Wavelets and Applications

Kévin Polisano

kevin.polisano@univ-grenoble-alpes.fr

M2 MSIAM & Ensimag 3A MMIS

January 13, 2020

Course Wavelets and Applications – Outline

1. Introduction: from Fourier to Wavelets
2. The 1D Continuous Wavelet Transform
3. Wavelet zoom: a local characterization of functions
4. Lab1: 1D Continuous Wavelet Transform
5. The 2D Continuous Wavelet Transform
6. Wavelet Bases (Haar, multiresolution, orthogonal wavelet bases)
7. Fast Wavelet Transform (the 1D and 2D cases)
8. Lab 2: Fast Wavelet Transform, image compression and denoising
9. Approximation in wavelet bases (sparsity, compression, denoising)
10. Application 1: The dual-tree complex wavelet transform and the scattering transform for deep learning
11. Application 2: Introduction to wavelets on graphs
12. Lab 3: Dedicated to the project

# Course evaluation

1. 2 lab sessions using the library PyWavelet (results to be sent by email)

2. **Examination**. 1 project consisting on:
   - Choosing a research article using wavelets for a given application.
   - Reading and understanding the article, writing a summary of what you expect to implement (**due to December 9th**).
   - Practical work: implementation of the method presented in the article (Matlab or Python).
   - Writing a report including figures of results.

3. **Ensimag students must form pairs** for lab sessions and the project (precise both your names on the summary)

Course materials

Books

- **S. Mallat**, *A wavelet tour of signal processing*, Academic press, third edition, 2009.
- **P. Flandrin**, *Explorations in Time-Frequency Analysis*, Cambridge University Press, Cambridge (UK), 2018.
- **J-P. Antoine**, R. Murenzi, P. Vandergheynst and S.T. Ali, *Two-dimensional Wavelets and Their Relatives*, Cambridge University Press, Cambridge (UK), 2004.
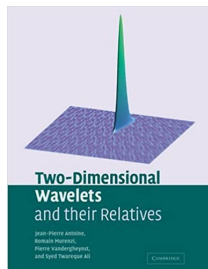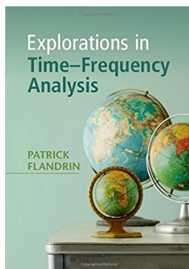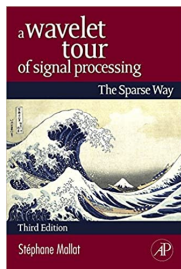
Course materials

Links

- **WaveLab** (free Matlab toolbox)

  http://www-stat.stanford.edu/ ~wavelab/

- **A numerical tour of Signal/Image Processing** (by Gabriel Peyré)

  http://www.numerical-tours.com/

- **PyWavelets** (python)
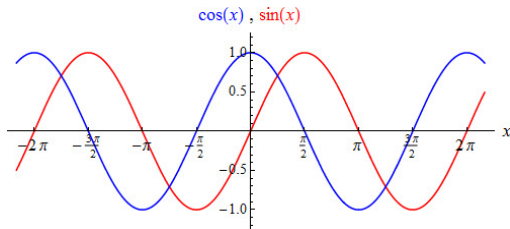
  https://pywavelets.readthedocs.io/

Introduction
From Fourier to Wavelets

What is a wavelet?
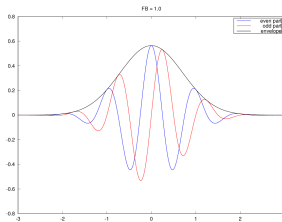
Examples of waves

- Electromagnetic wave
- Radio wave
- Microwave
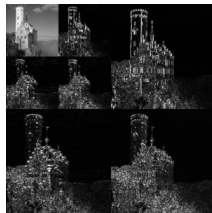- Sound wave

Wavelet = "short wave"
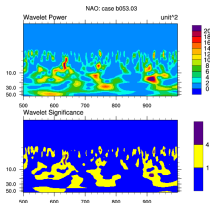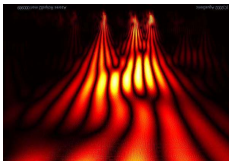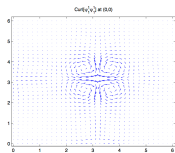

sinusoidal waves
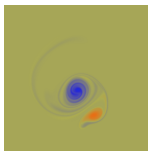

wavelet

Credits: Valérie Perrier

A success story

- *Wavelets for Data representation*

- *Wavelets for numerical simulation*

Divergence-free wavelet *Direct Simulation of Turbulence*

Credits: Valérie Perrier

A success story

WSQ (1993) is the **FBI**'s **Wavelet** Scalar Quantization: it is a national standard for the collecting, encoding, storing, and retrieving digitized fingerprint images.

JPEG 2000 is an image coding system that uses state-of-the-art compression techniques based on **wavelet technology**.

Academy Sci-Tech Award 2013

Awarded to **Theodore Kim**, **Nils Thuerey**, **Dr. Markus Gross** and **Doug James** for the invention, publication and dissemination of "**Wavelet Turbulence**" software.

Credits: Valérie Perrier

A success story

▶ **Abel Prize (2017)** : `Yves Meyer`, for his pivotal role in the development of the mathematical theory of wavelets.

"**Wavelet analysis** *has been applied in a wide variety of arenas as diverse as applied and computational harmonic analysis, data compression, noise reduction, medical imaging, archiving, digital cinema, deconvolution of the Hubble space telescope images, and the recent LIGO detection of gravitational waves created by the collision of two black holes.*"
`[http://www.abelprize.no/]`

The Abel Lecture (Yves Meyer)

www.youtube.com/watch?v=wxmzHwd3z34
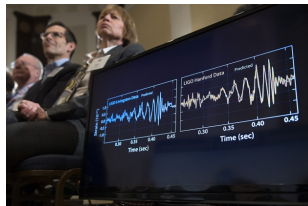
Wavelet theory helped LIGO to detect gravitational waves.

Credits: Valérie Perrier

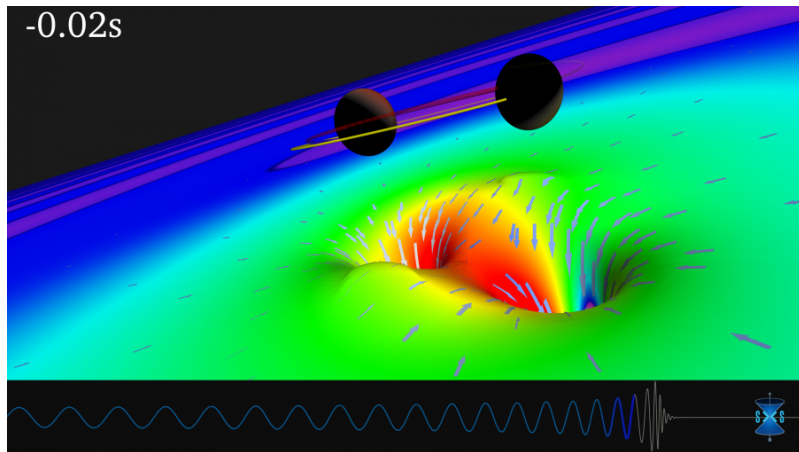# Gravitational waves detection

-0.02s

Credits: LIGO (http://www.black-holes.org/gw150914)

Gravitational waves detection
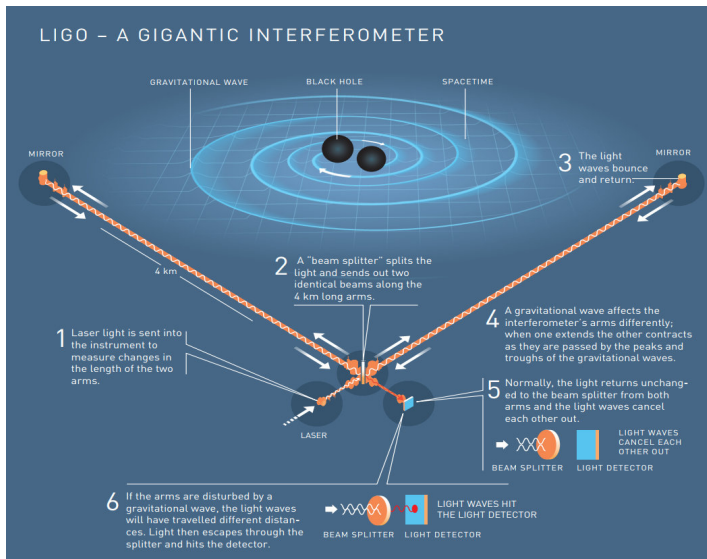
Credits: LIGO (http://directory.eoportal.org/web/eoportal/satellite-missions/l/ligo)

Gravitational waves detection

Analytic wave form derived from Einstein's equations (Thibault Damour et al.)
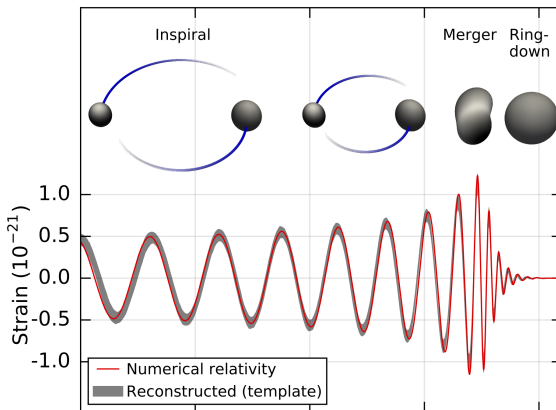
$$s(t) = c|t - t_0|^{-\frac{1}{4}}\cos(\omega|t - t_0|^{\frac{5}{8}} + \phi)$$

with c, ω, ϕ and t_0 are constants (t_0 = time when the stars are merging)

Credits : B.P. Abbott et al., Observation of Gravitational Waves from a Binary Black Hole Merger

Gravitational waves detection

Credits: B.P. Abbott et al., Observation of Gravitational Waves from a Binary Black Hole Merger

Gravitational waves detection

▶ Gravitational waves are perturbation of the space-time metric, predicted by Einstein and steadily sent to the Earth

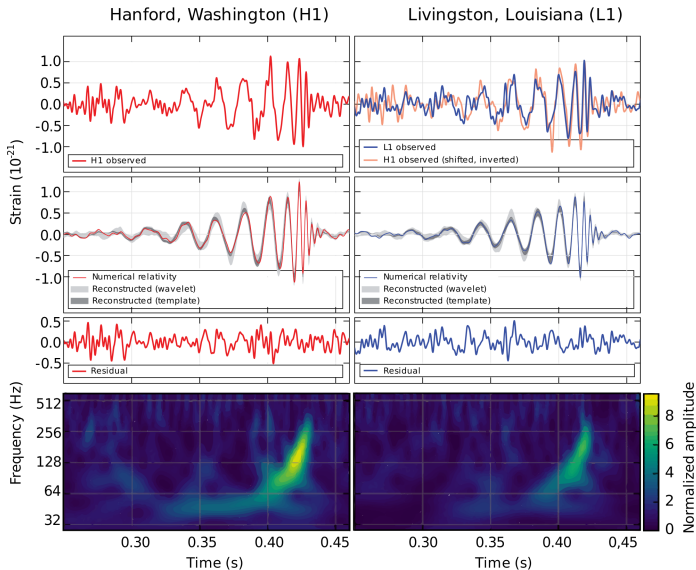▶ Gravitational waves, produced during the final fraction of second of the merger of two black holes into a single, are chirps like a modulated song of a bird, which are very tenious (noise/signal$>10^3$)

▶ Pattern detected by the 2 arms of an interferometer, separared by 3000km and 4km long, use laser lights and distant mirrors to be sensitive to derformations smaller than 10^{-19}m (!)

