Statistical Signal Processing
A (Very) Short Course

Episode IV: Deconvolution, Wavelets and related problems

Thomas Trigano[1]

[1]Hebrew University
Department of Statistics

February 13, 2007

Outline

Outline

First examples
Real-life situations

- In a crowd speaking in Hebrew in HUJI, all the buzz sounds like noise...
- ... Yet, if someone speaks in French in this crowd next to me, I will only listen to that... at least before starting Hebrew lessons...
- In this example, the buzz would be (stationary) noise and the French (as always) should be local (transitory) information.

What does it mean ? One of the main assumptions we used all the time until now, stationary, does not cover all the ground

First examples
Real-life situations

- In a crowd speaking in Hebrew in HUJI, all the buzz sounds like noise...
- ... Yet, if someone speaks in French in this crowd next to me, I will only listen to that... at least before starting Hebrew lessons...
- In this example, the buzz would be (stationary) noise and the French (as always) should be local (transitory) information.

What does it mean ? One of the main assumptions we used all the time until now, stationary, does not cover all the ground

Transition phenomena and signal processing

Some examples of applications

- Speech and audio processing: how to “karaoke” ? How to isolate nonstationary harmonics ?
- Sismology: High-frequency modulated impulsions ?
- Image processing: How to isolate a pattern in an image ?
- Signal: how to denoise a signal corrupted by a stationary noise ?

What we will look at on this course

and also what we will not talk about

- Recalling a few properties of the Fourier transform and show why this extremely powerful tool isn't the panacea.
- Introduce some time-frequency tools: local Fourier transform and wavelets.
- Study some applications for denoising and deconvolution.
- All the compression aspects related to wavelets will only be introduced here.
- We will see some examples on images, but mainly image processing using wavelets will be skipped.

Let us now detail the examples we will study more in detail.

What we will look at on this course

and also what we will not talk about

- Recalling a few properties of the Fourier transform and show why this extremely powerful tool isn't the panacea.
- Introduce some time-frequency tools: local Fourier transform and wavelets.
- Study some applications for denoising and deconvolution.
- All the compression aspects related to wavelets will only be introduced here.
- We will see some examples on images, but mainly image processing using wavelets will be skipped.

Let us now detail the examples we will study more in detail.

Signal and Image denoising
The problem

Consider a deterministic function f being monodimensional (signal) or bidimensional (image), and assume that we observe a noisy version of f:

$$y(t) = f(t) + \varepsilon(t) \ ,$$

where ε is a noise function. The problem of finding f given y and some information about the noise is called *signal (image) denoising*

Density Deconvolution

The problem

Instead of considering a deterministic function f, we may replace it by the realisation of a random variable X; the problem is now, given a series of observations:

$$Y_k = X_k + \varepsilon_k \ , \ k = 1 \dots n$$

find information on the distribution of X. This is called a *density deconvolution problem*.

Signal compression and coding

Example: AAC encoding

- Representation of a signal on a given basis allows to make compression (that is, selection of the "most representative" basis coefficients and suppression of the "least representative").
- Example: AAC (Advanced Audio Coding) uses a decomposition of the signal on a local cosine basis (we'll see later what it means...).
- Good choices of the basis and of component selection allows to reduce the size of a given file. This problematic is called *signal compression*.
- The same methodology can be appplied to images (JPEG and JPEG-2000 encoding, MPEG-2 on DVDs, etc...)

Outline

Definition of the Fourier Transform

"Mr. Fourier, there is no future in your theory..."

Fourier transform

For a function $f \in L^1(\mathbb{R})$, the Fourier transform is defined by

$$\hat{f}(\omega) \triangleq \int_{\mathbb{R}} f(t)\, \mathrm{e}^{-\mathrm{i}\omega t}\, dt \ .$$

If $\hat{f} \in L^1(\mathbb{R})$, we also have that:

$$f(t) \triangleq \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\omega)\, \mathrm{e}^{\mathrm{i}\omega t}\, d\omega \ .$$

For the physicist, the Fourier transform quantifies the number of oscillations of f at the frequency ω. A density argument ($\overline{L^1 \cap L^2} = L^2$) allows to extend this definition to the functions of $L^2(\mathbb{R})$.

Properties of the Fourier Transform

Regularity

Regularity of the Fourier Transform

A function f is bounded and has continuous and bounded derivatives up to order p if

$$\int_{\mathbb{R}} |\hat{f}(\omega)|(1+|\omega|^p)\, d\omega < \infty.$$

For example, if $\hat{f}$ has compact support then f is C^∞.

Uncertainty Principle
Intuition

Question: Is is possible to write a function "localized in time AND in frequency ?"

First answer: No ! If I make a given function "more localized" in time like this:

$$f_s(t) = \frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right),$$

then I keep the energy in time ($\|f\| = \|f_s\|$), and we lose localisation in frequency ($\hat{f}_s(\omega) = \sqrt{s}\hat{f}(s\omega)$).

Uncertainty Principle

Intuition

Question: Is is possible to write a function "localized in time AND in frequency ?"

