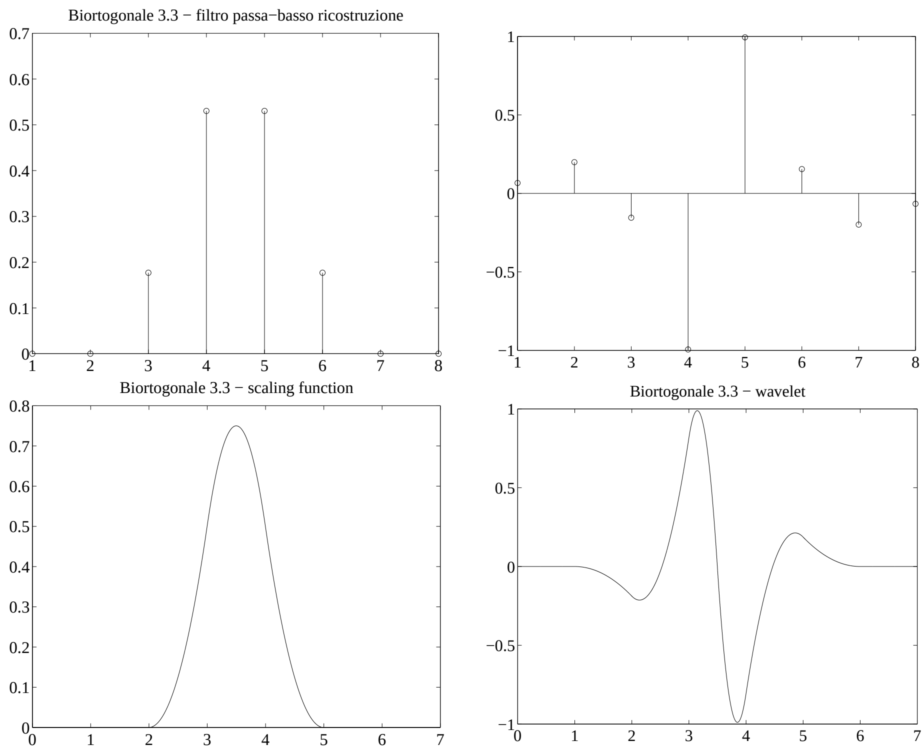
Biorthogonal basis (primal)



Biorthogonal basis (dual) (2)

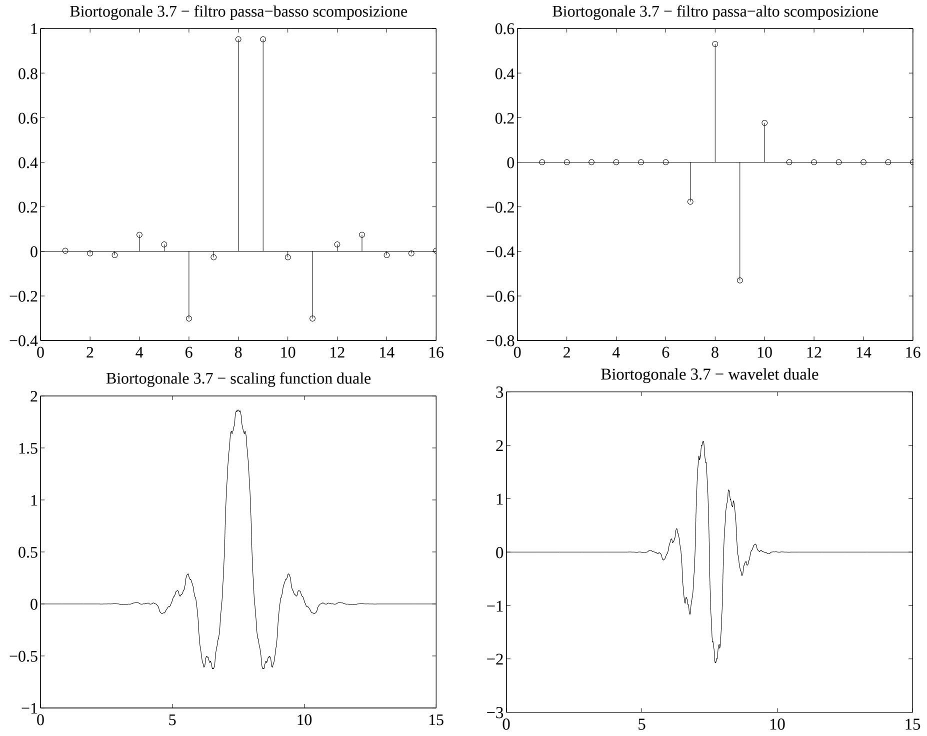


Biorthogonal basis (primal) (2)