▶ Detecting a short-lived chirp burried inside a very noisy signal cannot be extrated by Fourier analysis (adapted to stationary signals)

▶ Wavelet analysis extends and overcomes Fourier limitation by exploring the time-frequency / time-scale structure of the data. It explores what is beyond our senses by yielding details that cannot be perceived by our eyes (it acts like a "zoom")

Credits: Abel lecture (Yves Meyer)

Gravitational waves detection

▶ Projection on bases of functions reasonably localized in time and frequency: Wilson transforms (modification of Gabor transforms)

▶ The window can be Meyer scaling function. The signal processing is performed on 7 Wilson bases (and their quadrature bases) each obtained by a dilation of factor 2 of the window

▶ Several decompositions at different time scales. Inspiral requires good frequency resolution, merger requires good time resolution: compromise between time frequency and time scale analysis

▶ Retain time-frequency pixels that are "phase coherent". Can detect unexpected sources. Fast and robust algorithm.
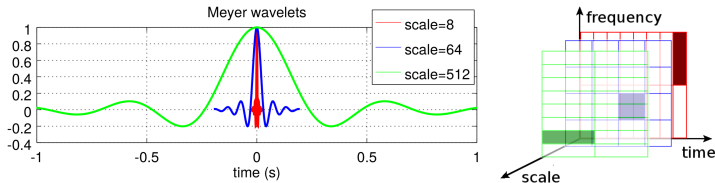
Credits: Eric Chassande-Mottin

Gravitational waves detection

Klimenko's algorithm: Coherent Wave Burst

Search for rare transients with low signal to noise ratio. Two ways:

▶ **Matched filtering**. Expected signal is known, targeted search signature of binary black-hole merger as predicted by general relativity

▶ **Time-frequency excess power**. Expected signal is unknown, search transients appearing in phase in all detectors with no waveform prior (general relativity not needed)

Y. Meyer & S. Klimenko

Non-parametric search

Credits: Yves Meyer & Eric Chassande-Mottin
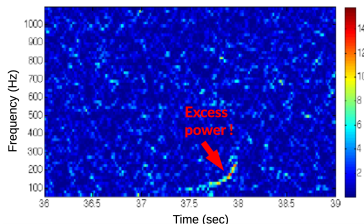
Gravitational waves detection

Astronomical objects

Physical device

Reconstruction

During the wave

Transient signals extraction

Klimenko's algorithm

Wavelet analysis explores beyond our senses
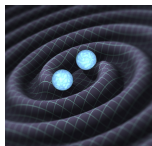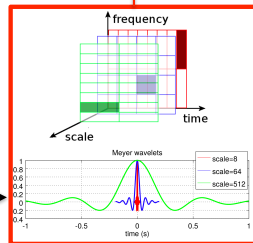
Hanford H1 raw data
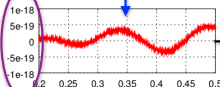
Credits: S. Klimenko et al., arXiv:1511.05999 + The economist

A new window on the universe

> *It's like Galileo pointing the telescope for the first time at the sky. You're opening your eyes — in this case, our ears — to a new set of signals from the universe that our previous technologies did not allow us to receive, study and learn from.*
>
> **Vassiliki Kalogera**

> *Up until now, we've been deaf to gravitational waves. What's going to come now is we're going to hear more things, and no doubt we'll hear things that we expected to hear... but we will also hear things that we never expected.*
>
> **David Reitze**

> *Writing the score while listening to the music, then analyzing and interpreting the score, is, in a sense, what Klimenko does to detect gravitational waves.*
>
> **Yves Meyer**

From the music of the spheres to the chirp of black holes

A small detour through old cosmology to meet Fourier

Back to the time of Ptolemy...

Who Wants to Be a Millionaire?
A pre-Copernican TV show

Credits: "Qui veut gagner des millions" (http://www.youtube.com/watch?v=ekmtqODjrSI)

The first heliocentric model by Aristarchus of Samos

Extract from The Sand Recknoner, Arenarius (Archimedes, c. 230 BC)

> " *You are now aware that the "universe" is the name given by most astronomers to the sphere, the centre of which is the centre of the earth, while its radius is equal to the straight line between the centre of the sun and the centre of the earth. This is the common account as you have heard from astronomers. But **Aristarchus** has brought out a book consisting of certain hypotheses, wherein it appears, as a consequence of the assumptions made, that the universe is many times greater than the "universe" just mentioned. His hypotheses are that the fixed stars and the sun remain unmoved, that the earth revolves about the sun on the circumference of a circle, the sun lying in the middle of the orbit, and that the sphere of the fixed stars, situated about the same centre as the sun, is so great that the circle in which he supposes the earth to revolve bears such a proportion to the distance of the fixed stars as the centre of the sphere bears to its surface.* "

Archimedes

Three competing models of the solar system

Ptolemy, Copernic and Tycho-Brahe models
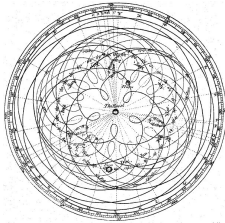
(g) Ptolemy
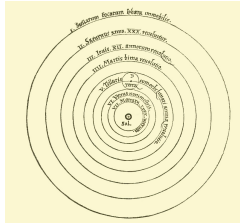
(h) Copernic
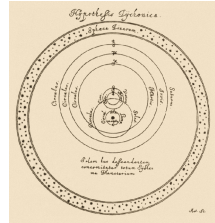
(i) Tycho-Brahe
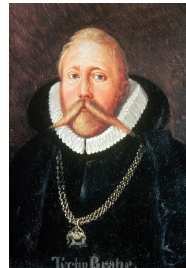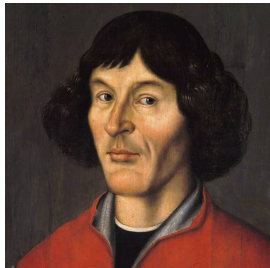
Heliocentrism versus Geocentrism
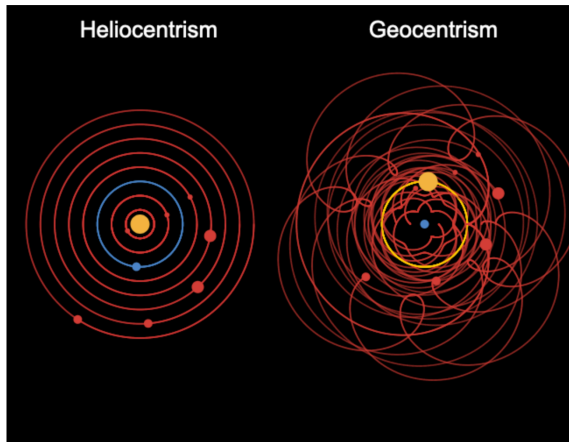
Remark: not realistic (the Copernican model is much more complex), it just serves to illustrate the relativity of motion.

Credits: Malin Christersson (http://www.malinc.se/math/trigonometry/geocentrismen.php)

Mars apparent retrograde motion
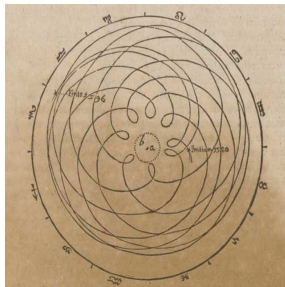
Mars apparent retrograde motion

" *If anyone were to believe that the sun truly moves in the space of a year through the zodiac, which **Ptolemy** and **Tycho Brahe** believed, then it is necessary to concede that the paths of the three superior Planets through ethereal space, composed as they are of several motions, are in reality spirals in the figure of a pretzel, in the general fashion that follows.* "

Johannes Kepler
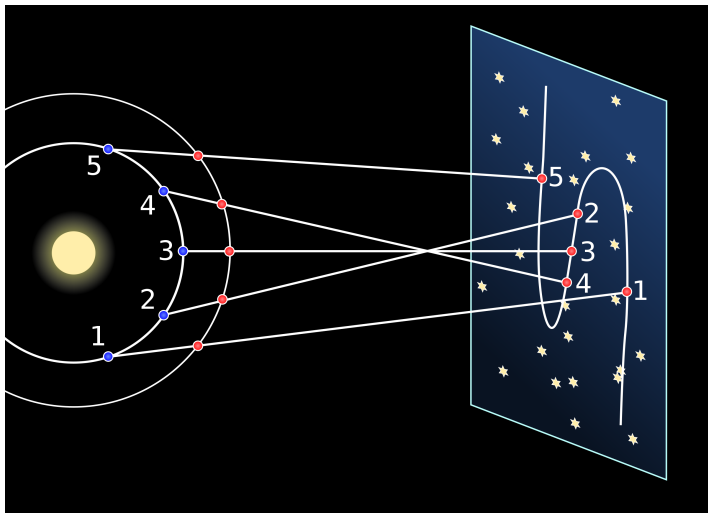
Mars apparent retrograde motion

Spirograph

Epicycles

Épicycloïde

Point n°1
Point n°2

Credits: El Jj (http://www.youtube.com/watch?v=uazPP0ny3XQ)
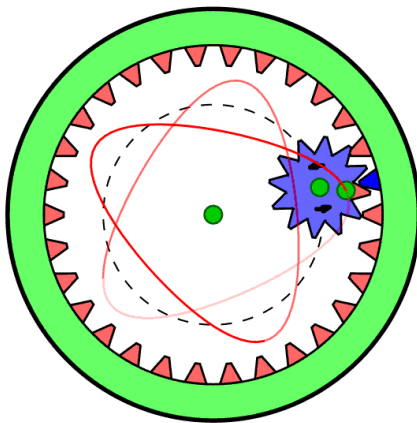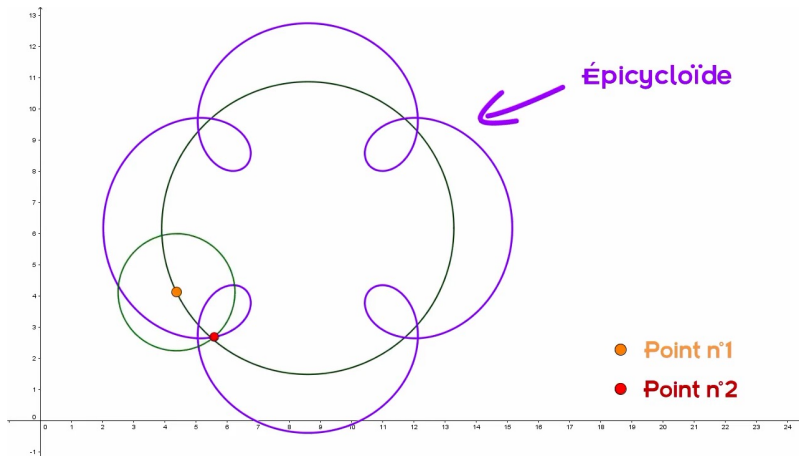
Roots of epicycles: from Hipparque to Kepler

▶ **Hipparque**'s epicycles theory based on **Apollonius of Perga**'s works (at the end of the 3rd century BC) + **Seleucus of Seleucia**

▶ **Ptolemy** refined the deferent-and-epicycle concept and introduced the equant as a mechanism for accounting for velocity variations in the motions of the planets

▶ **Copernicus** and his contemporaries were therefore using **Ptolemy**'s methods and finding them trustworthy well over a thousand years after **Ptolemy**'s original work was published

▶ In keeping with past practice, **Copernicus** used the deferent/epicycle model in his theory but his epicycles were small and were called *epicyclets*

▶ **Copernicus** eliminated **Ptolemy**'s somewhat-maligned equant but at a cost of additional epicycles. Various 16th-century books based on **Ptolemy** and **Copernicus** use about equal numbers of epicycles

Credits: Wikipedia (http://en.wikipedia.org/wiki/Deferent_and_epicycle)

