Institut
Mines-Telecom

Wavelet denoising

Marco Cagnazzo,
cagnazzo@telecom-paristech.fr

TSIA-SD205

Outline

Outline

Introduction

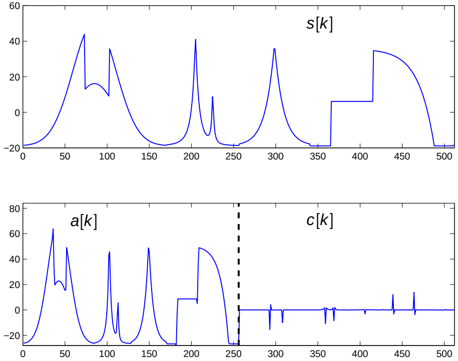
- Wavelet transform
- Principles of denoising

Denoising

Filterbanks 1D

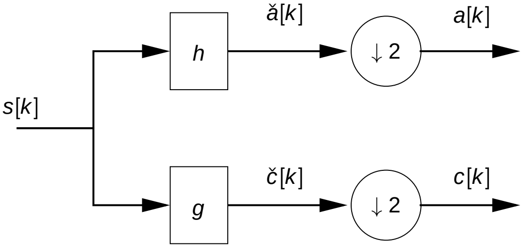
Characteristics of wavelet transform

- Energy concentration
- Contours representation
- Multiple resolution analysis
  - Low resolution version
  - “Details”



Filterbanks 1D

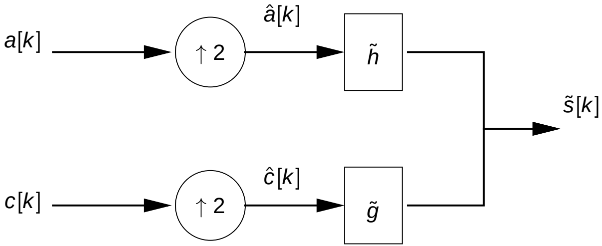
Decomposition



Analysis Filterbank

$\downarrow 2$: decimation: $a[k] = \check{a}[2k]$

Reconstruction



Synthesis Filterbank

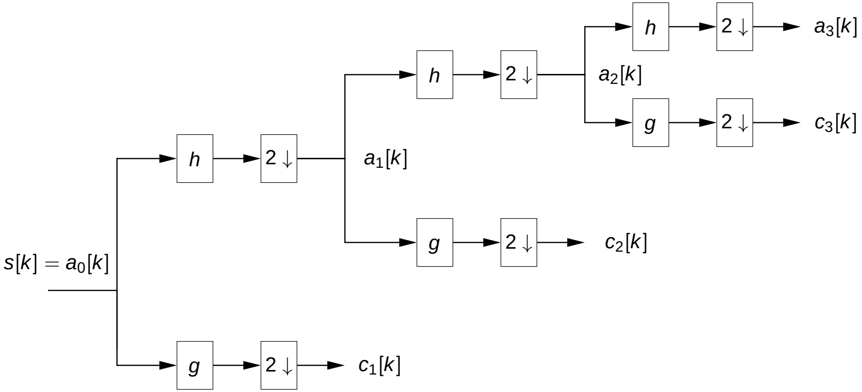
$\uparrow 2$: interpolator, doubles the number of samples

$$\hat{a}[k] = \begin{cases} a[k/2] & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}$$

Properties of the filters

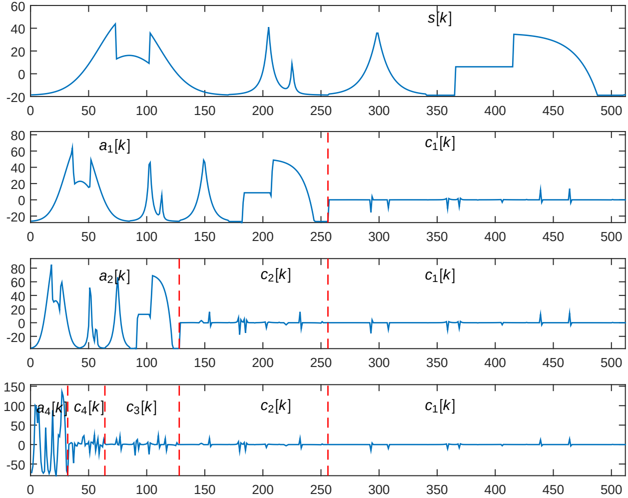
- Perfect reconstruction
  - It is possible to reconstruct the signal from its coefficients
- Finite Impulse Response
  - Finite operation implementation
- Orthogonality
  - Coefficients' energy equal to signal energy
- Vanishing moments
  - Capacity of reproducing polynomial signals with zero details
- Symmetry
  - Implementation by symmetrization and periodization