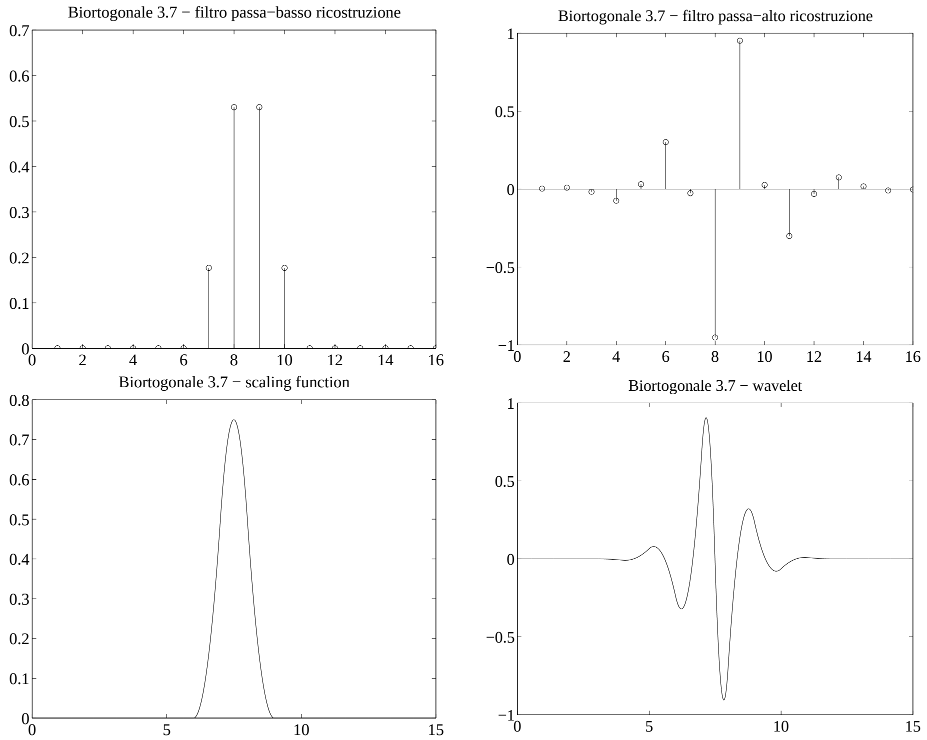


.

Biorthogonal basis (dual) (3)



Biorthogonal basis (primal) (3)



Fast Wavelet Transform (FWT)

orthogonality or biorthogonality

$\downarrow$

fast algorithm for wavelet transform computation

function projection	$\rightarrow$	FIR filters filtering

$$f \rightarrow \boxed{\begin{matrix} g \\ h \end{matrix}} \begin{matrix} \rightarrow \{\lambda_{j,k}\} \\ \rightarrow \{\gamma_{j,k}\} \end{matrix}$$

transform

$$\begin{matrix} \{\lambda_{j,k}\} \rightarrow \\ \{\gamma_{j,k}\} \rightarrow \end{matrix} \boxed{\begin{matrix} g \\ h \end{matrix}} \rightarrow f$$

inverse transform

orthogonal MRA

$$f \rightarrow \boxed{\begin{matrix} \tilde{g} \\ \tilde{h} \end{matrix}} \begin{matrix} \rightarrow \{\lambda_{j,k}\} \\ \rightarrow \{\gamma_{j,k}\} \end{matrix}$$

transform

$$\begin{matrix} \{\lambda_{j,k}\} \rightarrow \\ \{\gamma_{j,k}\} \rightarrow \end{matrix} \boxed{\begin{matrix} g \\ h \end{matrix}} \rightarrow f$$

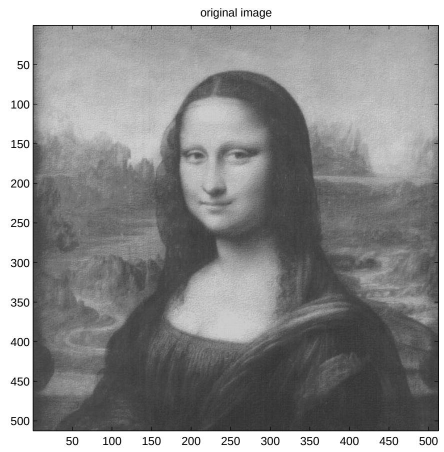
inverse transform

biorthogonal MRA

Applications to images

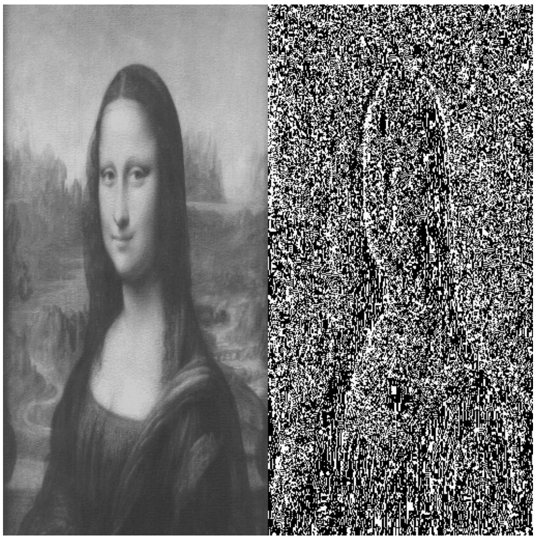
- Wavelet and scaling function can be defined also on a bidimensional domain, by using the tensor product.
  - They are defined as the product over the two dimensions.
- Hence they can be applied to the two dimensions independently.
  - Like the Fourier transform.

Example:

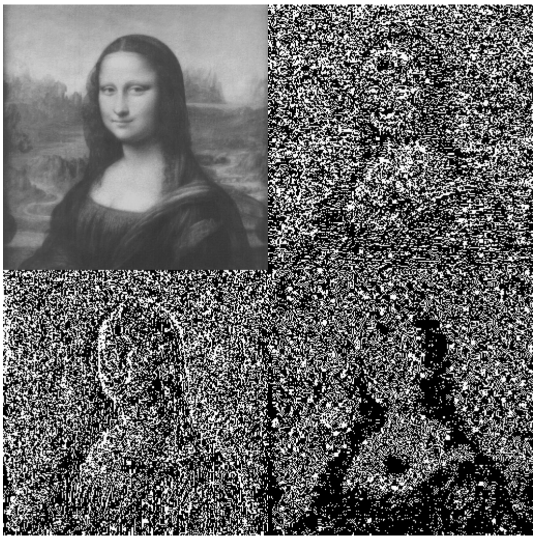


.