Heliocentrism versus Geocentrism

Credits: Malin Christersson (http://www.malinc.se/math/trigonometry/geocentrismen.php)
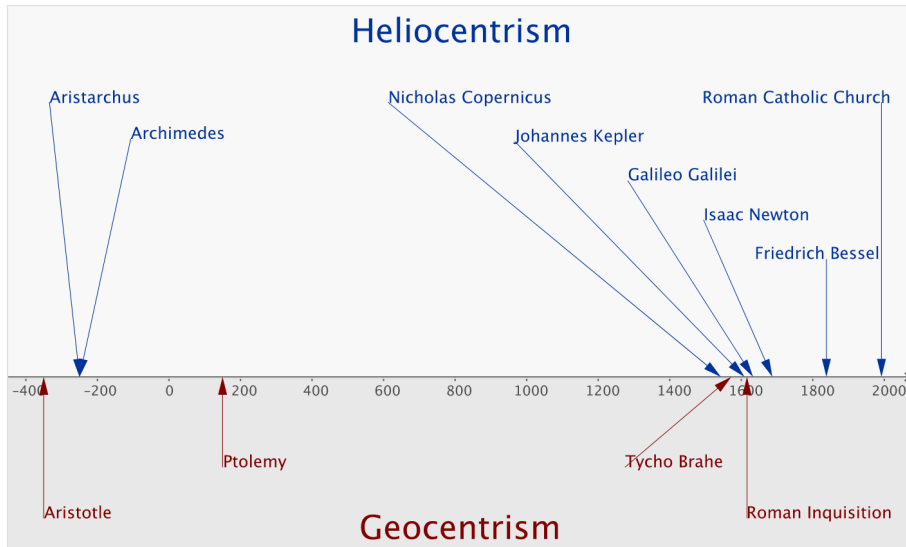
Heliocentrism versus Geocentrism

Why wasn't the heliocentric model capable of replacing the geocentric one?

Heliocentrism	Geocentrism
++ Explains retrograde motion	−− Cannot
++ Smaller epicycles used and avoiding the equants	−− Artificial equants used for differences of velocities observed
== Good accuracy for determining (mostly) planets location	== Good accuracy for determining (mostly) planets location
== Circular and uniform motions	== Circular and uniform motions
−− Fails to explain the divergence of Mars to observations	−− Fails to explain the divergence of Mars to observations
−− Movements of the Earth around the sun and on its axis contradict some observations (parallax, stars size, stability, ...)	++ Earth is stationary. **Tycho Brahe** proposed a mixed model called geoheliocentric observing deferents of Mars and the Sun crossing.
−− Philosophical breakthrough: the Earth is not a unique body anymore, which contradicts both Aristote and biblical arguments	++ Philosophically compliant with Aristote's conception of sublunary/aether distinction and with holy scriptures

The Paradigm shift: Kepler, Galileo and Newton discoveries
Newton standing on the shoulders of giants

▶ The intellectual climate of the time "remained dominated by Aristotelian philosophy and the Ptolemaic astronomy. At that time there was no reason to accept the Copernican theory, except for its mathematical simplicity." **Tycho Brahe**'s system ("that the earth is stationary, the sun revolves about the earth, and the other planets revolve about the sun") also directly competed with Copernicus.
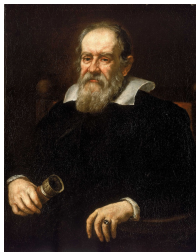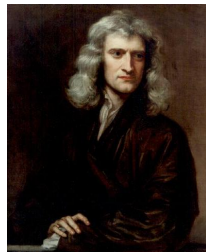
Kepler   Galileo   Newton

Credits: Wikipedia (http://en.wikipedia.org/wiki/Nicolaus_Copernicus)

The Paradigm shift: Kepler, Galileo and Newton discoveries
Newton standing on the shoulders of giants

▶ **Johannes Kepler** developed his laws of planetary motion using measurements made at Tycho's observatory. In Astronomia nova (1609), **Kepler** made a diagram of the movement of Mars in relation to Earth if Earth were at the center of its orbit, which shows that Mars'orbit would be completely imperfect and never follow along the same path. To solve the apparent derivation of Mars'orbit from a perfect circle, **Kepler** derived both a mathematical definition and, independently, a matching ellipse around the Sun to explain the motion of the red planet. Kepler's laws were born.
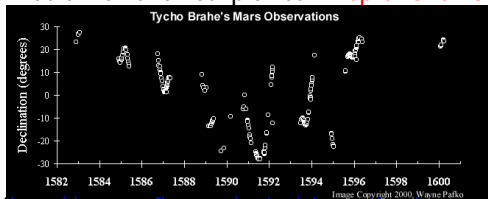
Tycho Brahe's Mars Observations

Credits: Wayne Pafko (http://www.pafko.com/tycho/observe.html)

The Paradigm shift: Kepler and Galileo discoveries

Newton standing on the shoulders of giants

▶ **Galileo** was able to look at the night sky with the newly invented telescope. He published his discoveries that Jupiter is orbited by moons and that the Sun rotates in his Sidereus Nuncius (1610)[93] and Letters on Sunspots (1613), respectively. Around this time, he also announced that Venus exhibits a full range of phases (satisfying an argument that had been made against Copernicus). Finally he discovered that the moon presents mountains, valleys and craters which depreciates the Aristotelian conception.

Credits: Pierce Wilcox

The Paradigm shift: Kepler and Galileo discoveries

Newton standing on the shoulders of giants

▶ **Galileo** was able to look at the night sky with the newly invented telescope. He published his discoveries that Jupiter is orbited by moons and that the Sun rotates in his Sidereus Nuncius (1610)[93] and Letters on Sunspots (1613), respectively. Around this time, he also announced that Venus exhibits a full range of phases (satisfying an argument that had been made against Copernicus). Finally he discovered that the moon presents mountains, valleys and craters which depreciates the Aristotelian conception.

▶ **Galileo** formulated the principle of inertia which helped to explain why everything would not fall off the earth if it were in motion.

▶ **Isaac Newton** formulated the universal law of gravitation and the laws of mechanics in his 1687 Principia, which unified terrestrial and celestial mechanics, was the heliocentric view generally accepted.

Credits: Wikipedia (http://en.wikipedia.org/wiki/Heliocentrism)

Why was Ptolemy's system so efficient?
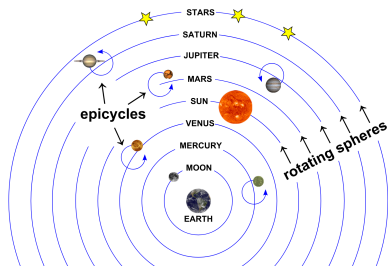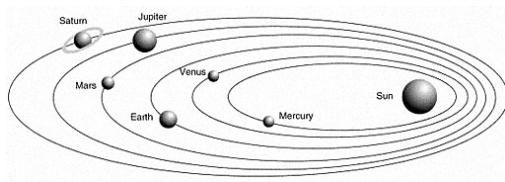
Credits: Joshua Hershey & Universe Today

Because of that...

... or more exactly thanks to him: Joseph Fourier

Just like Mr Jourdain speaking prose, astronomers made Fourier series without realizing it

Credits: 3blue1brown (http://www.youtube.com/watch?v=-qgreAUpPwM)
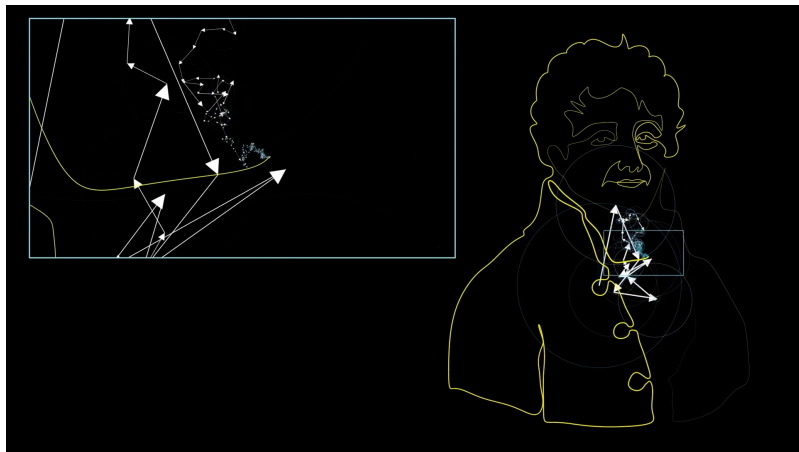
Fourier series: an intuition behind the decomposition

Periodic signals can be decomposed onto the Fourier basis

$$f(t) = \cdots + c_{-2}e^{-2\cdot2\pi it} + c_{-1}e^{-1\cdot2\pi it} + c_0e^{0\cdot2\pi it} + c_1e^{1\cdot2\pi it} + c_2e^{2\cdot2\pi it} + \cdots$$

Credits: 3blue1brown (http://www.youtube.com/watch?v=r6sGWTCMz2k)
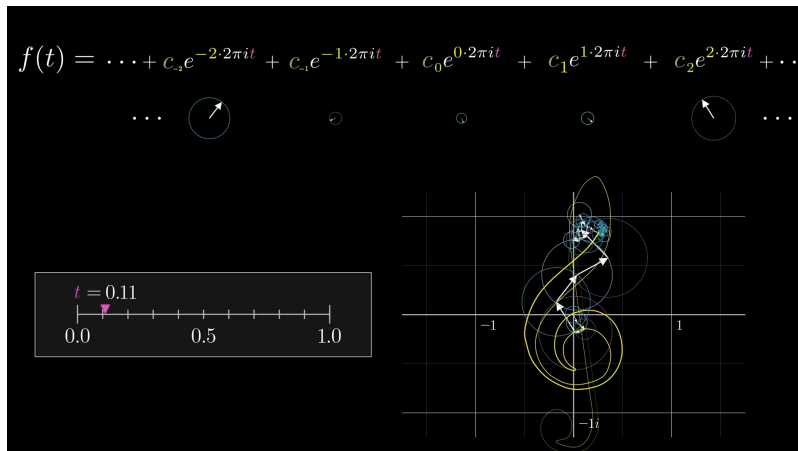
Draw me a (light weight) elephant

The fewer parameters the better
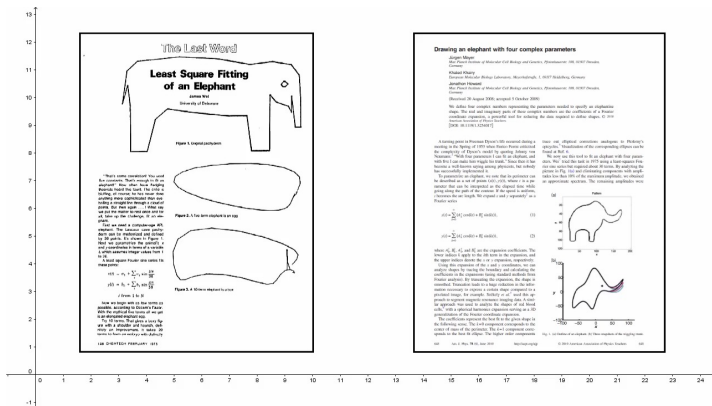
" With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.

John von Neumann

Fourier series: is it only useful for drawing?