Multiresolution Analysis 1D

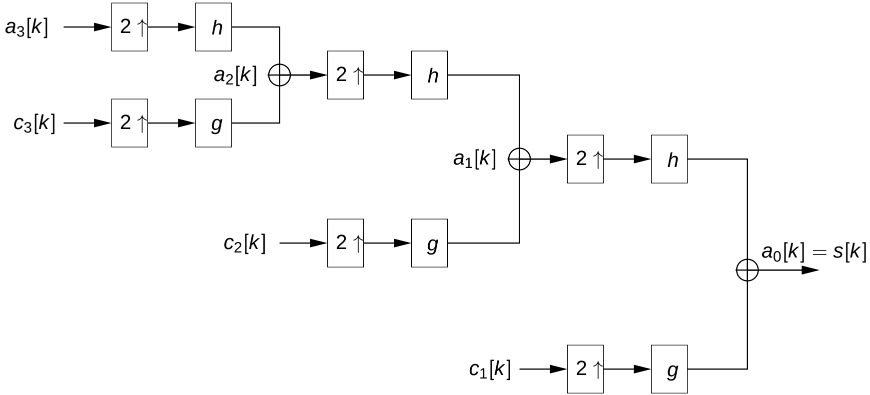


Structure of wavelets decomposition with 3 levels of resolution

Multiresolution Analysis 1D



Multiresolution synthesis 1D

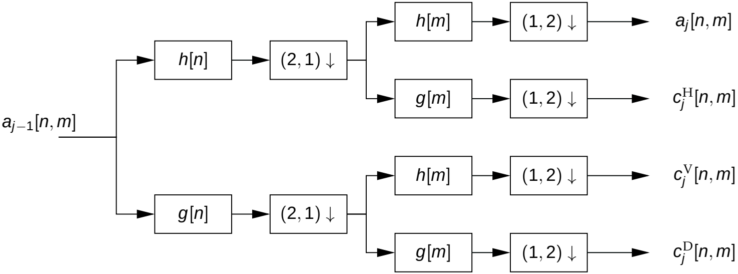


Reconstruction from the wavelet coefficients

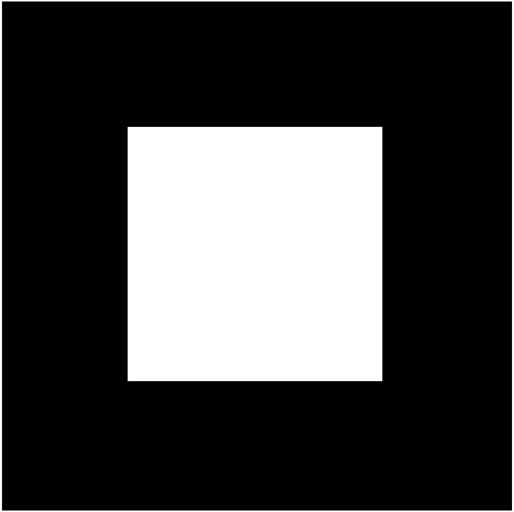
2D MRA

2D separable filterbanks

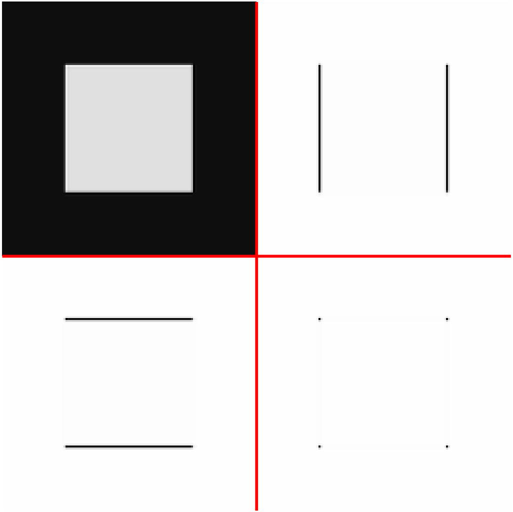
1 level of decomposition



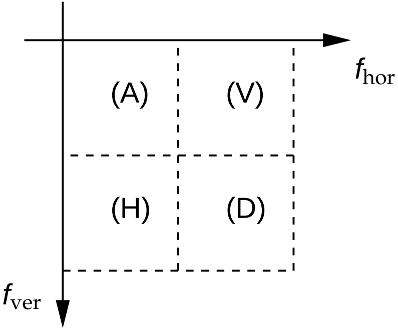
Example



Example

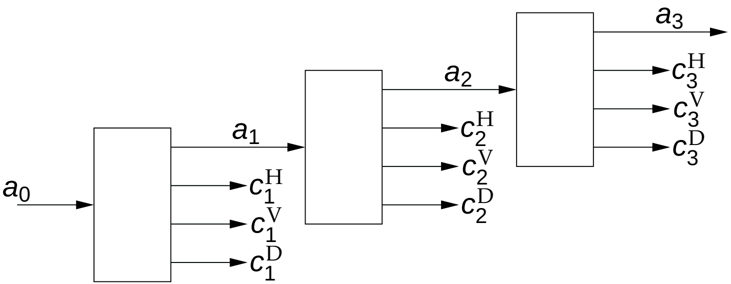


Frequency Interpretation



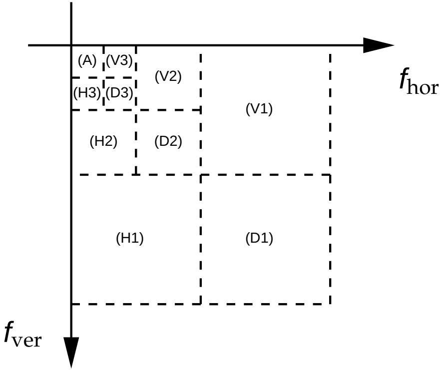
(A), (H), (V) and (D) areas correspond to the approximation subband, the horizontal, vertical and diagonal detail subbands