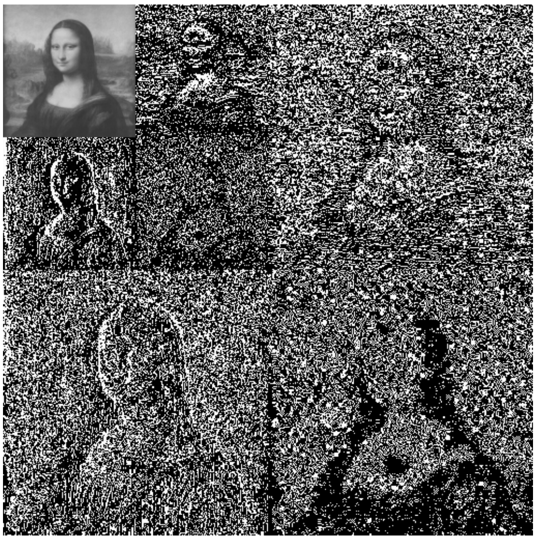
Applications to images (2)



Applications to images (3)



Applications to images (4)



Applications

- Signal representation (e.g., compression)
- Signal processing (e.g., filtering, anomalies detection)
- Pattern recognition (e.g., for feature selection)
- Hybrid models (e.g., Wavelet neural networks)

Image compression

Wavelet based image compression algorithms are based on some considerations:

- small detail coefficients (probably) carry unimportant information or noise;
  - if a detail occurs, the coefficients of all the levels corresponding to its position should be meaningful;
  - thresholding is used to set to zero unimportant coefficients;
  - quantization and encoding (e.g. Huffman) can then be realized.
- the shorter the wavelet support, the smaller the number of non-zero coefficients generated by an edge;
- orthogonality (and biorthogonality) decorrelates the coefficients.

Image compression (2)

- $f_M = \sum_{j.k} b_{j,k}\psi_{j,k}(x)$ with M non-zero coefficients, $b_{j,k}$
- From the orthogonality, the reconstruction error is:

$$||f - f_M||_{L^2} = \left(\sum_{j,k} |\langle f, \psi_{j,k}\rangle - b_{j,k}|^2 \right)^{\frac{1}{2}}$$

- Hence, the larger the $b_{j,k}$'s, the smallest the error.
- Besov space characterization allows a better estimate of the compression rate wrt. M.

.

Image compression (3)

- Encoding can take advantage of long sequences of zeros.
- The scanning order of the coefficients is critical for maximizing the length of zeros sequences.
- If a coefficient is zero, also the corresponding coefficients at the higher scales are probably zero.

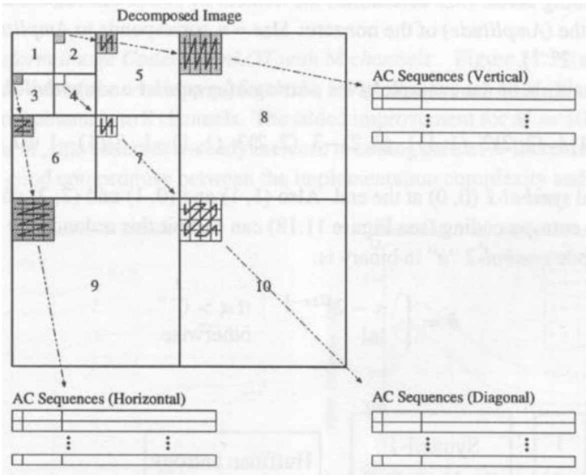


Image compression (4)

- Compression of image sequences can be realized using the 3D wavelet transform.
- Quality can be improved by considering not only a single coefficients, but the value of the coefficients in a neighborhood of each position.

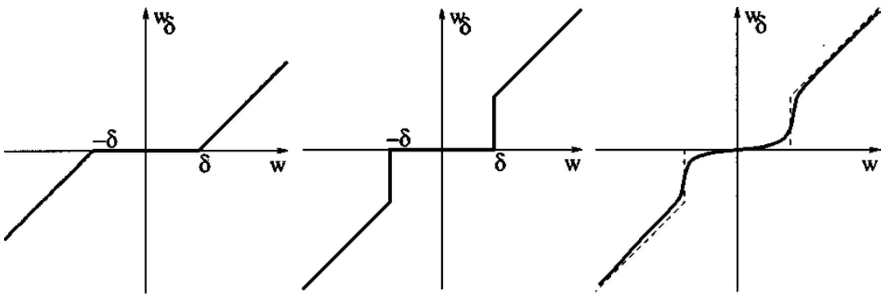
Image denoising

Wavelet denoising is based on three assumptions.

- Additive, stationary, and zero-mean noise affects the coefficients of all resolution levels.
- Large coefficients describe a good approximation of the original image.
- Noise should be relatively small.
  - Small influence on the large coefficients.

Image denoising – shrinking

- Shrinking is the approach generally used for denoising:
  - a threshold for each level and component is chosen;
  - the coefficients under threshold are set to zero.



- Soft and hard thresholding, and a sophisticated shrinking function.
- Soft thresholding is often used.

Image denoising – threshold selection

- For a given level and component (horizontal, vertical, diagonal), the optimal threshold should optimize (MSE):

$$\frac{1}{N}||w_\delta - v||^2$$

  where:
  - w_δ are the coefficients after shrinking
  - v are the unknown noise-free coefficients

- The Donoho and Johnstone threshold:

$$\delta = \sqrt{2 \log N}\sigma$$

  where:
  - N is the number of coefficients
  - σ is the noise standard deviation

Image denoising – threshold selection (2)

- Generalized cross validation can be used for estimating the threshold, by minimizing:

$$\text{GCV}(\delta) = \frac{\frac{1}{N}||w - w_\delta||^2}{\left(\frac{N_0}{N}\right)^2}$$

  where:
  - N_0 is the number of zero coefficients
- It mimics the MSE criterion.
- No estimate for the noise energy, σ, is needed.