How Joseph Fourier solved the heat equation

1) Sine = Nice
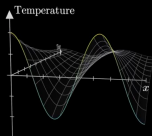
2) Linearity
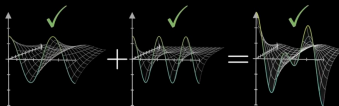
3) Fourier series

Credits: 3blue1brown (http://www.youtube.com/watch?v=r6sGWTCMz2k)
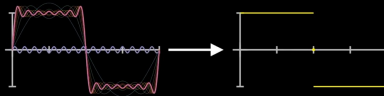
Fourier series analysis

The **Fourier analysis** decomposes a signal (function) $f(x)$ ($x = times$) into a sum of sinusoidal functions:

- For a T-<u>periodic function</u> f, with $f \in L^2(0, T)$:

$$f(x) = \sum_{n \in \mathbb{Z}} c_n(f) e^{2i\pi \frac{n}{T} x} \quad \text{(synthesis)}$$

  where the Fourier coefficients are:

$$c_n(f) = \frac{1}{T} \int_0^T f(x) e^{-2i\pi \frac{n}{T} x} \, dx \quad \text{(analysis)}$$

  are related to the *frequency* $\frac{n}{T}$ (in Hz).

Parseval equality:

$$\sum_{n \in \mathbb{Z}} |c_n(f)|^2 = \frac{1}{T} \int_0^T |f(x)|^2 dx \quad \text{(energy conservation)}$$

Fourier series limitations

Discontinuities require a lot of sinusoids to be described
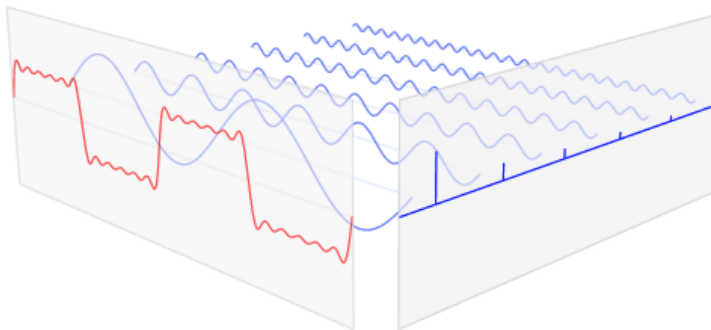
$$f(x) = \begin{cases} -1 \text{ if } -\pi \leqslant x < 0 \\ +1 \text{ if } 0 \leqslant x < \pi \end{cases} = \sum_{n=1}^{+\infty} \frac{4}{\pi(2n-1)} \sin((2n-1)x)$$

Credits: Wikipedia (https://en.wikipedia.org/wiki/Fourier_series)

Fourier transform

- For a function $f \in L^2(\mathbb{R})$:

$$f(x) = \int_{-\infty}^{+\infty} \hat{f}(\nu) e^{2i\pi\nu x} \, d\nu \quad \text{(synthesis)}$$

  where the Fourier transform of f is:

$$\hat{f}(\nu) = \int_{-\infty}^{+\infty} f(x) e^{-2i\pi\nu x} \, dx \quad \text{(analysis)}$$

  gives information on f for the frequency ν.

Plancherel-Parseval equality:

$$\int_{-\infty}^{+\infty} |\hat{f}(\nu)|^2 d\nu = \int_{-\infty}^{+\infty} |f(x)|^2 dx \quad \text{(energy conservation)}$$

Fourier transform: an intuition behind the transformation

aperiodic signals can also be decomposed onto the continuous dictionary of exponentials

Credits: 3blue1brown (https://www.youtube.com/watch?v=spUNpyF58BY)
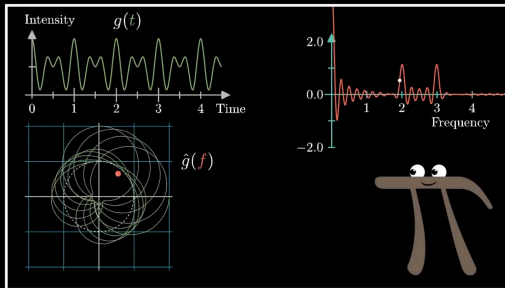
Fourier Transform visualization

Wrap the signal around a circle

5 Beats/Second Signal

3.12 Cycles/Second
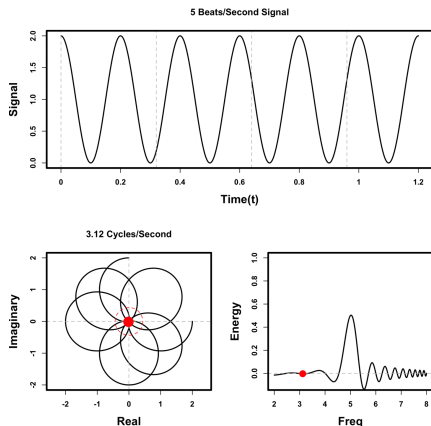
$$\widehat{g}(f) = \frac{1}{N} \sum_{k=1}^{N} g(t_k) e^{-2\pi i f t_k}$$

To find **the energy** at a particular frequency, **the signal** is wrapped around a circle at the particular frequency and the points along the path are averaged.

Credits: Elan Ness-Cohn

Fourier transform limitations

Example: two musical notes played at the same time
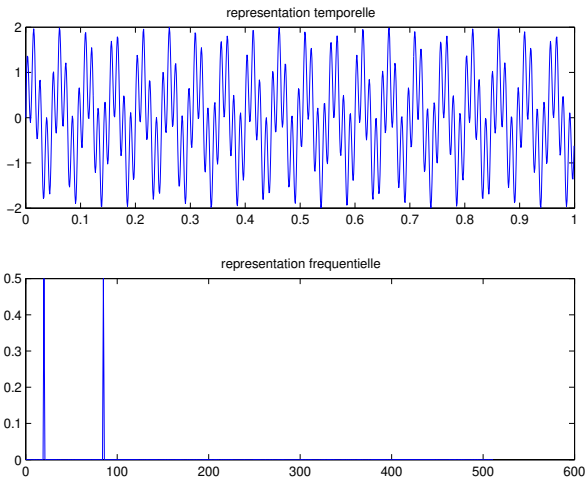
Figure: Signal $f(x) = \sin(40\pi x) + \sin(170\pi x)$ (top), and modulus of its Fourier Transform $\widehat{f}(\nu)$ (buttom)

Fourier transform limitations

Example of two musical notes played one after the other

The frequency analysis do not inform on the transient phenomenon in the signal $\Rightarrow$ Loss of temporal localization
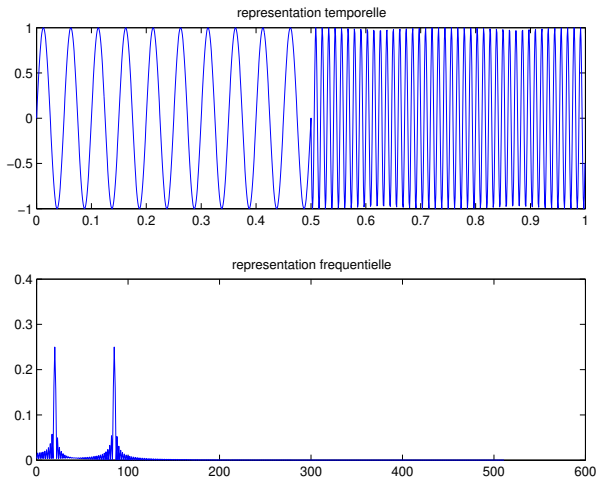
Fourier cat transformation
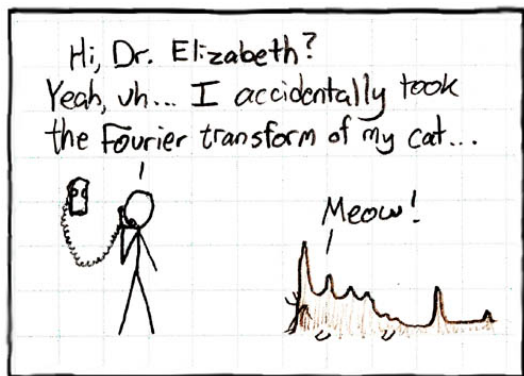
Take home message

To sum up

- Periodic functions (as planets motion along closed orbit) can be approximated by epicycles, that is by Fourier series.

- Fitting data do not necessarily mean that the mechanics behind is understood, and saving the phenomena can lead to a kind of overfitting. What is a good model or a good theory?

- The relativity of motion makes possible to consider different coordinate systems to describe trajectories. Something which is well known by physicists: the choice of the frame of reference can greatly simplify mathematical calculations.

- An appropriate representation of the signal can also reduced the number of parameters needed to encode its information.

⇒ **Toward a sparse representation of signals**

Take home message

To sum up

- Fourier series decomposition allowed Joseph Fourier to solve partial differential equations (heat equation).

- Extension to the Fourier transform for aperiodic signal also reveals the frequency contents of the signal, but suffer of the same issues:

- Discontinuities involve a lot of significant coefficients in the decomposition, whose the decrease in amplitudes encodes the *global* regularity of the signal.

- Losing the temporal localization, the Fourier transform does not allow to capture transient phenomena in the signal.

$\Rightarrow$ **Toward a time-frequency representation of signals**

The Continuous Wavelet Transform

Short Time Fourier Transform (STFT)
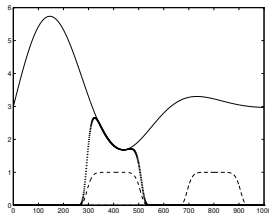
Multiplication of the signal $f(x)$ by a window $w(x - b)$ (real and of size a_0) and computation of the Fourier transform of this product:

$$Sf(\nu, b) = \int_{-\infty}^{+\infty} f(x)w(x - b)e^{-2i\pi\nu x}\,\mathrm{d}x$$

where b represents *time* and ν *frequency*. f can be recovered from its STFT coefficients:

$$f(x) = C_h \iint_{\mathbb{R}^2} Sf(\nu, b)w(x - b)e^{2i\pi\nu x}\,\mathrm{d}\nu\,\mathrm{d}b$$

Special case: the Gabor Transform

- In the Short Time Fourier Transform

$$Sf(\nu, b) = \int_{-\infty}^{+\infty} f(x)w(x-b)e^{-2i\pi\nu x}\,\mathrm{d}x = \langle f, \psi_{\nu,b}\rangle$$

  the analyzing functions are:

$$\psi_{\nu,b} = w(x-b)e^{2i\pi\nu x}$$

- In the **Gabor transform** (1946) the window w is a Gaussian of scale σ: $w(x) = \frac{1}{\sigma}e^{-\pi(\frac{x}{\sigma})^2}$ and the Gabor functions are then ($\sigma = 1$):

$$g_{\nu,b} = e^{-\pi(x-b)^2}e^{2i\pi\nu x}$$
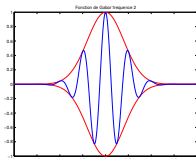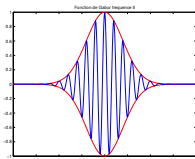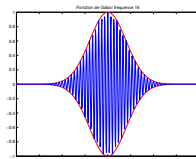
(a) $\nu = 2$ (b) $\nu = 5$ (c) $\nu = 15$

Short Time Fourier Transform

Example of two musical notes

The time-frequency analysis allows to recover both frequencies (the notes) and temporal information (the temporal order) of the signal f_1:
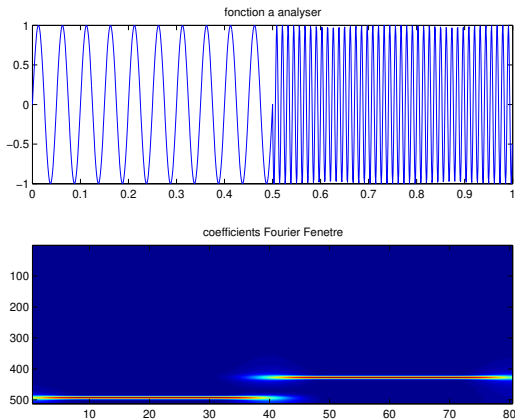
Figure: Time-frequency plane with b on the x-axis and ν on the y-axis, representing the density energy $|Sf(\nu, b)|^2 = |\langle f, g_{\nu,b} \rangle|^2$ called the spectrogram.