2D MRA

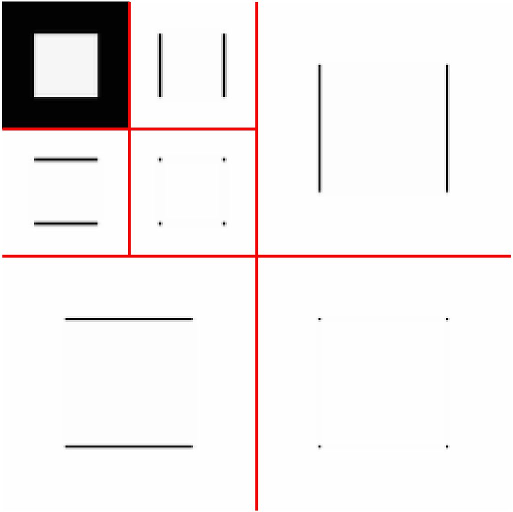


2D separable MRA with 3 resolution levels

AMR 2D - frequency interpretation

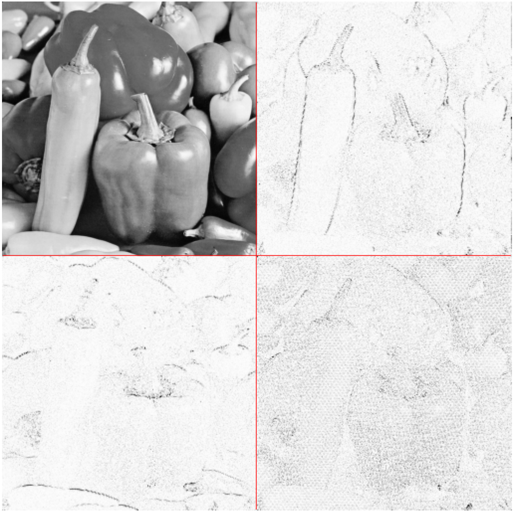


Example

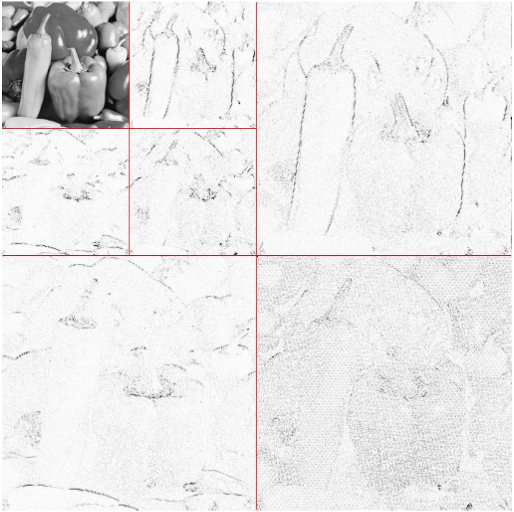


Example

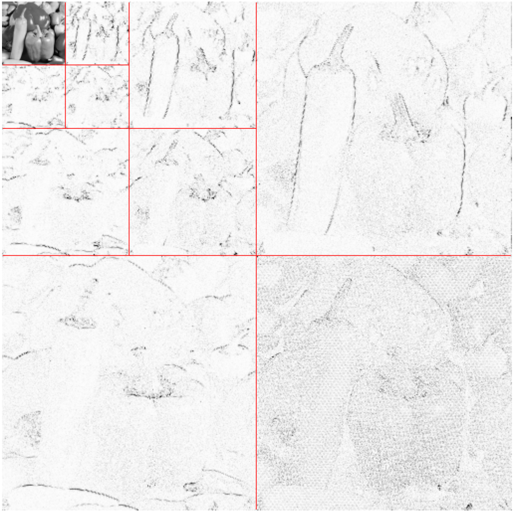
Example



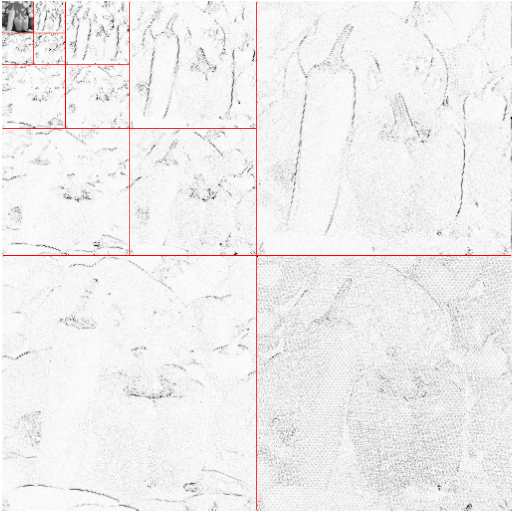
Example



Example



Example



Denoising

Principles

Model: Observation: $r(t)$; sum of an unknown signal $s(t)$ and random noise $b(t)$.
After decomposition on a wavelet basis :

$$c_j^r[k] = c_j^s[k] + c_j^b[k]$$

Hypothesis:

- Orthonormal basis, periodical decomposition
- Original signal (resolution $j = 0$) has size multiple of $2^{j_{\max}}$
- High MSE in the approximation subband: $a^s_{j_{\max}} \approx a^r_{j_{\max}}$

Denoising

Principles

- $c_j^r[k] = c_j^s[k] + c_j^b[k]$
- High MSE in the approximation subband: $a_{j_{\max}}^s \approx a_{j_{\max}}^r$
- **Estimator**: $\hat{s}$
- **Criterion**: MSE minimization: $\mathcal{E}^2(s) = \mathrm{E}\{\|s - \hat{s}\|^2\}$
- Without denoising, $\hat{s} = r$

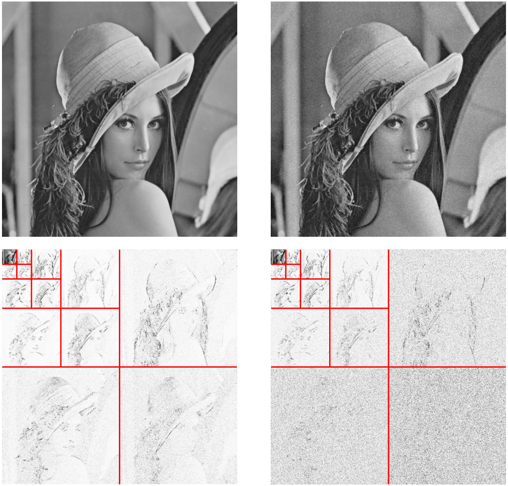
$$\mathcal{E}^2(s) = \mathrm{E}\{\|s - r\|^2\} = \mathrm{E}\{\|c^s - c^r\|^2\} = \mathrm{E}\{\|c^b\|^2\} = K_m \sigma^2$$
$$K_m = (1 - 2^{j_{\max}})K$$

Denoising

Principles

- Regular signal
  - Energy concentrated in low frequencies
  - Sparse signal in high frequencies
  - Many very small coefficients
  - A few large coefficients (information!)
- Noise is often white and stationary
  - Model: white, stationary, centered and with power σ^2
  - Noise power is equally shared among subbands
- What do we find in the subbands?