- Adaptive techniques for estimating δ from the data can be found in literature.

.

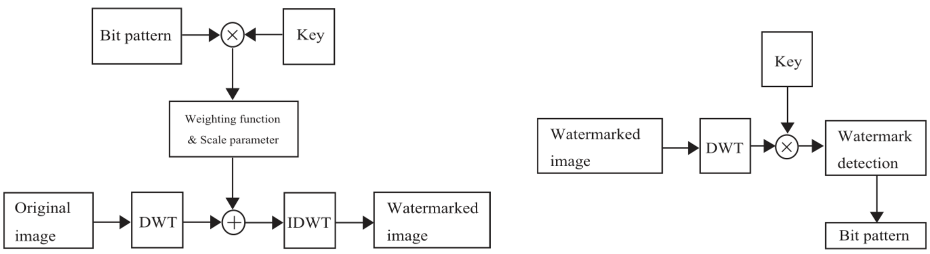
Image denoising – neighboring

- Correlation between neighboring coefficients can be exploited:
  1. compute $s_{j,k} = \sum_{t \in \mathcal{N}(k)} w_{j,t}$
  2. shrink $w_{j,k}$:

$$w_{j,k} = \begin{cases} 0, & s_{j,k} < \delta \\ w_{j,k}(1 - \delta/s_{j,k}), & \text{otherwise} \end{cases}$$

Image watermarking

- Wavelet coefficients can be perturbed in order to insert a watermark.
- The key can than be used in detecting the presence of the watermark.



.

Image fusion

- Wavelet coefficients of registered images can be averaged.

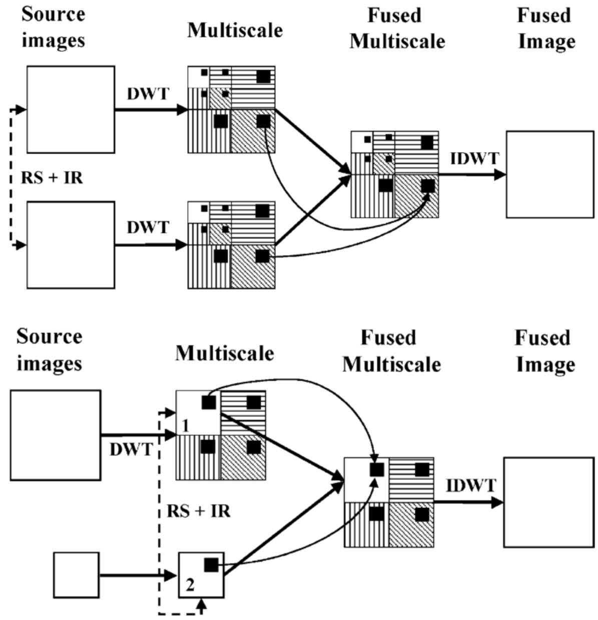
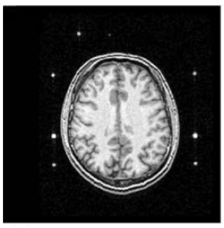
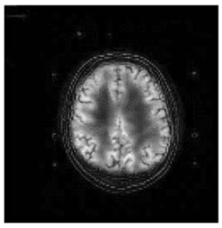


Image fusion (2)

- Example: MRI and PET images of the same subject.





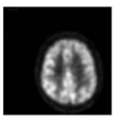
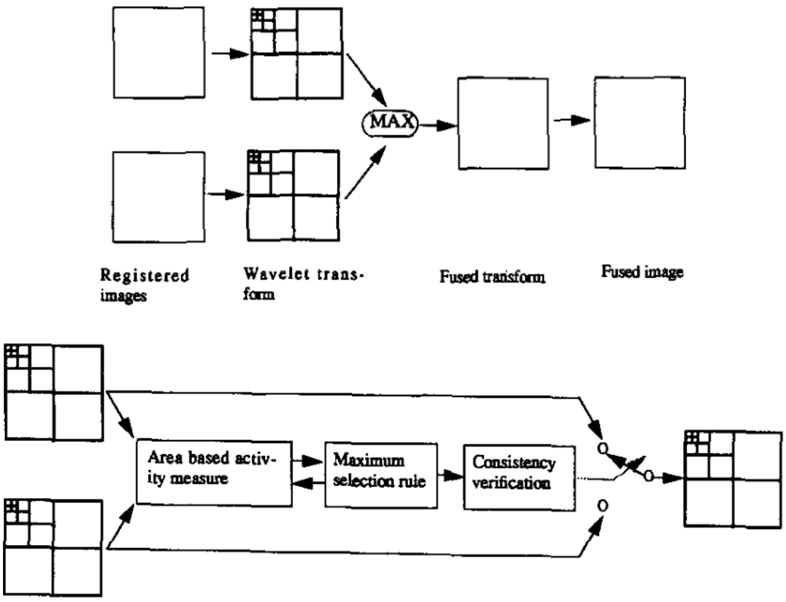


Image fusion (3)

- A suitable rule for averaging have to be devised.