Short Time Fourier Transform

Analogy with music scores: an example with a piano
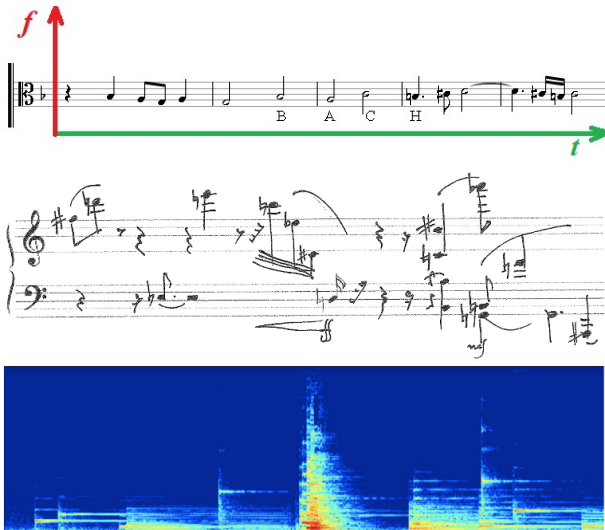
Credits: Patrick Flandrin, "Au-delà de Fourier, un monde qui vibre" (interstices.info)

Heisenberg boxes

Time-frequency localization and spread

$$g_{\xi,b}(t) = w(t-b)e^{i\xi t} \longleftrightarrow \widehat{g}_{\xi,b}(\omega) = \widehat{w}(\omega - \xi)e^{-ib(\omega - \xi)}$$

$$\sigma_t^2 = \int_{-\infty}^{\infty} (t-b)^2 |g_{\xi,b}(t)|^2 \mathrm{d}t = \int_{-\infty}^{\infty} t^2 |w(t)|^2 \mathrm{d}t$$

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\omega - \xi)^2 |\widehat{g}_{\xi,b}(\omega)|^2 \mathrm{d}\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |\widehat{w}(t)|^2 \mathrm{d}\omega$$
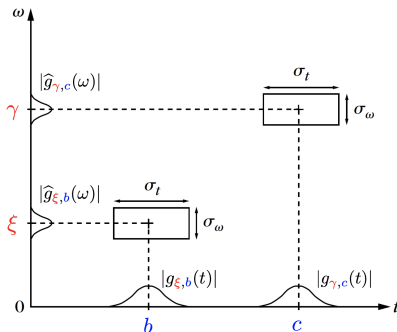
Heisenberg boxes

Example: Gabor limits
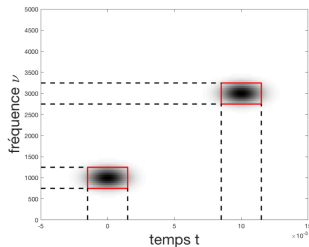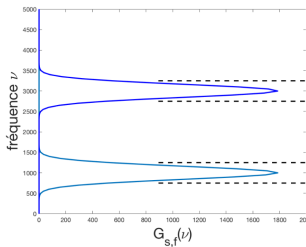
$$g_{s,f}(t) = w(t-s)e^{j2\pi ft}$$

$$w(t) = (\pi\sigma^2)^{-1/4}e^{-t^2/2\sigma^2}$$
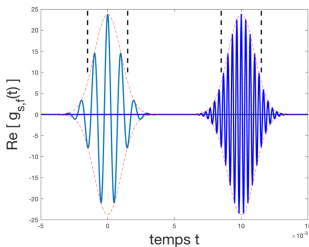
Credits: Pierre Chainais, "De la transformée de Fourier à l'analyse temps-fréquence bivariée"

Heisenberg boxes

Time-frequency localization and spread

- Can we construct a function f, with an energy that is highly localized in time and with a Fourier transform $\widehat{f}$ having an energy concentrated in a small-frequency interval?

- To reduce the time spread of f, we can scale it by $a < 1$, while keeping its total energy constant:

$$f_a(t) = \frac{1}{\sqrt{a}} f\left(\frac{t}{a}\right), \quad \|f_a\|^2 = \|f\|^2$$

- The corresponding Fourier transform is dilated by a factor $1/a$:

$$\widehat{f}_a(\omega) = \sqrt{a}\,\widehat{f}(a\omega)$$

$\Rightarrow$ So we lose in frequency localization what we gained in time. Underlying is a trade-off between time and frequency localization.

Credits: S. Mallat (Wavelet tour)

Heisenberg's indeterminacy relations

Defining the average location and frequency respectively by:

$$b = \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} t|f(t)|^2 \mathrm{d}t, \quad \xi = \frac{1}{2\pi\|f\|^2} \int_{-\infty}^{\infty} \omega|\widehat{f}(\omega)|^2 \mathrm{d}\omega$$

The variances around these average values are respectively:

$$\sigma_t^2 = \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} (t-b)^2|f(t)|^2 \mathrm{d}t, \quad \sigma_\omega^2 = \frac{1}{2\pi\|f\|^2} \int_{-\infty}^{\infty} (\omega-\xi)^2|\widehat{f}(\omega)|^2 \mathrm{d}\omega$$

Theorem (Heisenberg's indeterminacy relations)

The temporal variance and the frequency variance of $f \in L^2(\mathbb{R})$ satisfy

$$\boxed{\sigma_t\,\sigma_\omega \geqslant \frac{1}{2}}$$

This inequality is an equality iff $\exists(b, \xi, c_1, c_2) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that

$$f(t) = c_1 \mathrm{e}^{i\xi t - c_2(t-b)^2}$$

Credits: S. Mallat

Heisenberg's indeterminacy relations

Proof *(Weyl)*: this proof supposes that $\lim_{|t|\to+\infty} \sqrt{t} f(t) = 0$ $(*)$ but the theorem is valid for any $f \in L^2(\mathbb{R})$. The average time and frequency location of $e^{-i\xi t} f(t+b)$ is zero. Thus, it is sufficient to prove the theorem for $b = \xi = 0$. Since $\widehat{f'(t)}(\omega) = i\omega \widehat{f}(\omega)$, the Plancherel identity applied to $i\omega \widehat{f}(\omega)$ yields

$$\sigma_t^2 \sigma_\omega^2 = \frac{1}{\|f\|^4} \left(\int_{-\infty}^{\infty} |t\, f(t)|^2 \mathrm{d}t \right) \left(\int_{-\infty}^{\infty} |f'(t)|^2 \mathrm{d}t \right) \ (**)$$

Schwarz's inequality and the assumption $(*)$ [for the last equality] imply

$$
\begin{aligned}
\sigma_t^2 \sigma_\omega^2 &\geqslant \frac{1}{\|f\|^4} \left(\int_{-\infty}^{\infty} |t\, f'(t) f^*(t)| \mathrm{d}t \right)^2 \qquad \forall z \in \mathbb{C}, \ |z| \geqslant \mathrm{Re}(z) = \frac{z + z^*}{2} \\
&\geqslant \frac{1}{\|f\|^4} \left(\int_{-\infty}^{\infty} \frac{t}{2} (f'(t) f^*(t) + {f'}^*(t) f(t)) \, \mathrm{d}t \right)^2 \\
&\geqslant \frac{1}{4\|f\|^4} \left(\int_{-\infty}^{\infty} t(|f(t)|^2)' \, \mathrm{d}t \right)^2 \overset{\text{IBPF}}{=} \frac{1}{4\|f\|^4} \left(\int_{-\infty}^{\infty} |f(t)|^2 \mathrm{d}t \right)^2 = \frac{1}{4}
\end{aligned}
$$

Heisenberg's indeterminacy relations

Proof: To obtain an equality, Schwarz's inesquality applied to $(**)$ must be an equality. This implies that there exists $c_2 \in \mathbb{C}$ such that

$$f'(t) = -2c_1 t f(t)$$

Thus, there exists $c_1 \in \mathbb{C}$ such that

$$f(t) = c_1 e^{-c_2 t^2}$$

When $b \neq 0$ and $\xi \neq 0$ a time and frequency translation yield the result.

Remark: motivated by quantum mechanics, Gabor proposed time-frequency atoms that have a minimal spread in a time-frequency plane. By showing that signal decompositions over the dictionary of Gabor atoms are closely related to our perception of sounds, and that they exhibit important structures in speech and music recordings, he demonstrated the importance of localized time-frequency signal processing.
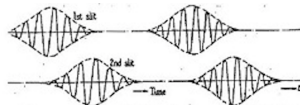
Credits: S. Mallat (Wavelet tour)

Heisenberg's indeterminacy relations
Some intuitions behind
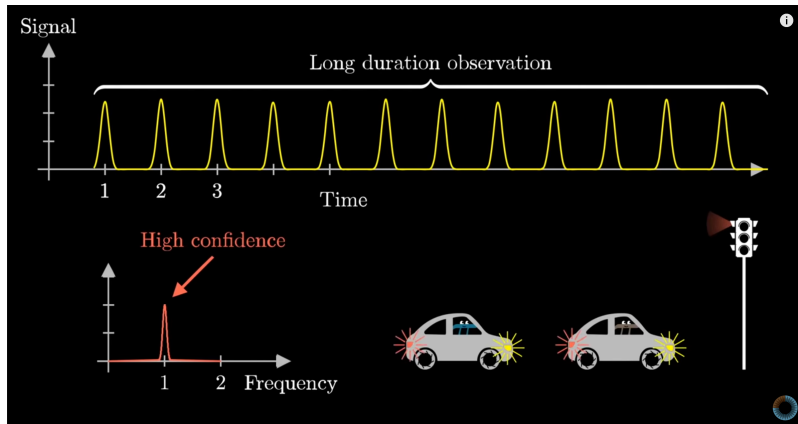
$$\mathrm{TF}(f \cdot \Pi_{[-a/2, a/2]}) = \mathrm{TF}(f) * \mathrm{TF}(\Pi_{[-a/2, a/2]}) = \mathrm{TF}(f) * a\,\mathrm{sinc}(\pi a \cdot)$$

Credits: 3blue1brown (http://www.youtube.com/watch?v=MBnnXbOM5S4)

Heisenberg's indeterminacy relations

Some intuitions behind
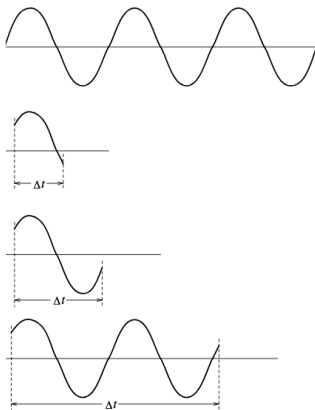
Figure: Improved frequency measurement over longer time intervals. The uncertainty in the frequency Δf decreases as the measurement interval Δt increases, and vice versa.

Credits: Bruce MacLennan (Gabor Representation)

Heisenberg's indeterminacy relations

Some intuitions behind
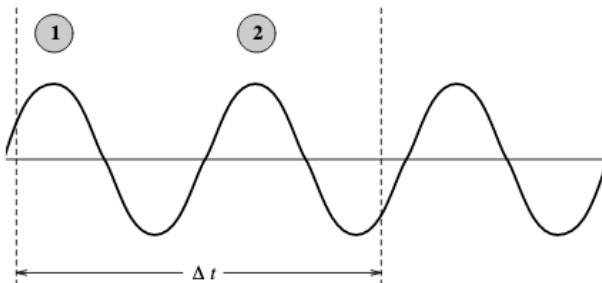
Figure: Measuring frequency by counting maxima in a given time interval. The circled numbers indicate the maxima counted during the measurement interval Δt. Since signals of other frequencies could also have the same number of maxima in that interval, there is an uncertainty Δf in the frequency.