Examples

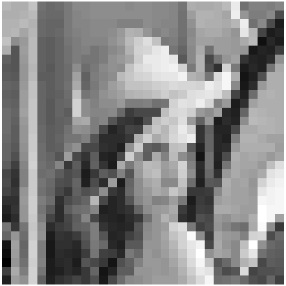
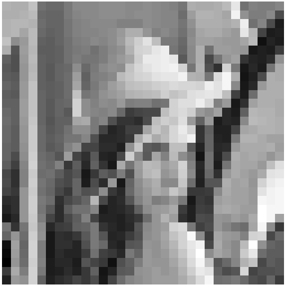


SNR: 22.4 dB;

$\sigma = 10$

Examples

Approximation subband

SNR: 46.4 dB

Examples

Detail subband

SNR: 15.2 dB

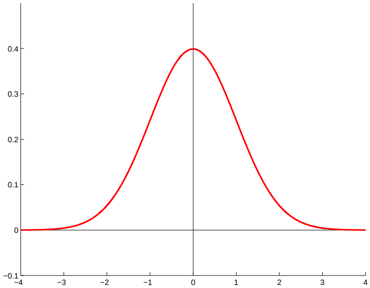
Noise variance estimation

Hypothesis:

- In the highest resolution subband the coefficients come only from the noise:

$$\{c_1^s[k]\}_{0\leq k<K/2} \approx 0$$

- Zero-mean Gaussian noise, i.i.d.



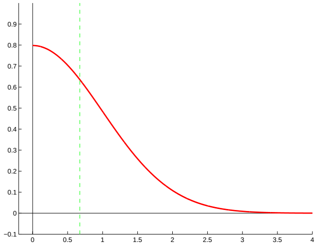
Noise variance estimation

We consider the distribution of $|Z|$, when Z is normal.

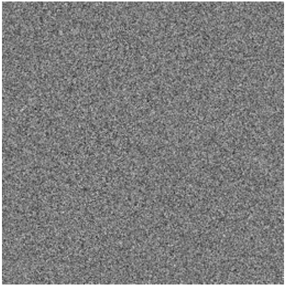
We know that the median of $|Z|$ is $0.6745\,\sigma$.

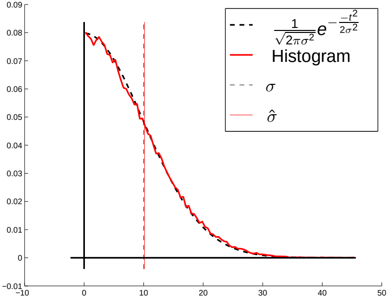
then we choose:

$$\hat{\sigma} = \frac{1}{0.6745} \mathrm{Med}|c_1^r|$$



Noise variance estimation





Outline

Attenuation estimator

Definition:

$$c_j^{\hat{s}}[k] = \theta_j[k] c_j^r[k]$$

MSE:

$$\mathcal{E}_a^2(s) = \sum_{j=1}^{j_{\max}} \sum_{k=0}^{K2^{-j}-1} \mathrm{E}\left[\left(c_j^s[k] - \theta_j[k] c_j^r[k]\right)^2\right]$$

$$(c^s - \theta c^r)^2 = (c^s(1-\theta) - \theta c^b)^2$$

$$\mathcal{E}_a^2(s) = \sum_{j=1}^{j_{\max}} \sum_{k=0}^{K2^{-j}-1} (c^s)^2(1-\theta)^2 + \sigma^2\theta^2$$

Attenuation estimator

General term of the previous sum:

$$J = (c^s)^2(1 - 2\theta + \theta^2) + \sigma^2\theta^2$$

Minimizing wrt θ:

$$\frac{\partial J}{\partial \theta} = -2(c^s)^2 + 2\theta((c^s)^2 + \sigma^2)$$

Then:

$$\theta^* = \frac{(c^s)^2}{(c^s)^2 + \sigma^2}$$

Attenuation estimator

Oracle

$$\theta^* = \frac{(c^s)^2}{(c^s)^2 + \sigma^2}$$

Oracle: the estimator depends on the signal. Useful for theoretical bounds

$$J = (c^s)^2(1-\theta)^2 + \sigma^2\theta^2 \qquad = \frac{\sigma^2 (c^s)^2}{\sigma^2 + (c^s)^2}$$

$$\mathcal{E}_a^2(s) = \sum_{j=1}^{j_{\max}} \sum_{k=0}^{K2^{-j}-1} \frac{\sigma^2 (c_j^s[k])^2}{\sigma^2 + (c_j^s[k])^2}$$

Attenuation estimator

Binary oracle

- If we constrain θ to be binary: $\theta_j[k] \in \{0, 1\}$, then
  - $J = (c^s)^2$ if $\theta = 0$; otherwise $J = \sigma^2$
  - then we choose $\theta = 0$ if $(c^s)^2 < \sigma^2$
- In this case, the MSE is:

$$\mathcal{E}_o^2(s) = \sum_{j=1}^{j_{\max}} \sum_{k=0}^{K2^{-j}-1} \min\left[\sigma^2, (c_j^s[k])^2\right]$$

Attenuation estimator

Binary oracle

$$\mathcal{E}_o^2(s) = \sum_{j=1}^{j_{\max}} \sum_{k=0}^{K2^{-j}-1} \min\left[\sigma^2, (c_j^s[k])^2\right]$$

$$\mathcal{E}_a^2(s) = \sum_{j=1}^{j_{\max}} \sum_{k=0}^{K2^{-j}-1} \frac{\sigma^2 (c_j^s[k])^2}{\sigma^2 + (c_j^s[k])^2}$$

$$0 < x \leq y \Rightarrow \frac{xy}{x+y} \geq \frac{xy}{2y} = \frac{1}{2}x = \frac{1}{2}\min(x,y)$$