First answer: No ! If I make a given function "more localized" in time like this:

$$f_s(t) = \frac{1}{\sqrt{s}} f\left(\frac{t}{s}\right),$$

then I keep the energy in time ($\|f\| = \|f_s\|$), and we lose localisation in frequency ($\hat{f}_s(\omega) = \sqrt{s}\hat{f}(s\omega)$).

Uncertainty Principle

Quantum mechanics answer

In Quantum Mechanics, a particle is described in dimension 1 by a wave function $f \in L^2(\mathbb{R})$. The mean position of a particle is given by

$$u = \frac{1}{\|f\|^2} \int_{\mathbb{R}} t|f(t)|^2 \, dt$$

and its mean quantity of movement is given by

$$\xi = \frac{1}{2\pi\|f\|^2} \int_{\mathbb{R}} \omega|\hat{f}(\omega)|^2 \, d\omega.$$

The variance around these mean values is given by

$$\sigma_u^2 = \frac{1}{\|f\|^2} \int_{\mathbb{R}} (t-u)^2|f(t)|^2 \, dt$$

$$\sigma_\xi^2 = \frac{1}{\|f\|^2} \int_{\mathbb{R}} (\omega-\xi)^2|\hat{f}(\omega)|^2 \, d\omega$$

Uncertainty Principle
Heisenberg Inequality

Heisenberg's Inequality

For $f \in L^2(\mathbb{R})$, we have

$$\sigma_u^2 \sigma_\xi^2 \geq \frac{1}{4} ,$$

with equality if and only if there exists $(u, \xi, a, b) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that

$$f(t) = a \exp\left(\mathrm{i}\xi t - b(t-u)^2\right) .$$

Compact Support constraints

Don't place your hopes too high

Support constraints

If $f \neq 0$ has compact support, then $\hat{f}$ cannot be equal to 0 on an interval. Conversely, if $\hat{f} \neq 0$ has compact support, then f cannot be equal to 0 on an interval.

Use of the Fourier tranform for deconvolution

A first answer

- Recall that the problem of deconvolution is given by $Y_k = X_k + \varepsilon_k,\ k = 1 \dots n$. Assume that we know the noise distribution.
- From the pdf point of view, we get that: $f_Y = f_X \star f_\varepsilon$, then a first answer to deconvolution would be given by the inverse Fourier transform of

$$\hat{f}_X = \frac{\hat{f}_Y}{\hat{f}_\varepsilon}$$

Use of the Fourier tranform for deconvolution
A first answer

- However, this approach is not stable numerically (e.g., if the noise is assumed to be gaussian, the Fourier transform of the pdf of the noise decays quickly to 0).
- A possible correction is to "threshold" the Fourier transform

$$\hat{f}_X = \frac{\hat{f}_Y}{\hat{f}_\varepsilon \wedge c} \,.$$

- Still, the method remains "ad hoc". In fact, this problem is encountered in deconvolution whatever the method employed.

Issues for Fourier basis and Fourier transform

Advantages...

- Fourier transform is a powerful and simple tool.
- It is quite well fitted for stationary signals.
- It suffers however from “structural limitations” (Heisenberg's inequality, compact support properties).

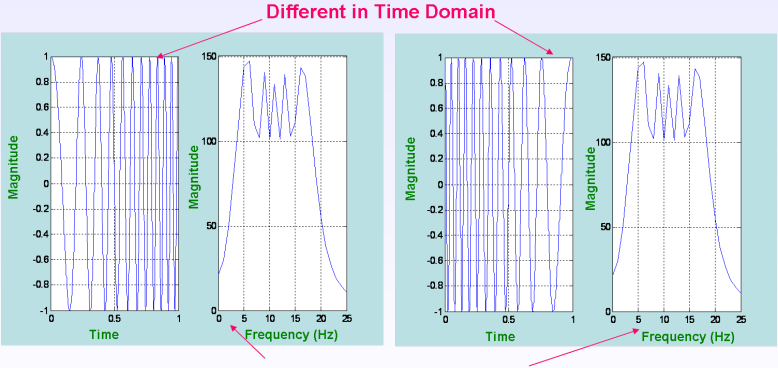
Issues for Fourier basis and Fourier transform

... and drawbacks

- Fourier Transform only gives which frequency components exist in the signal.
- The time and frequency information can not be seen at the same time.

$\Rightarrow$ Fourier transform is not an adapted tool to deal with non-stationary signals. For this, time-frequency representation of the signal is needed.

Issues for Fourier basis and Fourier transform



Outline

Time-frequency atoms

Decomposition

- A linear time-frequency transform decomposes the signal in a family of functions "well localized in time and energy".
- Such functions are called "time-frequency atoms". Consider a general family of time-frequency atoms $\{\phi_\gamma\}_\gamma$, where γ may be multidimensional, and assume that $\phi_\gamma \in L^2(\mathbb{R})$ and $\|\phi_\gamma\| = 1$.
- In that case,

$$Tf(\gamma) = \int_{\mathbb{R}} f(t)\phi_\gamma^*(t)\,dt = \langle f, \phi_\gamma \rangle$$

carries local information on time. Moreover, due to Plancherel Theorem:

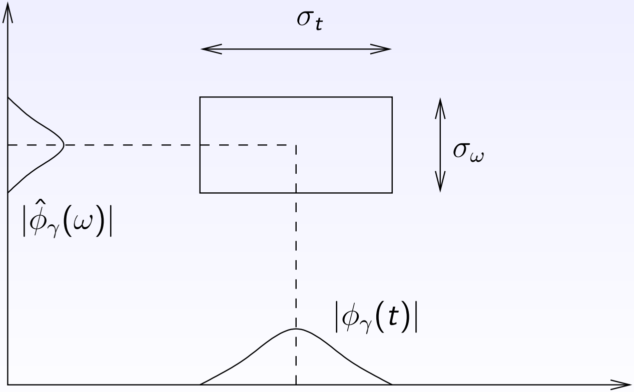
$$Tf(\gamma) = \int_{\mathbb{R}} \hat{f}(\omega)\hat{\phi}_\gamma^*(\omega)\,d\omega$$

thus giving frequency information.

Time-frequency atoms

Representation as Heisenberg boxes

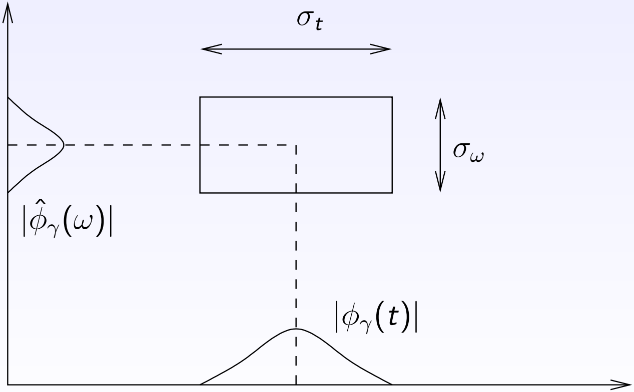
On the time-frequency plane, an atom is not a point of the plan, but a rectangle according to uncertainty principle.



Time-frequency atoms

Representation as Heisenberg boxes

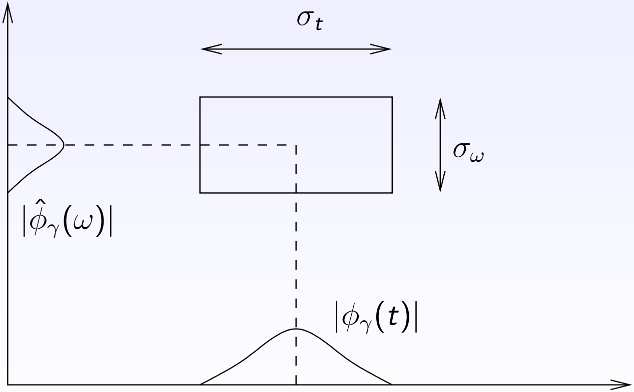
Due to Heisenberg inequality, only rectangles with surface $\geq 1/2$ can be time-frequency atoms.



Time-frequency atoms

Representation as Heisenberg boxes

There is a lower bound, but no upper bound, consequently the time-frequency plane can be split using different methods.



A first answer: short-term Fourier transform

Definition

Short-Time Fourier transform

Let be g an even real-valued function, such that $\|g\| = 1$. A time-frequency atom $g_{u,\xi}$ is obtained by translation and modulation:

$$g_{u,\xi}(t) = \mathrm{e}^{\mathrm{i}\xi t} g(t-u) \; .$$

The Short-term Fourier transform is obtained as:

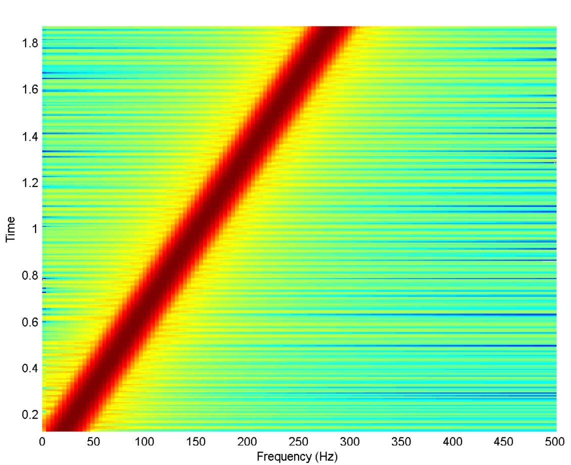
$$Sf(u,\xi) = \langle f, g_{u,\xi} \rangle = \int_{\mathbb{R}} f(t) g(t-u) \mathrm{e}^{-\mathrm{i}\xi t} \; .$$

We define the *spectrogram* as the PSD associated to the Short-ter Fourier transform:

$$P_S f(u,\xi) = \|Sf(u,\xi)\|^2.$$

A first answer: short-term Fourier transform

Example: the chirp



A first answer: short-term Fourier transform

Related Heisenberg's boxes

We get

$$\sigma_t^2 = \int_{\mathbb{R}} (t-u)^2 |g_{u,\xi}(t)|^2 \, dt = \int_{\mathbb{R}} t^2 |g(t)|^2 \, dt$$

and

$$\sigma_\xi^2 = \int_{\mathbb{R}} (\omega - \xi)^2 |\hat{g}(\omega - \xi) \exp(-\mathrm{i} u(\omega - \xi))|^2 \, d\omega = \int_{\mathbb{R}} \omega^2 |g(\omega)|^2 \, d\omega$$

Consequently, the atom $g_{u,\xi}$ has an Heisenberg box with surface $\sigma_t \sigma_\xi$, centered at (u, ξ)

A first answer: short-term Fourier transform

First conclusion

- Short-term Fourier transform allows to study the time-frequency plane more easily.
- The whole plane is covered by this boxes, thus allowing to retrieve the signal by inverse transformation.
- However, the size of Heisenborg boxes remain the same, which can hide some transitory states
- Moreover, the short-term Fourier transform assumes local stationarity.