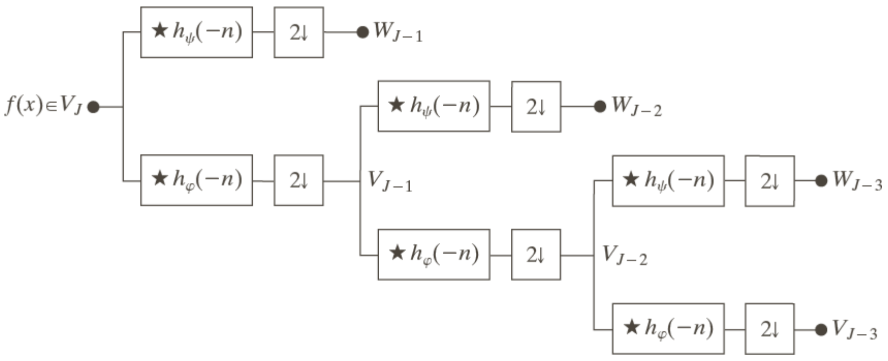
Modern wavelets

- Wavelets packets
- Lifting schema

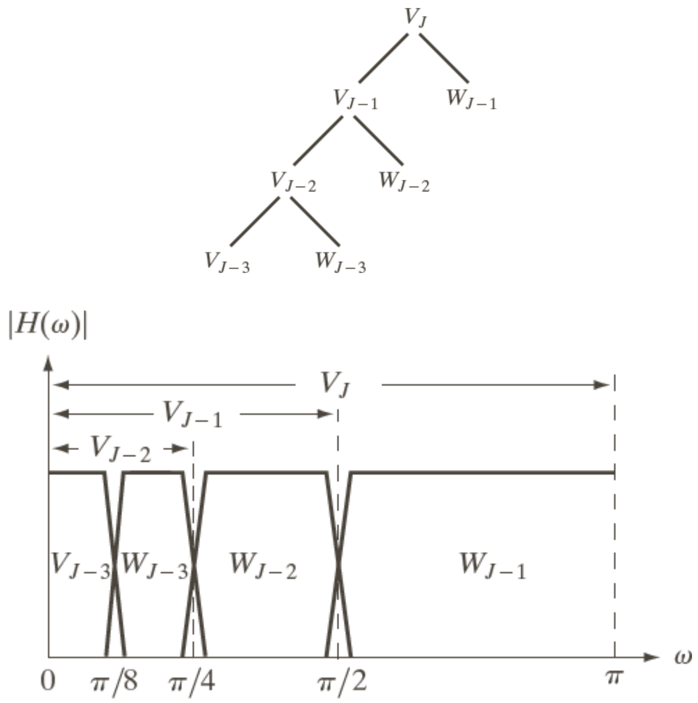
.

Wavelet packets

- FWT provides a decomposition of a signal f in element of several subspaces (with $O(N)$ computational cost).



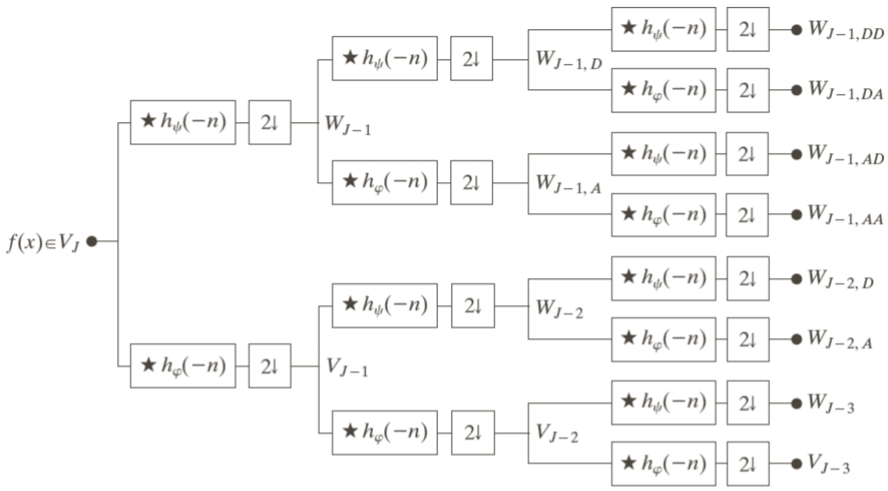
Wavelet packets (2)



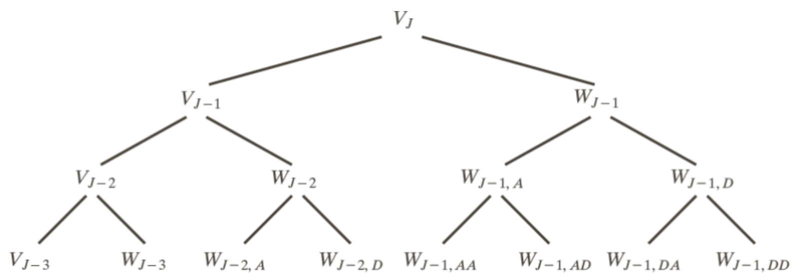
.

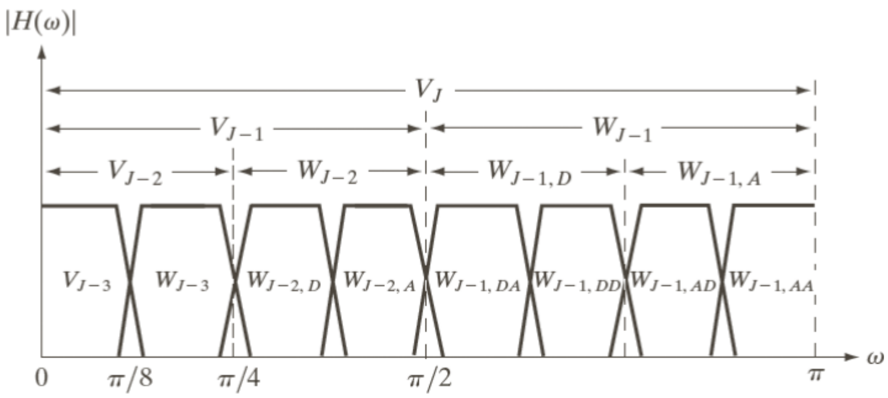
Wavelet packets (3)

- FWT machinery can be extended for decomposing also the detail coefficients.
- This transforms is called *wavelet packet* (and have an $O(N \log N)$ cost).