$$\frac{1}{2}\mathcal{E}_o^2(s) \leq \mathcal{E}_a^2(s)$$

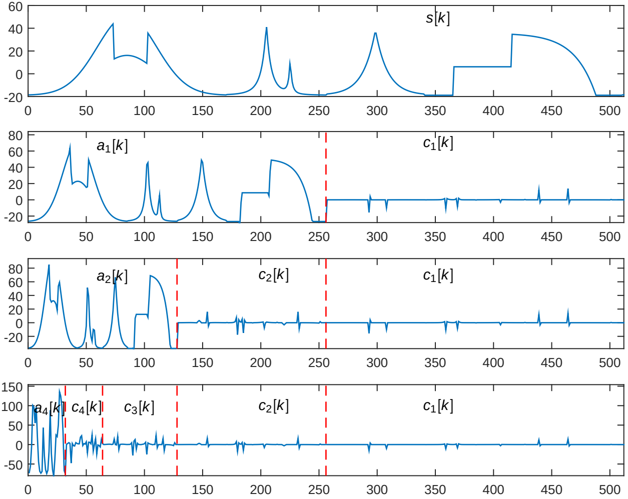
$$\mathcal{E}_a^2(s) \leq \mathcal{E}_o^2(s) \leq 2\mathcal{E}_a^2(s)$$

Attenuation estimator

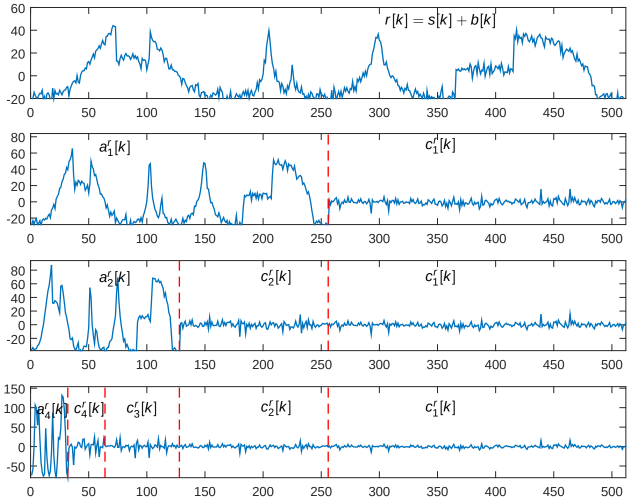
Binary oracle

- The MSE of the binary oracle is at most the double of the attenuation oracle
- Conclusion: We keep the wavelet coefficients where the signal can be supposed to be large, and we set to zero the others
- Simplified model: there are Q non-zero coefficients c^s and they are greater than σ; the others are zeros
- The binary oracle has in this case an MSE of $Q\sigma^2$
- Without denoising the MSE is $K_m\sigma^2$, where $K_m = K(1 - 2^{-j_{\max}})$ is the number of available wavelet coefficients

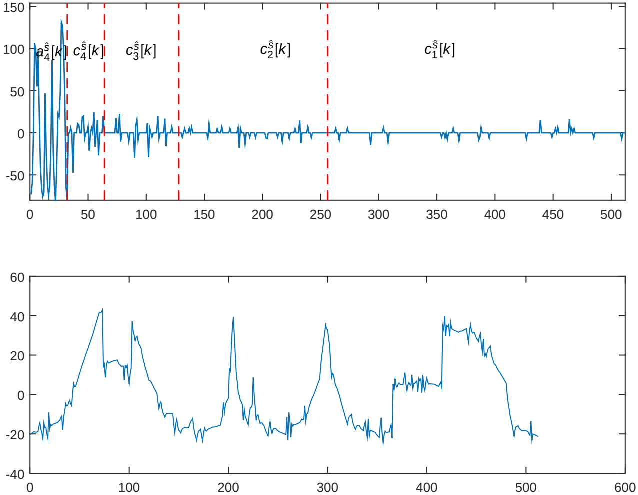
Attenuation estimator



Attenuation estimator



Attenuation estimator



Attenuation estimator

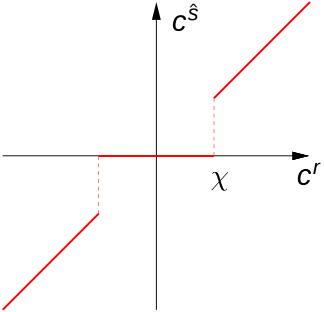
Binary oracle

- In conclusion, the binary oracle allows to reduce error by a factor:

$$\frac{K_m}{Q}$$

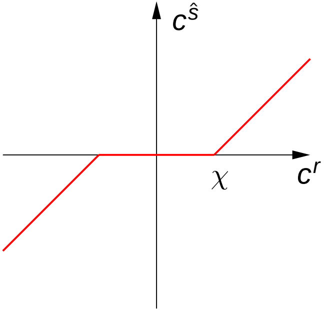
- The, a *good* wavelet basis is one making Q as small as possible
- The wavelet basis must generate a few large coefficients and many small ones
- In other words, it should concentrate energy in as few coefficients as possible

Thresholding



Hard Thresholding

$$c^{\hat{s}} = \begin{cases} c^r & \text{if } |c^r| > \chi \\ 0 & \text{if } |c^r| \leq \chi \end{cases}$$



Soft Thresholding

$$c^{\hat{s}} = \begin{cases} c^r - \chi & \text{if } c^r > \chi \\ 0 & \text{if } |c^r| \leq \chi \\ c^r + \chi & \text{if } c^r < \chi \end{cases}$$

Thresholding

- Hard thresholding is not continuous near the threshold $\pm\chi$
- Soft thresholding introduces a bias $\mp\chi$ on the estimation of the large coefficients
- Main problem: **value of the threshold** χ
  - *Minimax* method
  - *Visushrink* method (universal threshold)
  - *SURE* method
  - *Hybrid* method

Minimax method

Definitions:

- $K_m = K(1 - 2^{-j_{\max}})$ is the number of available wavelet coefficients
- $\mathcal{E}_\chi$ is the MSE associated to the soft thresholding with threshold χ
- $\tilde{\mu}$ is a symmetric probability density and $\bar{\mu}$ its normalized version (variance 1)

Hypothesis: The noise wavelets coefficients $c_j^b[k]$ have all the same symmetric PDF $\tilde{\mu}$ (of variance σ^2)