A first answer: short-term Fourier transform

Theorem of reconstruction

"Reconstruction" of a signal, given its short-term Fourier transform

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

$$f(t) = \frac{1}{2\pi} \iint_{\mathbb{R}^2} \langle f, g_{u,\xi} \rangle g_{u,\xi}(t) \, d\xi \, du$$

Why the brackets ?This formula appears as a decomposition on an orthogonal basis, but it is not, since $\{g_{u,\xi}\}_{(u,\xi)\in\mathbb{R}^2}$ is redundant. This redundancy implies that a given function of $L^2(\mathbb{R}^2)$ is not necessarily the short-term Fourier transform of a signal.

A first answer: short-term Fourier transform

Theorem of reconstruction

"Reconstruction" of a signal, given its short-term Fourier transform

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

$$f(t) = \frac{1}{2\pi} \iint_{\mathbb{R}^2} \langle f, g_{u,\xi} \rangle g_{u,\xi}(t) \, d\xi \, du$$

Why the brackets ?This formula appears as a decomposition on an orthogonal basis, but it is not, since $\{g_{u,\xi}\}_{(u,\xi)\in\mathbb{R}^2}$ is redundant. This redundancy implies that a given function of $L^2(\mathbb{R}^2)$ is not necessarily the short-term Fourier transform of a signal.

A first answer: short-term Fourier transform

Theorem of reconstruction

“Reconstruction” of a signal, given its short-term Fourier transform

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

$$f(t) = \frac{1}{2\pi} \iint_{\mathbb{R}^2} \langle f, g_{u,\xi} \rangle g_{u,\xi}(t)\, d\xi\, du$$

Why the brackets ?This formula appears as a decomposition on an orthogonal basis, but it is not, since $\{g_{u,\xi}\}_{(u,\xi)\in\mathbb{R}^2}$ is redundant. This redundancy implies that a given function of $L^2(\mathbb{R}^2)$ is not necessarily the short-term Fourier transform of a signal.

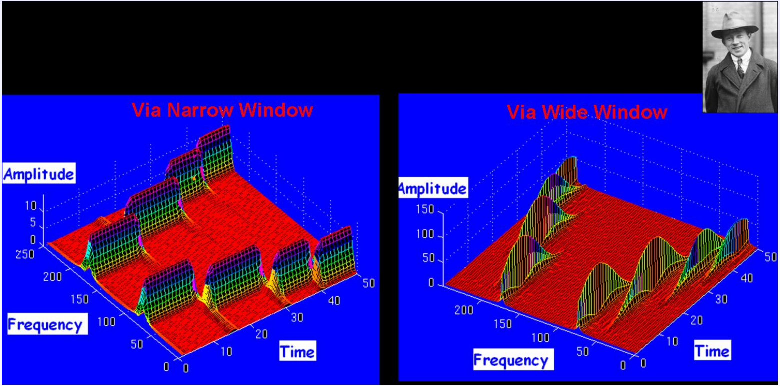
The short-term Fourier transform

The problems to overcome at this stage

- The fixed resolution is a problem to deal with brutal transitions of a signal.
- The redundancy bothers us to reconstruct the signal from its time frequency representation.

The short-term Fourier transform

Heisenberg 1 - Signal processing 0



Second answer: multiresolution analysis

A way to circumvent the resolution problem

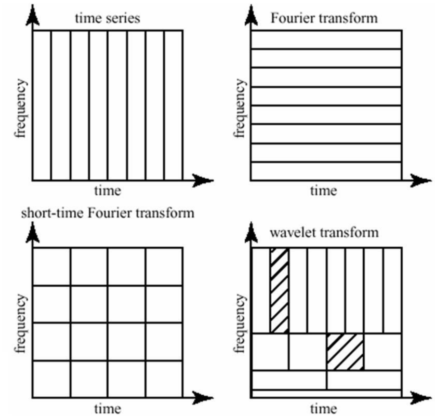
Multiresolution Analysis:

- Analyze the signal at different frequencies with different resolutions
- Good time resolution and poor frequency resolution at high frequencies
- Good frequency resolution and poor time resolution at low frequencies

$\Rightarrow$ More suitable for short duration of higher frequency; and longer duration of lower frequency components

Second answer: multiresolution analysis

Intuitive view for wavelet: a cunning way to split the time-frequency plane



The Continuous Wavelet Transform (CWT)