Credits: Bruce MacLennan (Gabor Representation)

Heisenberg's indeterminacy relations

Some intuitions behind
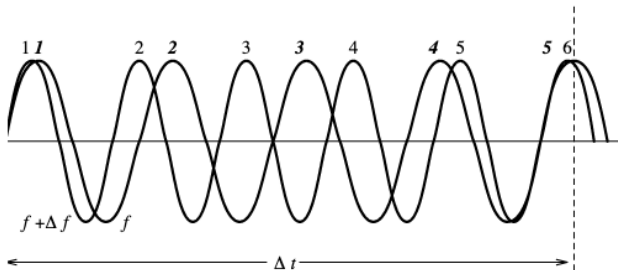
Figure: Minimum time interval Δt to detect frequency difference Δf. If two signals differ in frequency by Δf, then a measurement of duration $\Delta t \geqslant 1/\Delta f$ is required to guarantee a difference in counts of maxima. (Italic numbers indicate maxima of signal of frequency f, roman numbers indicate maxima of signal of higher frequency $f + \Delta f$)

$$(f + \Delta f)\Delta t - f\Delta t \geqslant 1 \iff \Delta f\,\Delta t \geqslant 1$$

Credits: Bruce MacLennan (Gabor Representation)

Short Time Fourier Transform

Examples

1. A sinusoidal wave $f(t) = e^{i\xi_0 t}$ whose Fourier transform is a Dirac $\widehat{f}(\omega) = 2\pi\delta(\omega - \xi_0)$ has a STFT:

$$Sf(\xi, b) = \widehat{w}(\xi - \xi_0)e^{-ib(\xi - \xi_0)}$$

Its energy is spread over the frequency interval

$$\xi \in [\xi_0 - \sigma_\omega/2, \xi_0 + \sigma_\omega/2]$$

2. A Dirac $f(t) = \delta(t - b_0)$ has a STFT:

$$Sf(\xi, b) = w(b - b_0)e^{-i\xi b_0}$$

Its energy is spread in the time interval

$$b \in [b_0 - \sigma_t/2, b_0 + \sigma_t/2]$$

Limitation of the Short Time Fourier Transform

The STFT cannot separate events of a distance smaller than a_0, that is to localize the two frequencies and the transient phenomena.
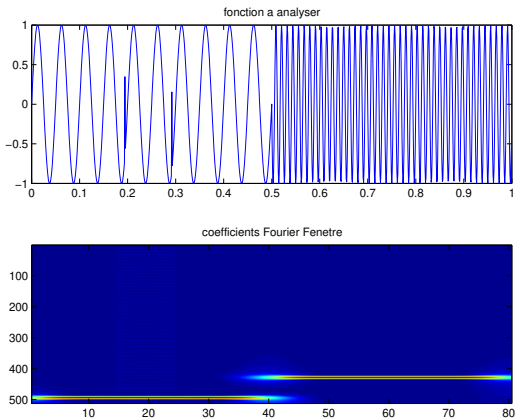
Figure: Signal $f_2 = f_1 + \delta_1 + \delta_2$ and its Gabor transform with $a_0 = 0.05$

Limitation of the Short Time Fourier Transform

The STFT cannot separate events of a distance smaller than a_0, that is to localize the two frequencies and the transient phenomena.
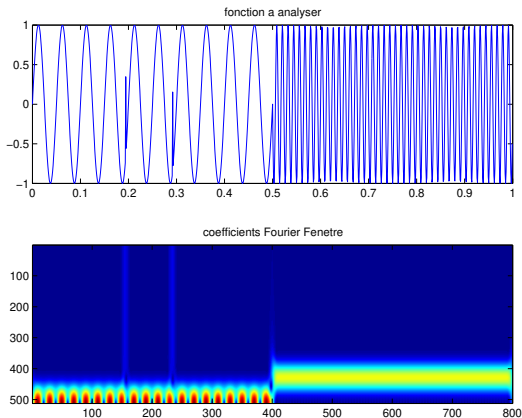
Figure: Signal $f_2 = f_1 + \delta_1 + \delta_2$ and its Gabor transform with $a_0 = 0.005$

Pioneer works on wavelets

- **Jean Morlet** research engineer at ELF Aquitaine discovered wavelets for solving signal processing problems arising from oil exploration.
- **Alex Grossmann** recognized in the Morlet wavelets something similar to coherent states formalism in quantum mechanics and developed an exact inversion formula for the wavelet transform.
- They developed the mathematics of the continuous wavelet transforms in their article: "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape" (1984)

Dennis Gabor   Alex Grossmann   Jean Morlet

"Gaborettes" vs Morlet wavelets
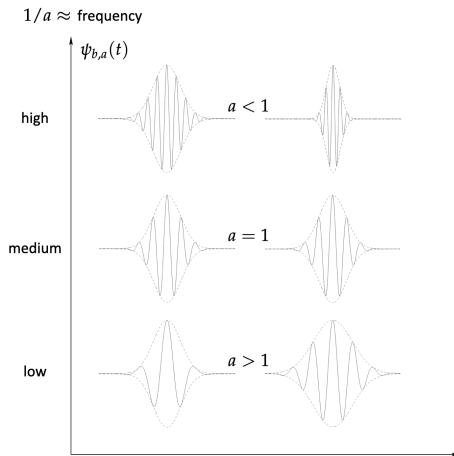
Figure: (Left) Gabor $\psi_{b,a}(t) = e^{it/a}\psi(t-b)$, (right) Morlet $\psi_{b,a}(t) = a^{-1/2}\psi\left(\frac{t-b}{a}\right)$

Gabor $\Rightarrow$ frequency modulation inside a constant window width
Wavelets $\Rightarrow$ shape of $\psi_{b,a}$ doesn't change, simply dilated or compressed

The Continuous Wavelet Transform (CWT) – Definition
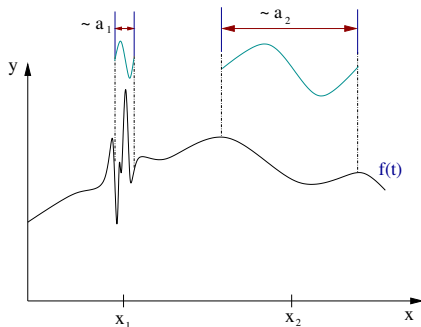
$$Wf(a, b) = \int_{-\infty}^{+\infty} f(x)\overline{\psi_{a,b}(x)}\,\mathrm{d}x = \langle f, \psi_{\nu,b}\rangle, \quad a > 0, \ b \in \mathbb{R}$$

The analyzing functions or **wavelets** are defined by:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x - b}{a}\right)$$

Wavelet family in physical space

Example: the Morlet wavelets

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$$

with mother wavelet

$$\psi(x) = \cos(x)\mathrm{e}^{-10\pi x^2}$$
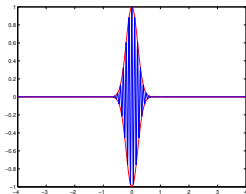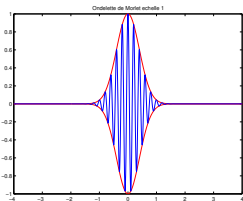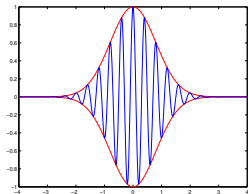
| (a) $a = 1/2$ | (b) $a = 1$ | (c) $a = 2$ |

Figure: Morlet wavelets of scale: $a = 1/2$, 1, 2 (real part). The scale a gives the support size (inverse of a frequency), whereas b gives the position.

Wavelet analysis of the toy signal with Morlet wavelets
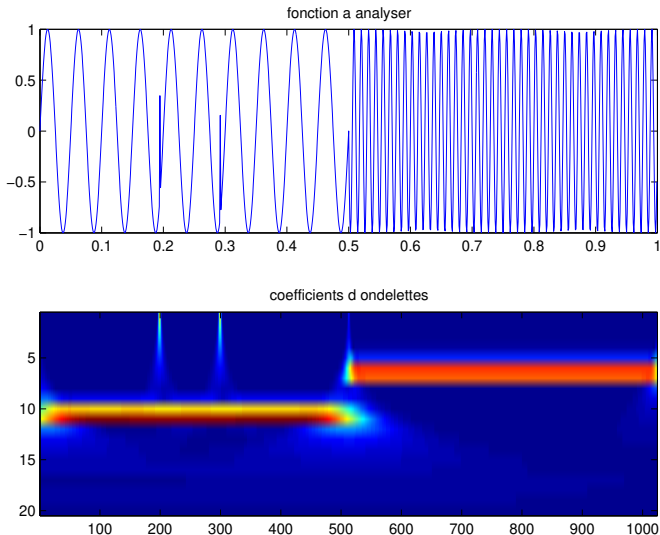
Figure: Signal f_2 (two notes + scratch) and its CWT

Wavelet definition

A function $\psi(x) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is a **wavelet** if it satisfies the following *admissiblity condition*:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{\left| \widehat{\psi}(\nu) \right|^2}{|\nu|} \, \mathrm{d}\nu \; < \; \infty$$

which implies $\int_{-\infty}^{+\infty} \psi(x) \, \mathrm{d}x = 0$ (and this is equivalent if $x\psi$ integrable).

Examples

1. The (complex) Morlet wavelet
   - Mother wavelet: $\psi(x) = \mathrm{e}^{-\pi x^2} \mathrm{e}^{10 i \pi x}$
   - Its Fourier Transform: $\hat{\psi}(\nu) = \mathrm{e}^{-\pi(\nu - 5)^2}$

2. Gaussian derivatives
   - Mother wavelet: $\psi_n(x) = \frac{\mathrm{d}^n}{\mathrm{d}x^n} \mathrm{e}^{-\pi x^2}, \quad n \geq 1$
     (for $n = 2$, ψ_2 is called the "Mexican Hat")
   - Its Fourier Transform: $\hat{\psi}_n(\nu) = (2 i \pi \nu)^n \mathrm{e}^{-\pi \nu^2}$

Wavelet analysis

A picture is worth a thousand words
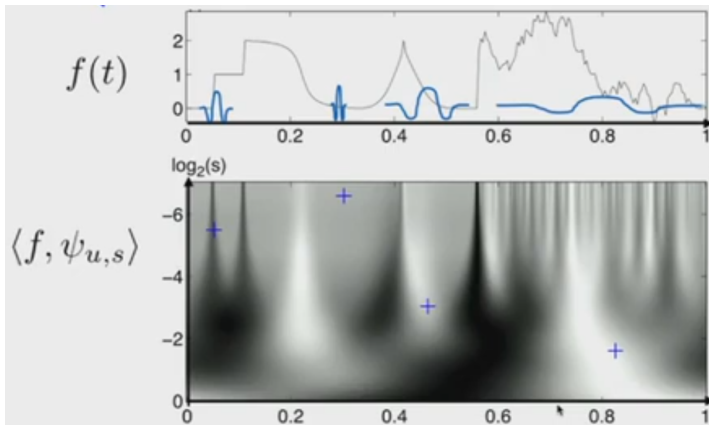
Figure: Correlations with Morlet wavelets translated and dilated

Fourier Transform of wavelets

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \longleftrightarrow \hat{\psi}_{a,b}(\nu) = \sqrt{a}\hat{\psi}(a\nu)e^{-2i\pi b\nu}$$

Transformee de Fourier des ondelettes echelle 2, 1, 1/2
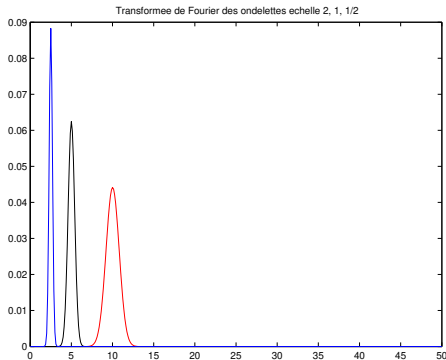
Figure: Fourier Transform (modulus) of Morlet wavelets of scales $a = 1/2,\ 1,\ 2$. Wavelets behaves as band-pass filters around frequency $\nu = \frac{\nu_0}{a}$, where ν_0 is the peak wavenumber (max of $\hat{\psi}$). For the Morlet wavelet, $\nu_0 = 5$.