Then

- There is an equation giving χ_m, the threshold minimizing the maximum MSE over signals s for the soft thresholding case.
- This MSE can be related to the binary oracle case

Minimax method

$$\inf_{\chi \geq 0} \sup_{s} \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^2 + \mathcal{E}_{o}^{2}(s)} = \sup_{s} \frac{\mathcal{E}_{\chi_m}^{2}(s)}{\sigma^2 + \mathcal{E}_{o}^{2}(s)}$$

$$= \Lambda_{\chi_m} = \frac{K_m(\chi_m^2 + \sigma^2)}{(K_m + 1)\sigma^2} \ ,$$

where χ_m is the unique solution in $\mathbb{R}_+$ of the equation

$$2(K_m + 1) \int_{\chi}^{\infty} (z - \chi)^2 \tilde{\mu}(z) dz = \chi^2 + \sigma^2 \ . \qquad (1)$$

Minimax method

Solution of the equation (1):

1. The noise probability density is normalized:

$$\bar{\mu}(z) = \sigma\tilde{\mu}(\sigma z)$$

2. The equation is solved for the normalized threshold $\bar{\chi}_m$

$$2(K_m + 1)\int_{\bar{\chi}_m}^{\infty}(z - \bar{\chi}_m)^2\bar{\mu}(z)dz = \bar{\chi}_m^2 + 1 \tag{2}$$

3. We find the threshold as: $\chi_m = \sigma\bar{\chi}_m$

Minimax method

In the Gaussian case, the equation (2) becomes:

$$\frac{1}{2}\,\mathrm{erf}\left(\frac{\bar{\chi}_m}{\sqrt{2}}\right) + \frac{\bar{\chi}_m}{\bar{\chi}_m + 1}\frac{1}{\sqrt{2\pi}}e^{-\frac{\bar{\chi}_m^2}{2}} = \frac{K_m}{2(1+K_m)}.$$

Table of numerical solutions
Is it reasonable to suppose a Gaussian PDF fore the WT coefficients?

- Yes if the original noise was Gaussian
- Otherwise, other laws can be used (different laws for different levels)

Asymptotical value of the optimal threshold

If the noise wavelet coefficients have PDF:

$$\forall z \in \mathbb{R}, \qquad \mu(z) = Ce^{-h(z)}$$

where $C \in \mathbb{R}_+^*$ and h is a even continuous function, strictly increasing on $\mathbb{R}_+$ and such that

$$\lim_{z \to \infty} z^{-\beta} h(z) = \gamma \in \mathbb{R}_+^* \ , \qquad \beta \geq 1$$

and

$$\forall (z_1, z_2) \in \mathbb{R}_+^2, \quad h(z_1 + z_2) \geq h(z_1) + h(z_2) \ .$$

When $K_m \to \infty$,

$$\chi_m \sim \chi_{\mathrm{U}} = h^{-1}(\ln K_m)$$

$$\Lambda_{\chi_m} \sim \frac{\chi_{\mathrm{U}}^2}{\sigma^2} + 1 \ .$$

Universal threshold

- The previous hypotheses hold for Gaussian and generalized Gaussian distributions
- In the both cases we have $\mu(z) = Ce^{-h(z)}$
- Gaussian:

$$\mu(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}} \qquad h(z) = \frac{z^2}{2\sigma^2}$$

$$h^{-1}(t) = \sqrt{2\sigma^2 t} \qquad \chi_U = \sigma\sqrt{2\ln K_m}$$

- GG:

$$\mu(z) = Ce^{-\gamma|z|^\beta} \qquad h(z) = \gamma|z|^\beta$$

$$\chi_U = \left(\frac{\ln K_m}{\gamma}\right)^{\frac{1}{\beta}}$$

Universal threshold

- It eases the computation of the universal threshold in the Gaussian case
  - It is easier than χ_m
- The difference between χ_U and χ_m can be large if K_m is not large
- Nevertheless, χ_U is called *universal threshold*
- Using the universal threshold is referred to as *visushrink*

Universal threshold

Minimax vs. universal threshold (Gaussian case)

K_m	χ_m	Λ_{χ_m}	χ_U	Λ_U
8	0.877372	1.573139	2.039334	5.158883
16	1.076456	2.031772	2.354820	6.545177
32	1.276276	2.549217	2.632769	7.931472
64	1.474135	3.124256	2.884054	9.317766
128	1.668605	3.754906	3.115134	10.704061
256	1.859020	4.438616	3.330218	12.090355
512	2.044916	5.171582	3.532230	13.476649
1024	2.226161	5.949982	3.723297	14.862944
2048	2.402888	6.770567	3.905027	16.249238
4096	2.575057	7.629058	4.078668	17.635532
8192	2.742753	8.521655	4.245212	19.021827
16384	2.906252	9.445722	4.405465	20.408121
32768	3.065703	10.398216	4.560089	21.794415
65536	3.221205	11.375990	4.709640	23.180710
131072	3.373025	12.377200	4.854586	24.567004
262144	3.521304	13.399528	4.995328	25.953299

Threshold optimality

In the Gaussian case, the threshold estimators are asymptotically *minimax* :
If the noise coefficients are i.i.d. $\mathcal{N}(0, \sigma^2)$ then

$$\lim_{K_m \to \infty} \inf_{\hat{s}} \sup_{s} \left(\frac{\mathcal{E}^2(s)}{\left(\sigma^2 + \mathcal{E}_o^2(s)\right) \Lambda_{\chi_m}} \right) = 1 \ .$$

where $\inf_{\hat{s}}$ is the infimum on the set of possible estimators of $s(t)$.
Asymptotically, threshold estimators are the best in the minimax sense

SURE method

Principles

- Minimax method is too pessimistic
- Idea: estimate the MSE in the average case, and then minimize it
- Problem: the estimation of the MSE depending on the signal
- Stein's lemma: allows to estimate the MSE *without bias*

SURE method

Principles

- We observe the r.v. $Y = x + Z$
- x, deterministic, is the quantity we want to measure
- Z, whose standard deviation is σ, is the noise
- We consider an estimator $T(Y) = Y + \gamma(Y)$
- The risk (the MSE) is:

$$\begin{aligned}
\epsilon^2(x) &= \mathrm{E}\{(x - T(Y))^2\} \\
&= \mathrm{E}\{(x - Y - \gamma(Y))^2\} \\
&= \sigma^2 - 2x\mathrm{E}[\gamma(Y)] + 2\mathrm{E}[Y\gamma(Y)] + \mathrm{E}[\gamma(Y)^2]
\end{aligned}$$

Problem: the risk depends on x!