Definition and first example

Wavelet transform

A *wavelet* is an even function Ψ of $L^2(\mathbb{R})$ such that

$$\int_{\mathbb{R}} \Psi = 0 \text{ . and } \|\Psi\| = 1.$$

From this, we define a time-frequency atom as follows:

$$\Psi_{u,s} \triangleq \frac{1}{\sqrt{s}} \Psi(\frac{t-u}{s})$$

and we define the Continuous Wavelet Transform of a function of L^2 as:

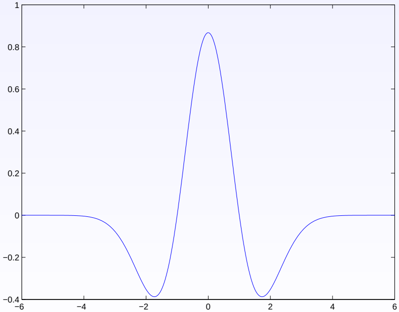
$$Wf(u,s) = \langle f, \Psi_{u,s} \rangle = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{s}} \Psi^*(\frac{t-u}{s})\, dt$$

The Continuous Wavelet Transform (CWT)

Definition and first example

Example: The "mexican hat" wavelet is the second derivative of the Gaussian probability density function.

$$mexh(x) = c \exp(-x^2/2)(1-x^2), c = \frac{2}{\sqrt{3} * pi^{1/4}}$$



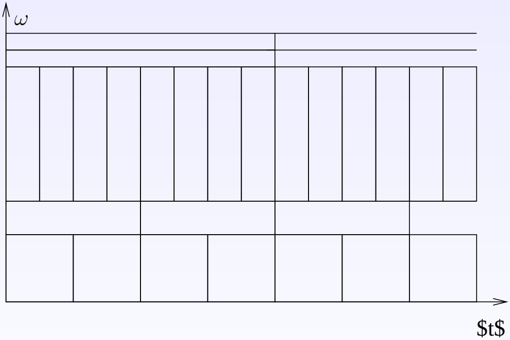
The Continuous Wavelet Transform (CWT)

Properties

The Continuous Wavelet Transform (CWT)

Properties

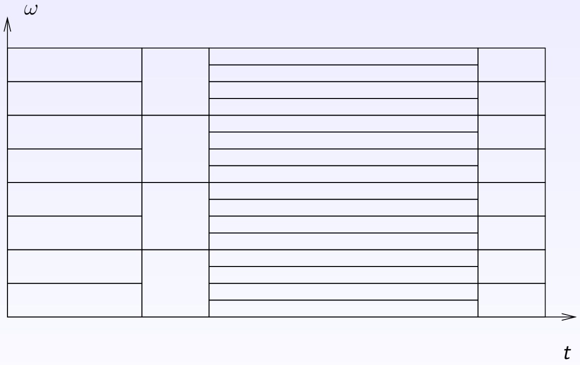
Other possible splits
Block wavelets



For block wavelets, the frequency domain is split in "boxes" with arbitrary lengths, and are translated in time.

Other possible splits

Local cosine decomposition



Local cosine decomposition is the opposite: decompose first the time domain and then translate.

Formalisation of multiresolution analysis

Definition of multiresolution

The following formalism is introduced by Mallat and Meyer:

Multiresolution

A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution if the following properties are verified:

1. $\forall j, k \; f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j$
2. $\mathbf{V}_{j+1} \in \mathbf{V}_j$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_j$
4. $\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = Adh\left(\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j\right) = L^2(\mathbb{R})$
6. There exists a Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_0$

Formalisation of multiresolution analysis

Definition of multiresolution

The following formalism is introduced by Mallat and Meyer:

Multiresolution

A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution if the following properties are verified:

1. $\forall j, k\ f(t) \in \mathbf{V}_j \Leftrightarrow f(t-2^j k) \in \mathbf{V}_j$
2. $\mathbf{V}_{j+1} \in \mathbf{V}_j$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_j$
4. $\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = Adh\left(\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j\right) = L^2(\mathbb{R})$
6. There exists a Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_0$

Formalisation of multiresolution analysis

Definition of multiresolution

The following formalism is introduced by Mallat and Meyer:

Multiresolution

A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution if the following properties are verified:

1. $\forall j, k \; f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j$
2. $\mathbf{V}_{j+1} \in \mathbf{V}_j$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_j$
4. $\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = Adh\left(\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j\right) = L^2(\mathbb{R})$
6. There exists a Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_0$

Formalisation of multiresolution analysis

Definition of multiresolution

The following formalism is introduced by Mallat and Meyer:

Multiresolution

A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution if the following properties are verified:

1. $\forall j, k \; f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j$
2. $\mathbf{V}_{j+1} \in \mathbf{V}_j$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_j$
4. $\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = Adh\left(\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j\right) = L^2(\mathbb{R})$
6. There exists a Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_0$

Formalisation of multiresolution analysis

Definition of multiresolution

The following formalism is introduced by Mallat and Meyer:

Multiresolution

A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution if the following properties are verified:

1. $\forall j, k \ f(t) \in \mathbf{V}_j \Leftrightarrow f(t-2^j k) \in \mathbf{V}_j$
2. $\mathbf{V}_{j+1} \in \mathbf{V}_j$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_j$
4. $\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = Adh\left(\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j\right) = L^2(\mathbb{R})$
6. There exists a Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_0$