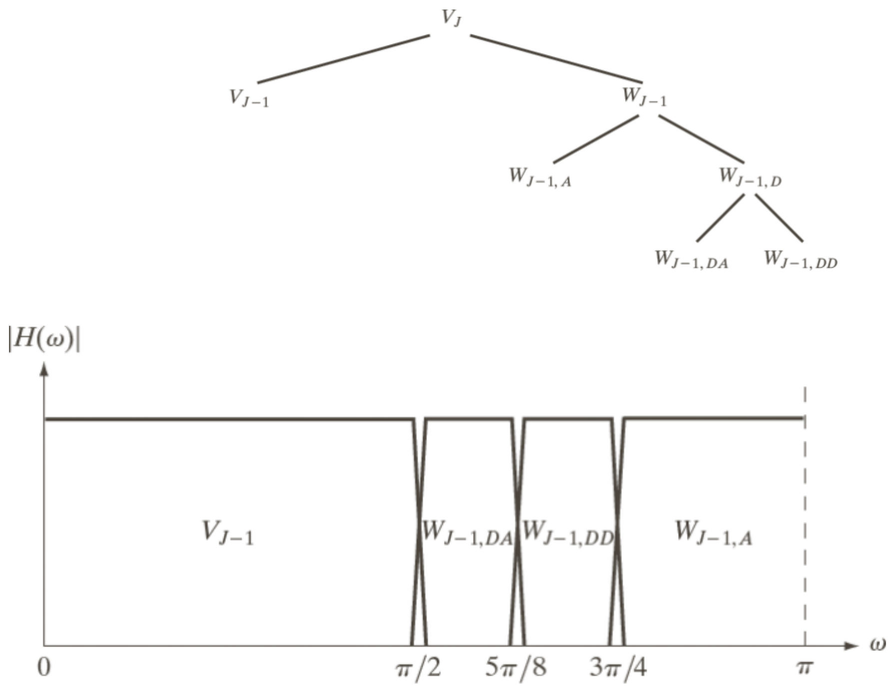


Wavelet packets (4)



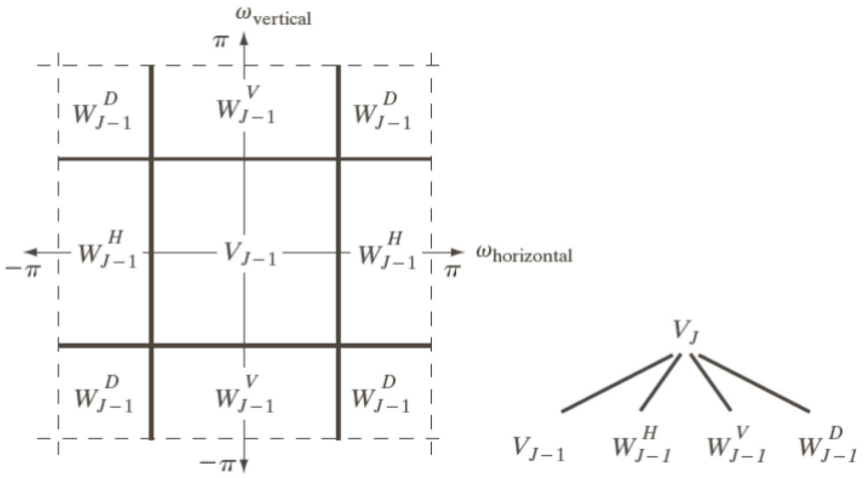


Wavelet packets (5)



Wavelet packets for images

- For modeling the effects of a n-dimensional wavelet packet transform, a 2^n-ary tree can be considered.



Wavelet packets – optimal decomposition

- For the FBI fingerprint archive, a three scales wavelet packets based compression is used.
  - The complete decomposition yields to 64 coefficients sets.

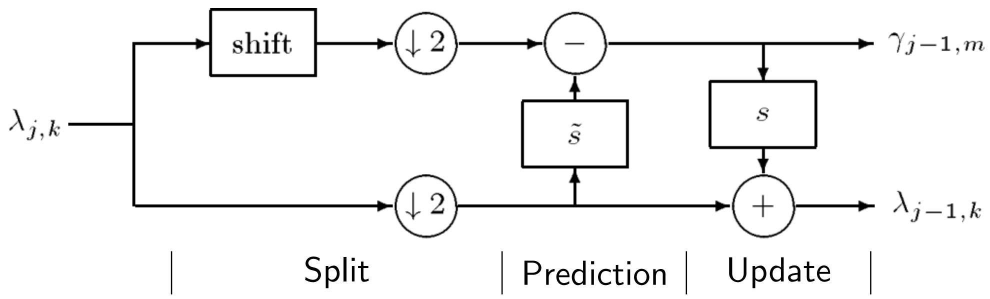

Wavelet packets – optimal decomposition (2)

- In order to optimize the storage requirement, the optimal decomposition (best basis selection) can be considered.