SURE method

Stein's lemma

We observe the r.v. $Y = x + Z$ with x deterministic and $Z \sim \mathcal{N}(0, \sigma^2)$; if γ is a continuous function, piecewise derivable and such that, for all $x \in \mathbb{R}$,

$$\lim_{|z|\to\infty} \gamma(z) \exp\left(-\frac{(z-x)^2}{2\sigma^2}\right) = 0$$

$$\mathrm{E}\{\gamma(x+Z)^2\} < \infty, \quad \mathrm{E}\{|\gamma'(x+Z)|\} < \infty$$

then

$$x\mathrm{E}\{\gamma(Y)\} = \mathrm{E}\{Y\gamma(Y)\} - \sigma^2\mathrm{E}\{\gamma'(Y)\} \qquad (3)$$

SURE method

Stein's lemma

- The conditions over $\gamma(\cdot)$ are quite general
- E.g., they hold for non-linear continuous and piecewise derivable function with polynomial or slower increase
- This means that it exists $m \in \mathbb{N}$ and $A \in \mathbb{R}_+$ such that:

$$\forall z \in \mathbb{R}, \ |\gamma(z)| \leq A\,|t|^m\,.$$

SURE method

Stein's lemma

Applying the Stein's lemma, we can write the equation (3):

$$\begin{aligned}\epsilon^2(x) &= \sigma^2 + \mathrm{E}[\gamma(Y)^2] + 2\sigma^2\mathrm{E}[\gamma'(Y)] \\ &= \mathrm{E}[J(Y)]\end{aligned}$$

where

$$J(Y) = \sigma^2 + 2\sigma^2\gamma'(Y) + \gamma^2(Y)$$

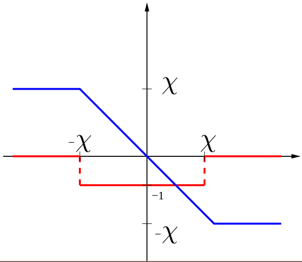
This is true also if $Y = X + Z$ with X r.v. independent from Z

SURE method

Application to wavelet coefficients

- The wavelet coefficients of the signal are r.v. with finite variance, and for a given level j, i.i.d.
- The noise coefficients , $c_j^b[k]$, are i.i.d., $\mathcal{N}(0, \sigma_j^2)$ independent from the signal
- We use soft thresholding, which verifies the hypothesis of Stein's lemma, with

$$\gamma(z) = \begin{cases} \chi & \text{if } z \leq -\chi, \\ -z & \text{if } |z| \leq \chi, \\ -\chi & \text{if } z \geq \chi \end{cases}$$



SURE method

Application to wavelets

Estimation of the MSE: $\epsilon_j^2(x) = \mathrm{E}[J(c_j^r[k])]$ with:

$$J(z) = \sigma_j^2 + 2\sigma_j^2\gamma'(z) + \gamma^2(z)$$
$$= \begin{cases} z^2 - \sigma_j^2 & \text{if } |z| \leq \chi \\ \chi^2 + \sigma_j^2 & \text{if } |z| > \chi \end{cases}$$

and finally:

$$\hat{\epsilon}_j^2(x) = \frac{1}{2^{-j}K} \sum_{k=0}^{K2^{-j}-1} J(c_j^r[k])$$

We only have to find the threshold χ minimizing $\hat{\epsilon}_j^2$

SURE method

Algorithm

We sort the wavelet coefficients:

$$A = |c_j^r[0]| \geq |c_j^r[1]| \geq \ldots \geq |c_j^r[K2^{-j}-1]| = B$$

and We consider the three cases: $\chi > A$, $A \geq \chi \geq B$, and $\chi < B$.

- In the first, the MSE does not depend on χ
- In the second, $\exists k_0$ such that $|c_j^r[k_0| \leq \chi < |c_j^r[k_0-1|$, then:

  $$2^{-j}K\hat{\epsilon}_j^2 = k_0\chi^2 + (2k_0 - K2^{-j})\sigma_j^2 + \sum_{k=k_0}^{K2^{-j}-1}(c_j^r[k])^2$$

  and the minimum is attained for $\chi = |c_j^r[k_0]|$
- In the third, $\hat{\epsilon}_j^2 = \chi^2 + \sigma^2$ and the minimum is for $\chi = 0$

SURE method

Algorithm

- In conclusion the optimal value of χ is among:

$$\{|c_j^r[0]|, |c_j^r[1]|, \ldots, |c_j^r[K2^{-j} - 1]|, 0\}$$

- An exhaustive search can be carried off.
- The risk can b computed with a recurrent equation
- Total complexity: $O(2^{-j}K)$ for the search and $O(2^{-j}K\log(2^{-j}K))$ for sorting
- Advantage: threshold automatically adapted to data

SURE method

Hybrid threshold

- If the signal power, at a given resolution level is too small with respect to the noise, the SURE estimator is not reliable
- Then we use for that level the universal threshold
- Estimator of the signal power:

$$\overline{(c_j^s)^2} = \frac{1}{K2^{-j}} \sum_{k=0}^{K2^{-j}-1} (c_j^r[k])^2 - \sigma_j^2$$

- Critical power level:

$$\lambda_{j,K} = \frac{\sigma_j^2}{\sqrt{(K2^{-j})}} \left[\ln(K2^{-j})\right]^{3/2}$$