Formalisation of multiresolution analysis

Definition of multiresolution

The following formalism is introduced by Mallat and Meyer:

Multiresolution

A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution if the following properties are verified:

1. $\forall j, k \; f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j$
2. $\mathbf{V}_{j+1} \in \mathbf{V}_j$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_j$
4. $\lim_{j\to+\infty} \mathbf{V}_j = \bigcap_{j\in\mathbb{Z}} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = Adh\left(\bigcup_{j\in\mathbb{Z}} \mathbf{V}_j\right) = L^2(\mathbb{R})$
6. There exists a Riesz basis $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_0$

Formalisation of multiresolution analysis

Definition of multiresolution

Example 1: Piecewise constant approximation

$$\mathbf{V}_j = \{g \in L^2(\mathbb{R}); g \text{ constant on } [n2^j; (n+1)2^j[\}$$

Example 1: Spline approximation

$$\mathbf{V}_j = \{g \in L^2(\mathbb{R}); g \text{ polynomial of degree m on } [n2^j; (n+1)2^j[, g\ \ C^{m-1}\}$$

Formalisation of multiresolution analysis

Construction of a wavelet orthogonal basis

Let W_j be the orthogonal complement of V_j:

$$V_{j-1} = V_j \oplus W_j$$

V_j is the approximation space, W_j is then the "detail" space. The following theorem, due to Mallat and Meyer, gives a construction of an orthonormal basis of V_j W_j by dilatation and translation of a wavelet ψ.

Formalisation of multiresolution analysis

Construction of a wavelet orthogonal basis

Mallat, Meyer construction of an orthonormal basis for V_j

Let ϕ the scale function whose Fourier transform is defined by

$$\hat{\phi}(\omega) = \frac{\hat{\theta}(\omega)}{\sqrt{\sum_{k\in\mathbb{Z}} \left|\hat{\theta}(\omega + 2k\pi)\right|^2}};$$

Then for all resolution level j, the family

$$\left\{\phi_{j,n} \triangleq \frac{1}{\sqrt{2^j}}\phi\left(\frac{t - 2^j n}{2^j}\right)\right\}_{n\in\mathbb{Z}}$$

is an orthonormal basis of V_j. $\phi_{j,n}$ is called the approximation wavelet.

Formalisation of multiresolution analysis

Construction of a wavelet orthogonal basis

Mallat, Meyer construction of an orthonormal basis for W_j

Let ϕ an integrable scale function and denote by $h(n) = \langle \phi(t/2)/\sqrt{2}, \phi(t-n) \rangle$. Let ψ the function defined by its Fourier transform:

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} \mathrm{e}^{-\mathrm{i}\omega/2} \hat{h}^*(\omega/2 + \pi) \times \hat{\phi}(\omega/2)$$

Then for all resolution level j, the family

$$\left\{ \psi_{j,n} \triangleq \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j n}{2^j} \right) \right\}_{n \in \mathbb{Z}}$$

is an orthonormal basis of W_j. $\phi_{j,n}$ is called the detail wavelet. Moreover, $\{\psi_{j,n}\}_{n,j \in \mathbb{Z}^2}$ is an orthonormal basis of $L^2(\mathbb{R})$.

Some criteria to build a basis

Main objectives

- Most applications use the fact that the signal can be expressed by a limited number of wavelet coefficients (parcimony)
- Consequently, we must build ψ in order to guarantee that $\langle f, \psi_{j,n} \rangle$ would be close to 0 for a large class of j, n.
- If at sharp scale, most of the wavelet coefficients are "small", f will have only a small number of non-negligible wavelet coefficients.

Some criteria to build a basis

Moment conditions

- Intuitive idea: if f is locally regular, it can be approximated by a high order polynomial (say, of order p).
- Consequently, a wavelet coefficient equal to zero at high resolution is equivalent to an orthogonality condition:
- A good criterion for the function ψ is thus a moment condition:

$$\int t^k \psi(t)\, dt = 0, 0 \leq k < p$$

Some criteria to build a basis

Support condition

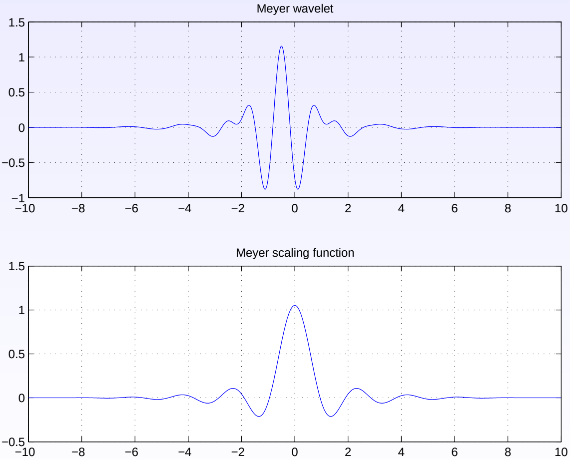
- Intuitive idea: minimizing the support of ψ should maximize the number of zeros.
- A good criterion for the functions ϕ and ψ is to take them with compact support.
- Indeed, we can show that if ϕ has $[N_1, N_2]$ for support, then the approximation wavelet built using Mallat-Meyer theorem has also compact support $[(N_1 - N_2 + 1)/2, (N_2 - N_1 + 1)/2]$.