Equivalent definition

Let $f \in L^2(\mathbb{R})$. Forall $a > 0$, $b \in \mathbb{R}$,

$$Wf(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \overline{\psi\left(\frac{x - b}{a}\right)} \, \mathrm{d}x$$

$$Wf(a, b) = \sqrt{a} \int_{-\infty}^{+\infty} \hat{f}(\nu) \overline{\hat{\psi}(a\nu)} \mathrm{e}^{2i\pi\nu b} \, \mathrm{d}\nu$$

Proof: *From the Parseval formula*

$$Wf(a, b) = \langle f, \psi_{a,b} \rangle = \langle \hat{f}, \hat{\psi}_{a,b} \rangle$$

- In the time domain (x), $Wf(a, b)$ provides information on the signal f around point b in a vicinity of size $\sim a$.
- In the frequency domain (ν), $Wf(a, b)$ provides information on the signal $\hat{f}$ around frequency $\sim \frac{1}{a}$.

$\Rightarrow$ Wavelet analysis is a time-scale analysis

Time-frequency resolution of wavelets

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \longleftrightarrow \hat{\psi}_{a,b}(\nu) = \sqrt{a}\hat{\psi}(a\nu)e^{-2i\pi b\nu}$$

We suppose that ψ is analytic, $\psi(0) = 0$ and $\eta = \frac{1}{2\pi}\int_0^\infty \omega|\widehat{\psi}(\omega)|^2 d\omega$

$$\int_{-\infty}^\infty (t-b)^2 |\psi_{a,b}(t)|^2 dt \stackrel{t \leftarrow \frac{t-b}{a}}{=} \int_{-\infty}^\infty a^2 t^2 |\psi(t)|^2 dt = a^2\sigma_t^2$$

$$\frac{1}{2\pi}\int_0^\infty \left(\omega - \frac{\eta}{a}\right)^2 |\widehat{\psi}_{\xi,b}(\omega)|^2 d\omega = \frac{1}{2\pi a^2}\int_0^\infty (\omega - \eta)^2 |\widehat{\psi}(\omega)|^2 d\omega = \frac{\sigma_\omega^2}{a^2}$$

The energy spread of a wavelet time-frequency atom $\psi_{a,b}$ corresponds to a Heisenberg box centered at $(b, \xi = \eta/a)$, of size $a\sigma_t$ along time and σ_ω/a along frequency. The area of the rectangle remains equal to $\sigma_t\sigma_\omega$ at all scales but the resolution in time and frequency depends on a. An analytic wavelet transform defines a local time-frequency energy density

$$P_W f(b, \xi) = |Wf(a,b)|^2 = \left|Wf\left(\frac{\eta}{\xi}, b\right)\right|^2 \quad (\text{scalogram})$$

Heisenberg boxes of two wavelets $\psi_{a,b}$ and ψ_{a_0,b_0}

$$\int_{-\infty}^{\infty} (t-b)^2 |\psi_{a,b}(t)|^2 \mathrm{d}t = a^2 \sigma_t^2$$

$$\frac{1}{2\pi} \int_0^{\infty} \left(\omega - \frac{\eta}{a}\right)^2 |\widehat{\psi}_{\xi,b}(\omega)|^2 \mathrm{d}\omega = \frac{\sigma_\omega^2}{a^2}$$
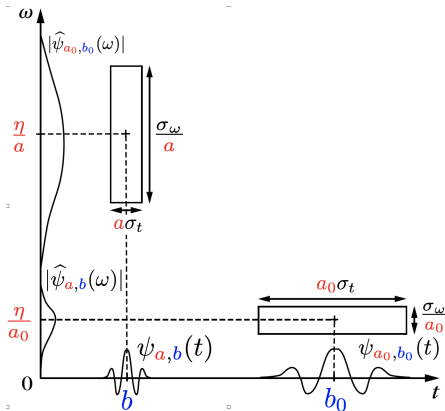
Inversion of the Continuous Wavelet Transform

Synthesis formula and energy conservation

Theorem (Calderón, Grossmann and Morlet)

Let $\psi \in L^2(\mathbb{R})$ be a real function such that

$$C_\psi = \int_{-\infty}^{+\infty} \frac{\left|\widehat{\psi}(\nu)\right|^2}{|\nu|} \, \mathrm{d}\nu \; < \; \infty$$

Any $f \in L^2(\mathbb{R})$ satisfies

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(a, b)\psi_{a,b}(x) \frac{\mathrm{d}a \, \mathrm{d}b}{a^2} \; (*)$$

and

$$\int_{-\infty}^{+\infty} |f(x)|^2 \mathrm{d}x = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(a, b)|^2 \frac{\mathrm{d}a \, \mathrm{d}b}{a^2} \; (**)$$

Proof *(Synthesis formula):* For a fixed a, the CWT can be written:

$$Wf(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \, \overline{\psi\left(\frac{x-b}{a}\right)} \, \mathrm{d}x = (f * \check{\psi}_a)(b)$$

where we have noted:

$$\psi_a(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x}{a}\right), \quad \check{\psi}_a(x) = \psi_a(-x)$$

The right integral $b(x)$ of $(*)$ can now be rewritten as a sum of convolutions:

$$b(x) = \frac{1}{C_\psi} \int_0^{+\infty} Wf(a, \cdot) * \psi_a(x) \frac{\mathrm{d}a}{a^2} = \frac{1}{C_\psi} \int_0^{+\infty} f * \check{\psi}_a * \psi_a(x) \frac{\mathrm{d}a}{a^2}$$

$$\widehat{b}(\omega) = \frac{1}{C_\psi} \int_0^{+\infty} \widehat{f}(\omega) \sqrt{a} \, \widehat{\psi}^*(a\omega) \sqrt{a} \widehat{\psi}(a\omega) \frac{\mathrm{d}a}{a^2} = \frac{\widehat{f}(\omega)}{C_\psi} \int_0^{+\infty} |\widehat{\psi}(a\omega)|^2 \frac{\mathrm{d}a}{a^2}$$

By the change of variable $\xi = a\omega$ we get $\widehat{b}(\omega) = \frac{\widehat{f}(\omega)}{C_\psi} \int_0^{+\infty} \frac{|\widehat{\psi}(\xi)|^2}{\xi} \, \mathrm{d}\xi = \widehat{f}(\omega)$

The equality of their Fourier transform leads to $b = f$. QED $\qquad\qquad \square$

Inversion with a different synthesis wavelet

- **Decomposition** with an *analysing wavelet g*: $a > 0$, $b \in \mathbb{R}$,

$$W_g f(a, b) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{a}} \bar{g} \left(\frac{x - b}{a} \right) \mathrm{d}x$$

- **Synthesis** with a *reconstruction wavelet h*:

$$f(x) = \frac{2}{c_{gh}} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_g f(a, b) \ \frac{1}{\sqrt{a}} h \left(\frac{x - b}{a} \right) \ \frac{\mathrm{d}a}{a^2} \mathrm{d}b$$

- **Cross-admissibility condition** on wavelets g et h ($g, h \in L^2(\mathbb{R})$):

$$c_{gh} = \int_{-\infty}^{+\infty} \frac{\bar{\hat{g}}(k)\hat{h}(k)}{|k|} \, \mathrm{d}k < +\infty$$

Remark: In this case, only h or g has to be a zero mean function.

Coding – In practice

For a fixed a, the CWT is a convolution product:

$$
\begin{aligned}
Wf(a, b) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \, \overline{\psi\left(\frac{x - b}{a}\right)} \, \mathrm{d}x \\
&= (f * \check{\psi}_a)(b)
\end{aligned}
$$

where we have noted:

$$
\check{\psi}_a(x) = \frac{1}{\sqrt{a}} \psi\left(-\frac{x}{a}\right)
$$
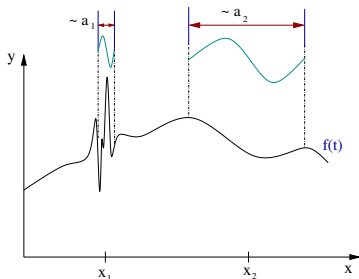
Coding – In practice

$$f * \check{\psi}_a : b \mapsto Wf(a, b)$$

3D plot of Wavelet Transform

2D plot of Wavelet Transform
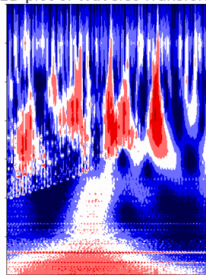
source: ataspinar.com

Examples and Interpretation

Example 1: pure cosine

If f is a **pure cosine** $f(x) = \cos(2\pi k x)$, then

$$
\begin{aligned}
Wf(a, b) &= \int_{-\infty}^{+\infty} \left(\frac{e^{2i\pi kx} + e^{-2i\pi kx}}{2} \right) \overline{\psi_{a,b}(x)} \, dx \\
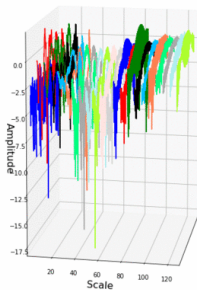&= \frac{1}{2} \left[\overline{\hat{\psi}_{a,b}(k)} + \overline{\hat{\psi}_{a,b}(-k)} \right] \\
&= \frac{\sqrt{a}}{2} \left[\overline{\hat{\psi}(ak)} e^{2i\pi kb} + \overline{\hat{\psi}(-ak)} e^{-2i\pi kb} \right]
\end{aligned}
$$

- If the wavelet ψ is analytic complex:

$$
Wf(a, b) = \frac{\sqrt{a}}{2} \, \overline{\hat{\psi}(ak)} e^{2i\pi kb}
$$

- If the wavelet ψ is real $Wf(a, b) = \sqrt{a} \, \mathrm{Re}\left(\hat{\psi}(ak) e^{-2i\pi kb} \right)$

Example 1: pure cosine

Pure cosine $f(x) = \cos(20\pi x)$ ($k = 10$)
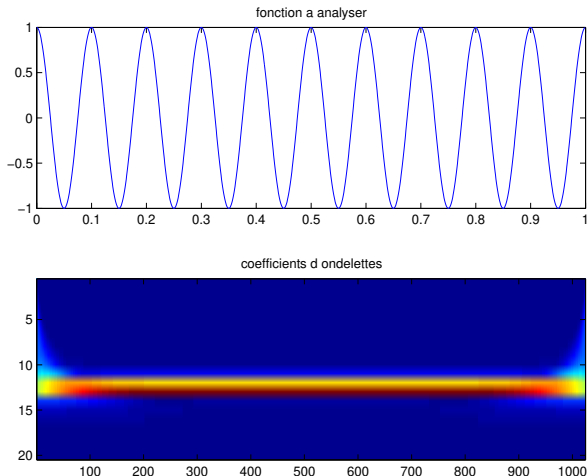
Figure: CWT (modulus), using the Morlet wavelet (analytic complex)

Example 1: pure cosine

Pure cosine $f(x) = \cos(20\pi x)$ ($k = 10$)
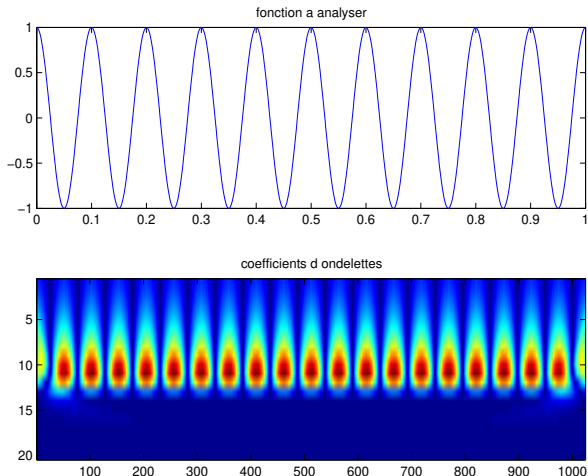
Figure: CWT (modulus), using the Gaussian derivatives (real)

Examples and Interpretation

If f is a **Dirac** $f(x) = \delta(x - x_0)$ (pointwise measure supported by x_0), then:

$$
\begin{aligned}
Wf(a, b) &= \int_{-\infty}^{+\infty} \delta(x - x_0) \, \overline{\psi_{a,b}(x)} \, \mathrm{d}x \\
&= \overline{\psi_{a,b}(x_0)} \\
&= \frac{1}{\sqrt{a}} \, \overline{\psi\left(\frac{x_0 - b}{a}\right)}
\end{aligned}
$$

Remark: At each scale a, $b \to Wf(a, b)$ is the wavelet of scale a centered on x_0 (up to a symmetry).