Lifting scheme

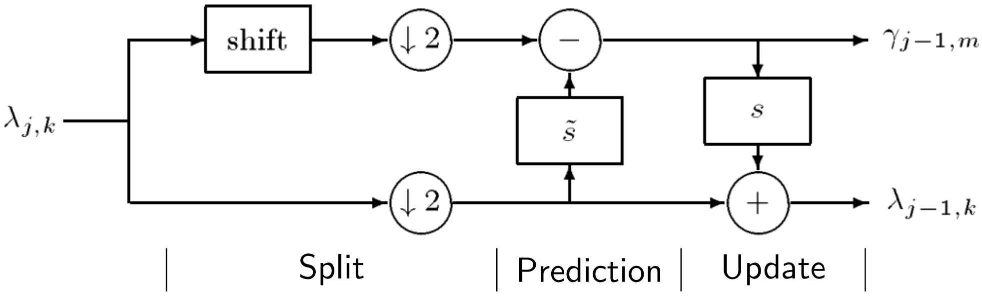
- The lifting scheme is a method for constructing the so-called second generation wavelets (orthogonal and biorthogonal):
  - they do not make use of Fourier transform (no regularly spaced samples are required);
  - they are not necessarily translates and dilates of the same function.
- Lifting scheme (LS) has the following advantages:
  - Faster implementation of the wavelet transform
    - FWT processes the same sequence with two filters and then subsample both the sequences;
    - LS splits the sequence before processing.
  - In-place processing (no additional memory requirement).
  - Inverse transform is realized inverting the transform operations.

Lifting scheme – analysis



- Split: $\{\lambda_{j,k}\}$ is split in $\{\lambda_{j-1,k}\}$ and $\{\gamma_{j-1,k}\}$;
  - the split can be done with any rule, but even and odd samples partition is a sensible choice.
- Prediction: $\{\lambda_{j-1,k}\}$ is used to predict $\{\gamma_{j-1,k}\}$ through $\tilde{s}$:
  - the value in the two sequence should be correlated;
  - this information is used to change the values of $\{\gamma_{j-1,k}\}$.

Lifting scheme – analysis (2)

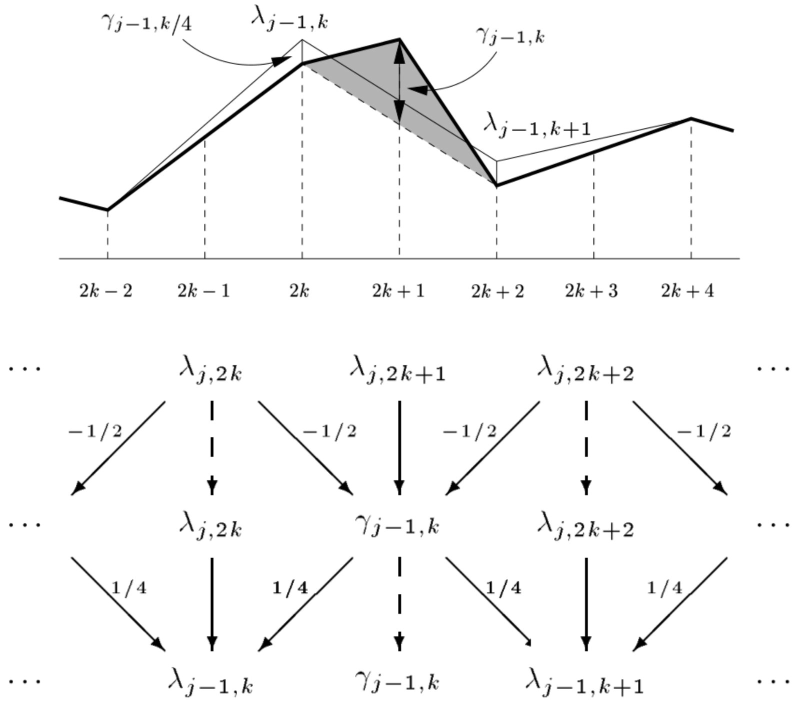


- Update: information in the original $\{\gamma_{j-1,k}\}$ that cannot be predicted by $\{\lambda_{j-1,k}\}$ is now in $\{\gamma_{j-1,k}\}$; this can be used to update the value of $\{\lambda_{j-1,k}\}$ through s:
  - downsampling can suffer of aliasing;
  - $\{\lambda_{j-1,k}\}$ can now preserve some features of $\{\lambda_{j,k}\}$ (e.g., the mean, or up to the n-th moment);
  - an ad-hoc operator could be hardly invertible.