Some criteria to build a basis

Moment-support trade-off

- Number of moments and support size are correlated: if ψ a has p moments equal to 0, then the size of its support is at least $2p - 1$.
- The wavelet has to be chosen with respect to the application, whether the number of singularities and the type of regularity between them.

Example: The Meyer wavelet family
The Meyer wavelet



Outline

Estimator by basis projection

Linear approximation error

Let $\{g_m\}_{n\in\mathbb{N}}$ be an orthonormal basis of an Hilbert space H. Any $f \in H$ is decomposed as

$$f = \sum_{m=0}^{+\infty} \langle f, g_m \rangle g_m \ .$$

A projection estimator is obtained by taking only the first components:

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m \ .$$

The approximation error tends to 0, but we don't know at each rate:

$$\varepsilon(M) \stackrel{\Delta}{=} \|f - f_M\|^2 = \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2 \longrightarrow_{M\to\infty} 0 \ .$$

Estimator by basis projection

Introduction of Sobolev spaces

The following theorem gives information on the decreasing rate of $\varepsilon(M)$.

Rate of convergence of the linear approximation error

For all $s > 1/2$, there exists $A > 0$ and $B > 0$ such that if

$$\sum_{m=0}^{+\infty} |m|^{2s} |\langle f, g_m \rangle|^2 < +\infty,$$

then

$$A \sum_{m=0}^{+\infty} m^{2s} |\langle f, g_m \rangle|^2 \leq \sum_{m=0}^{+\infty} M^{2s-1} \varepsilon(M) \leq B \sum_{m=0}^{+\infty} m^{2s} |\langle f, g_m \rangle|^2,$$

and then $\varepsilon(M) = o(M^{-2s})$.

Estimator by basis projection

Introduction of Sobolev spaces

The previous theorem gives a rate of convergence provided that

$$f \in \mathbf{W}_{B,s} \triangleq \left\{ f \in H;\ \sum_{m=0}^{+\infty} |m|^{2s} |\langle f, g_m \rangle|^2 < +\infty \right\}.$$

This kind of space defines the regularity of f is the considered basis is a Fourier or a wavelet basis in the sense of "general differentiability". We define the Sobolev space with index s:

$$\mathbf{W}_s(\mathbb{R}) \triangleq \left\{ f \in L^2(\mathbb{R});\ \int_{\mathbb{R}} |\omega|^{2s} |\hat{f}(\omega)|^2 \, d\omega < \infty \right\}$$

and

$$\mathbf{W}_s([0;1]) \triangleq \left\{ f \in L^2([0;1]); \exists g \in \mathbf{W}_s(\mathbb{R}), g_{|[0;1]} = f \right\}$$

Then, the error for a Fourier basis approximation decreases quickly if f is in a Sobolev space of big index s.

Estimator by basis projection

Problems related to Sobolev spaces

- If f has singularities, then it cannot belong to $\mathbf{W}_s([0;1])$ for all $s > 1/2$.
- The linear approximation error is localized around the discontinuities (Gibbs oscillations).
- The M first components are not necessarily the best to represent a function f (not the most representative)
- For linear approximation, a first answer to this issue is the Karhunen-Loeve decomposition (principal components).

Linear approximation
Example

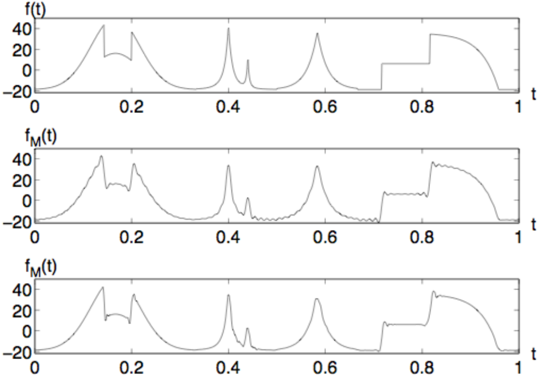


Figure 9.1: Top: Original signal f. Middle: Signal f_M approximated from lower frequency Fourier coefficients, with $M/N = 0.15$ and $\|f - f_M\|/\|f\| = 8.63\,10^{-2}$. Bottom: Signal f_M approximated from larger scale Daubechies 4 wavelet coefficients, with $M/N = 0.15$ and $\|f - f_M\|/\|f\| = 8.58\,10^{-2}$.

Nonlinear approximation
Main idea

- A projection estimate takes the first vectors to estimate a function.
- A threshold estimate (nonlinear approximation) takes some vectors belonging to a general subbasis I_M:

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m$$

- The indices in I_M should be chosen such that $|\langle f, g_m \rangle|$ are big (principal structures of f), in that case the nonlinear estimate is obtained by a thresholding operation.
- The approximation error is then

$$\varepsilon(M) = \|f - f_M\|^2 = \sum_{m \notin I_M} |\langle f, g_m \rangle|^2$$

Nonlinear approximation

Decreasing rate of the approximation error

We rearrange the basis coefficients in a decreasing order. Denote by $f_B^r(k) = \langle f, g_{m_k} \rangle$ the k-th term of this new sequence. The first theorem relates the

approximation error when M increases to the decreasing rate of the sequence $f_B^r(k)$

Let $s > 1/2$. If there exists $C > 0$ such that $|f_B^r(k)| \leq Ck^{-s}$, then

$$\sum_{k=M+1}^{+\infty} |f_B^r(k)|^2 \leq \frac{C^2}{2s-1} M^{1-2s}.$$

Nonlinear approximation

Decreasing rate of the approximation error

We rearrange the basis coefficients in a decreasing order. Denote by $f_B^r(k) = \langle f, g_{m_k} \rangle$ the k-th term of this new sequence. The second theorem relates

the decreasing rate of the error to the l^p-norm of f.