SURE method

Hybrid threshold

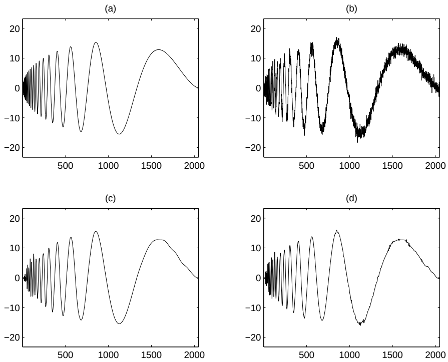
In conclusion we use the hybrid threshold:

$$\chi_{j,\mathrm{H}} = \begin{cases} \chi_{j,\mathrm{SURE}} & \text{if } \overline{(c_j^s)^2} > \lambda_{j,K} \\ \chi_{j,\mathrm{U}} & \text{otherwise} \end{cases}$$

the universal threshold is:

$$\chi_{j,\mathrm{U}} = \sigma_j \sqrt{2 \ln K 2^{-j}}$$

Example of denoising/debruit



Original signal (a), noisy, SNR = 18.86 dB (b), After universal threshold, SNR = 23.80 dB (c), denoised with *sureshrink*, SNR = 27.45 dB (d).

Bayes method

- We observe $Y = X + Z$, X r.v. with probability density p_X and Z r.v. with probability density μ, independent from X
- Having an observation y of the r.v. Y, we select the most probable value of X:

$$\hat{x} = \arg\max_x p_X(x|Y = y)$$

It is the **MAP** estimator (Maximum *A Posteriori* probability)

Bayes method

- MAP Estimator

$$\hat{x} = \arg\max_x p_X(x|Y=y)$$

- Using Bayes rules:

$$p_X(x|Y=y) = \frac{p_Y(y|X=x)p_X(x)}{p_Y(y)} = \frac{\mu(y-x)p_X(x)}{p_Y(y)}$$

- The MAP Estimator is then equivalent to:

$$\hat{x} = \arg\min_x [-\ln(\mu(y-x)) - \ln(p_X(x))]$$

Most of reasonable of the *a priori* give thresholding estimators

Bayes method

Hypotheses

- Noise and signal mutually independent
- Their wavelet coefficients are independent r.v.
- At resolution level j, the noise $\mathcal{N}(0, \sigma_j^2)$ and the signal coefficients are i.i.d. with pdf p_j
- If the wavelet basis fits well the signal, we expect $c_j^s[k]$ very small with high probability, and large with low probability

Bayes method

Laplacian PDF

$$p_j(u) = \frac{1}{\sqrt{2}\eta_j} \exp\Big(-\frac{\sqrt{2}|u|}{\eta_j}\Big)$$

It can be shown that
The MAP estimator corresponding to a zero-mean Laplacian PDF with standard deviation $\eta_j > 0$ and with noise $\mathcal{N}(0, \sigma_j^2)$ is the soft thresholding with threshold $\chi_{j,\mathrm{B}} = \sqrt{2}\sigma_j^2/\eta_j$.

Bayes method

Generalized Gaussian (GG) PDF

$\mathcal{GG}(\alpha_j, \beta_j)$, $(\alpha_j, \beta_j) \in (\mathbb{R}_+^*)^2$, with:

$$p_j(u) = \frac{\beta_j}{2\alpha_j \Gamma(1/\beta_j)} \exp\left(-\frac{|u|^{\beta_j}}{\alpha_j^{\beta_j}}\right)$$

where Γ the gamma function.
If $\beta_j \leq 1$, the MAP estimator for $\mathcal{GG}(\alpha_j, \beta_j)$ and noise $\mathcal{N}(0, \sigma_j^2)$ is a threshold estimator with

$$c_j^{\hat{s}}[k] = 0 \Longleftrightarrow |c_j^r[k]| \leq \chi_{j,\mathrm{B}}$$

où

$$\chi_{j,\mathrm{B}} = \frac{2-\beta_j}{2(1-\beta_j)} \left(\frac{2\sigma_j^2(1-\beta_j)}{\alpha_j^{\beta_j}}\right)^{1/(2-\beta_j)} .$$

Bayes method

Generalized Gaussian (GG) PDF

When

$$|c_j^r[k]| \geq \chi_{j,\mathrm{B}} \qquad \text{where} \qquad \beta_j > 1$$

the estimation of a coefficient is a *shrinkage* of the observed value

If $\beta_j < 1$, the estimator is rather close to the hard thresholding, since it is not continuous near the threshold

Bayes method

Bernoulli-Gaussian PDF

$q_j[k]$: hidden random variables binary, independent and such that each component $c_j^s[k]$ of $s(t)$ is:

- carrying information, if $q_j[k] = 1$: $P(q_j[k] = 1) = \epsilon_j$
- zero, if $q_j[k] = 0$

When $q_j[k] = 1$, we suppose that $c_j^s[k]$ is Gaussian, zero-mean, with variance σ_j^2

Bayes method

Bernoulli-Gaussian PDF

Signal estimation

Maximum A Posteriori Estimator of $q_j[k]$:

$$\hat{q}_j[k] = \begin{cases} 1 & \text{if } |c_j^r[k]| > \chi_{j,\mathrm{B}}, \\ 0 & \text{otherwise} \end{cases}$$

threshold $\chi_{j,\mathrm{B}} \geq 0$:

- depends on σ^2, σ_j^2 and ϵ_j
- independent from the signal length

$$c_j^{\hat{s}}[k] = \begin{cases} \dfrac{\sigma_j^2}{\sigma^2 + \sigma_j^2} c_j^r[k] & \text{if } \hat{q}_j[k] = 1 \\ 0 & \text{otherwise} \end{cases}$$

Bayes method

Determination of the *a priori* model parameters

The main problem is to find the parameters (ϵ_j and σ_j^2)

- iterative methods (generalized likelihood method, EM algorithm, MCMC, ...)
- performance/complexity trade-off