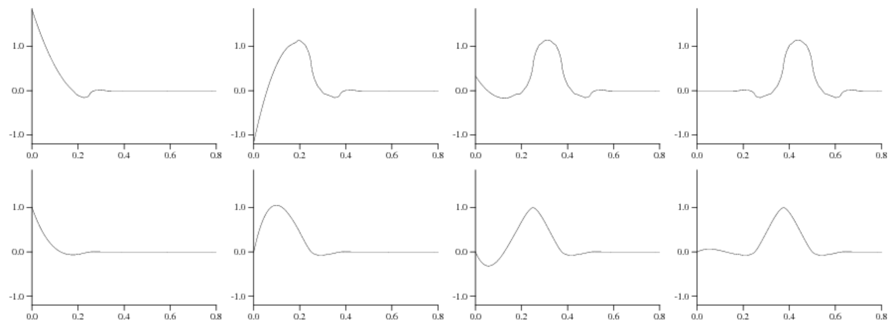
Lifting scheme – synthesis

- Since the analysis stage can be realized as:
  1. $[\{\lambda_{j-1,k}\}, \{\gamma_{j-1,k}\}] := \text{split}(\{\lambda_{j,k}\})$
  2. $\{\gamma_{j-1,k}\} := \{\gamma_{j-1,k}\} - \tilde{s}(\{\lambda_{j-1,k}\})$
  3. $\{\lambda_{j-1,k}\} := \{\lambda_{j-1,k}\} + s(\{\gamma_{j-1,k}\})$
- the synthesis stage can be obtained as:
  1. $\{\lambda_{j-1,k}\} := \{\lambda_{j-1,k}\} - s(\{\gamma_{j-1,k}\})$
  2. $\{\gamma_{j-1,k}\} := \{\gamma_{j-1,k}\} + \tilde{s}(\{\lambda_{j-1,k}\})$
  3. $\{\lambda_{j,k}\} := \text{join}(\{\lambda_{j-1,k}\}, \{\gamma_{j-1,k}\})$

Lifting scheme – linear

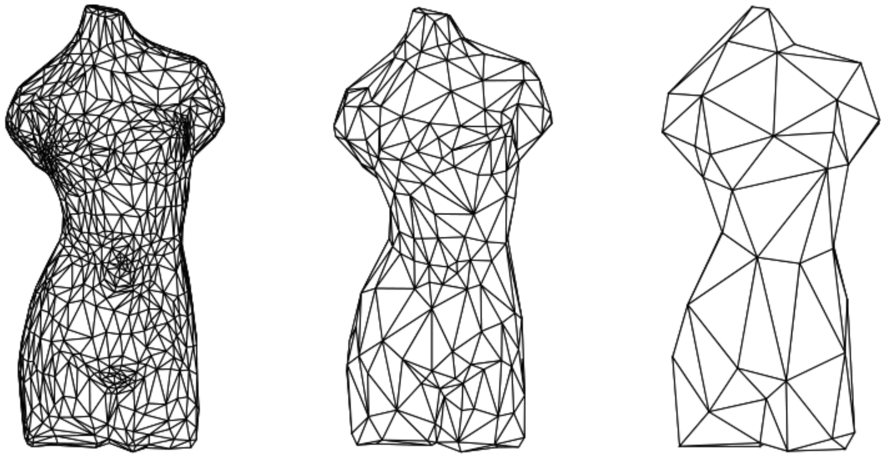


Lifting scheme – boundary



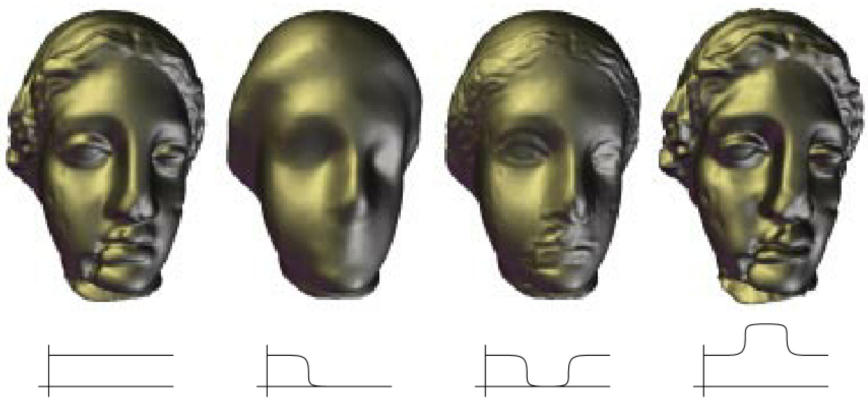
- Scaling function at boundary (quadratic and cubic spline)

Lifting scheme – irregular sampling



- Second generation wavelets can be applied on irregular sampled data.
- Low scale approximation of mesh produce a coarsification.

Lifting scheme – mesh processing



- Typical signal processing techniques are possible also on meshes:
  - original, smoothed, stop-banded, enhanced.

.

Other multiscale approaches

- Curvelets
- Beamlets
- Tetrolets
- Ridgelets

Bibliography

- *Digital image processing*, R.C. Gonzalez and R.E. Woods
- *Digital image processing*, B. Jähne
- "Wavelets on Irregular Point Sets," I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens
- "The lifting scheme: a construction of second generation wavelets," W. Sweldens, SIAM Journal on Mathematical Analysis
- "Image de-noising by integer wavelet transforms and generalized cross validation," M. Jansen, G. Uytterhoeven, and A. Bultheel, Medical Physics, Vol. 26, No. 4, April 1999
- "Image denoising with neighbour dependency and customized wavelet and threshold," G.Y. Chen, T.D. Bui, A. Krzyzak, Pattern Recognition, 2005
- "Image denoising with an optimal threshold and neighbouring window," Zhou Dengwen, Cheng Wengang, Pattern Recognition Letters, 2008

.

Bibliography (2)

- "Genetic Algorithm Approach for Wavelet-Based Image Watermarking," M. Nakamoto, S. Fujimoto, A. Doi, T. Hinamoto, ICSP2008 Proceedings
- "A wavelet-based image fusion tutorial," G. Pajares, J. Manuel de la Cruz, Pattern Recognition, 2004
- "Multisensor Image Fusion Using the Wavelet Transform," H. Li, S. Manjunath, and S.K. Mitra, Graphical Model and Image Processing, 1995
- "A Review of Curvelets and Recent Applications," J. Ma and G. Plonka
- "Tetrolet Shrinkage with Anisotropic Total Variation Minimization for Image Approximation," J. Krommweha, J. Ma
- "The Finite Ridgelet Transform for Image Representation," M.N. Do and M. Vetterli, IEEE Transactions on Image Processing