Let $p < 2$. If $\|f\|_{B,p} < \infty$, then

$$|f_B^r(k)| \leq \|f\|_{B,p} k^{-1/p} \text{ and } \sum_{k=M+1}^{+\infty} |f_B^r(k)|^2 = o(M^{1-2/p})$$

Nonlinear approximation

Combining nonlinear approximation and wavelet decomposition

- A nonlinear approximation on a wavelet basis defines an adaptative grid, so that the scale is refined around the singularities.
- It is possible to show that if the wavelet coefficients decrease fast enough, the approximation error is small.
- This is related to the study of *Besov spaces*

$$\mathbf{B}^s_{\beta,\gamma}([0;1]) \triangleq \left\{ f \in L^2([0.1]);\ \|f\|_{s,\beta,\gamma} < \infty \right\}$$

$$\|f\|_{s,\beta,\gamma} \triangleq \left(\sum_{j=-\infty}^{J+1} \left[2^{-j(s+0.5+1/\beta)} \left(\sum_{n=0}^{2^{-j}-1} |\langle f, \psi_{j,n} \rangle|^{\beta} \right)^{1/\beta} \right]^{\gamma} \right)^{1/\gamma}$$

($\beta > 2$: "uniformly regular functions", $\beta = \gamma = 2$: Sobolev space, $\beta < 2$: functions with irregularities)

Nonlinear approximation

Example

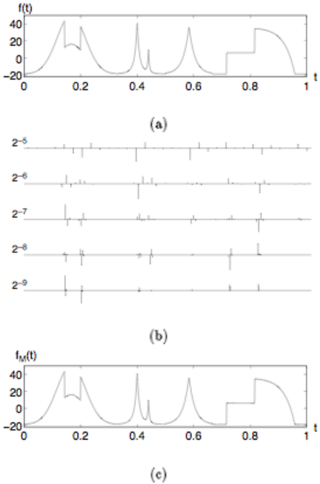
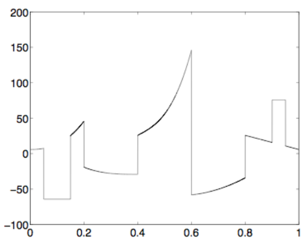
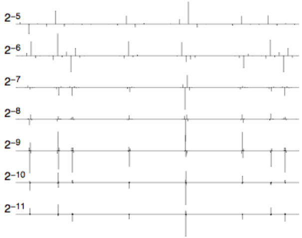


Figure 9.2: (a): Original signal f. (b): Larger $M = 0.15\,N$ wavelet coeffi-

Some notes on denoising and deconvolution problems

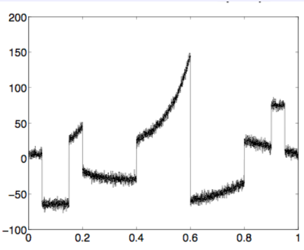
Examples of denoised signal by wavelet soft thresholding

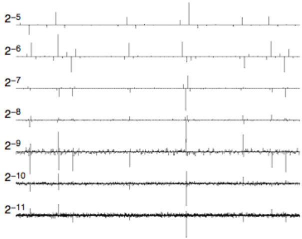




Some notes on denoising and deconvolution problems

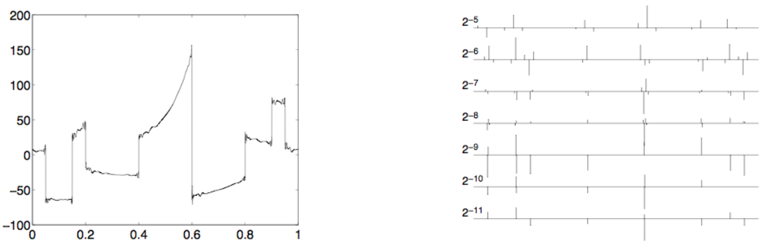
Examples of denoised signal by wavelet soft thresholding





Some notes on denoising and deconvolution problems

Examples of denoised signal by wavelet soft thresholding



Some notes on denoising and deconvolution problems

Remarks on the deconvolution problem

- An additive noise usually decreases rates of convergence of threshold estimates
- If the noise density is "smooth" (that is, its Fourier transform decays polynomially to 0), then the deconvolution can be done at standard rates.
- On the other hand, if the noise density is supersmooth (eg, gaussian), the convergence rates decrease.
- If furthermore, we know nothing on the variance of the noise, then the rates of convergence drastically decrease (relate to Wiener filter).

Thank You !