Example 2: a Dirac

Dirac δ_{x_0}
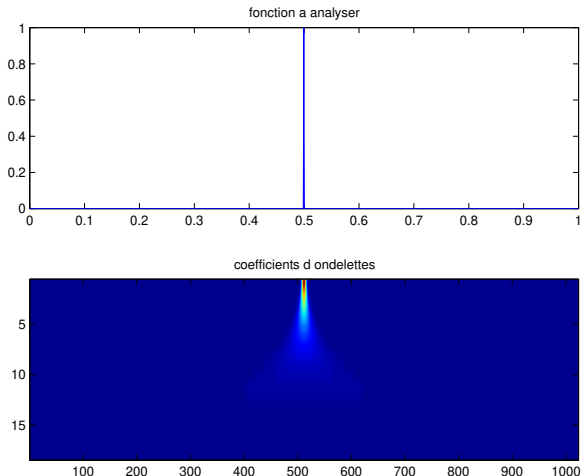
Figure: Signal "Dirac" and its CWT (modulus, Morlet wavelet, divided by $\sqrt{a}$)

Example 3: the periodic square wave

Periodic square wave
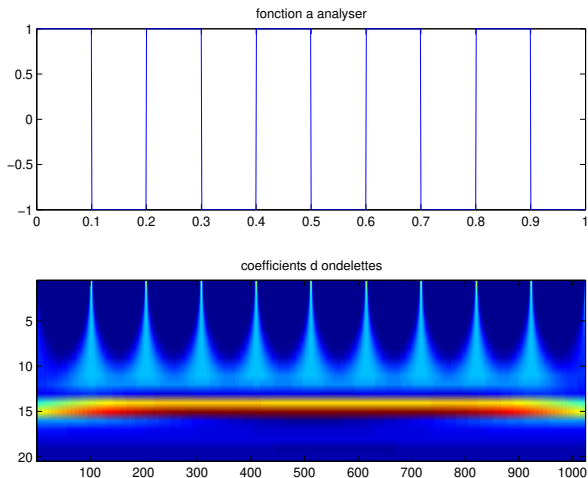
Figure: Square wave and its CWT (modulus, Morlet wavelet, divided by $\sqrt{a}$)

Example 4: a modulated wave

Modulated wave
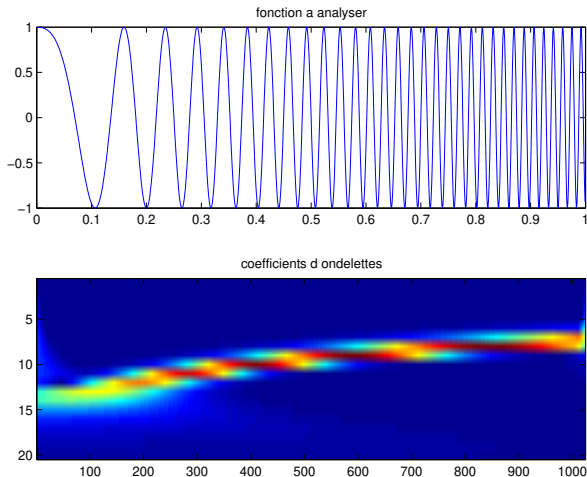
Figure: Modulated wave and its CWT (modulus, Morlet wavelet, divided by $\sqrt{a}$)

Example 5: 2 sinusoids with noise

2 sinusoids with noise
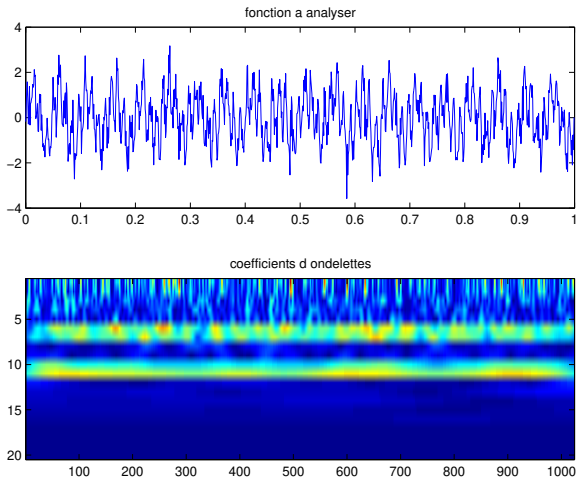
Figure: Signal and its CWT (modulus, Morlet wavelet, divided by $\sqrt{a}$)

Example 6: Holder function of exponant $\frac{1}{2}$

Holder function of exponant $\frac{1}{2}$
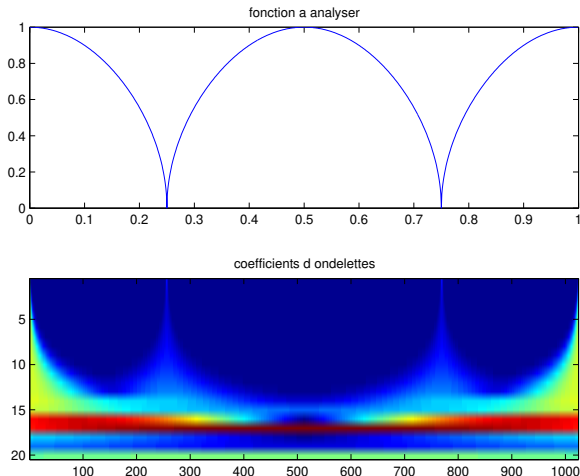
Figure: $f(x) = \sqrt{|\cos(2\pi x)|}$ and its CWT (modulus, Morlet wavelet, divided by $\sqrt{a}$)

The Dyadic Wavelet Transform
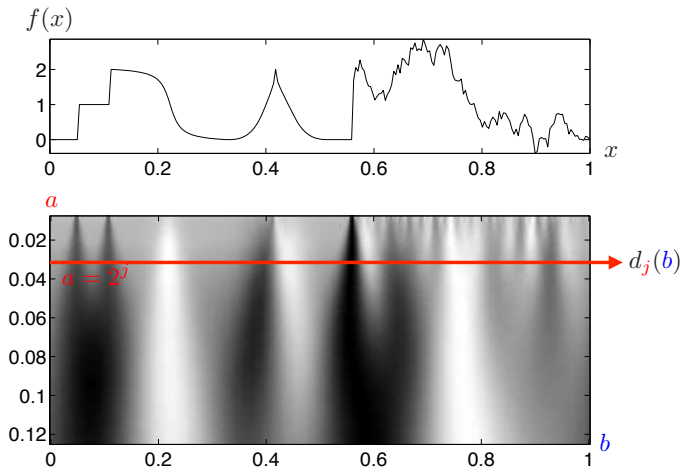
The Dyadic Wavelet Transform

For a fixed a, the CWT is a convolution product:

$$
\begin{aligned}
Wf(a, b) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\, \overline{\psi\left(\frac{x - b}{a}\right)}\, \mathrm{d}x \\
&= (f * \check{\psi}_a)(b)
\end{aligned}
$$

where we have noted:

$$
\psi_a(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x}{a}\right), \quad \check{\psi}_a(x) = \psi_a(-x)
$$

The Dyadic Wavelet Transform

For a fixed $a = 2^j$, the Dyadic Wavelet Transform is a convolution product:

$$
\begin{aligned}
Wf(2^j, b) &= \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} f(x) \, \overline{\psi\left(\frac{x-b}{2^j}\right)} \, \mathrm{d}x \\
d_j(b) &= (f * \check{\psi}_j)(b)
\end{aligned}
$$

where we have noted:

$$
\psi_j(x) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{x}{2^j}\right), \quad \check{\psi}_j(x) = \psi_j(-x)
$$
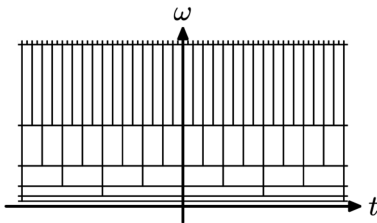
The Dyadic Wavelet Transform

For a fixed $a = 2^j$, the Dyadic Wavelet Transform is a convolution product:

$$
\begin{aligned}
Wf(2^j, b) &= \frac{1}{\sqrt{2^j}} \int_{-\infty}^{+\infty} f(x) \, \overline{\psi\left(\frac{x-b}{2^j}\right)} \, \mathrm{d}x \\
d_j(b) &= (f * \check{\psi}_j)(b)
\end{aligned}
$$

whose Fourier transform is

$$
\widehat{d}_j(\omega) = \widehat{f}(\omega)\sqrt{2^j}\,\widehat{\psi}^*(2^j\omega)
$$
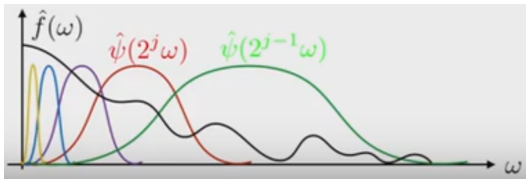
The Dyadic Wavelet Transform

Theorem (Littlewood-Paley, 1930)

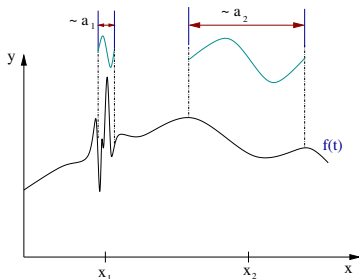If $\sum_j |\widehat{\psi}(2^j\omega)|^2 = 1$ then

$$f(x) = \sum_j 2^{-j} \int Wf(2^j, b)\psi_{2^j, b}(x)\,\mathrm{d}b$$

Proof: Remark that

$$\int Wf(2^j, b)\psi_{2^j, b}(x)\,\mathrm{d}b = d_j * \psi_j(x)$$

then take the Fourier transform

$$
\begin{aligned}
\sum_j 2^{-j}\widehat{d}_j(\omega)\widehat{\psi}_j(\omega) &= \sum_j 2^{-j}\widehat{f}(\omega)\sqrt{2^j}\widehat{\psi}^*(2^j\omega)\sqrt{2^j}\psi(2^j\omega) \\
&= \widehat{f}(\omega)\underbrace{\sum_j |\widehat{\psi}(2^j\omega)|^2}_{=1} = \widehat{f}(\omega)
\end{aligned}
$$

Take home message

Time-frequency vs time-scale
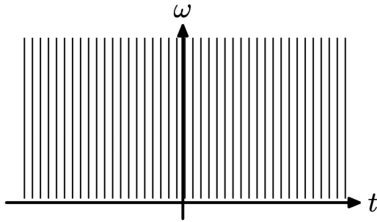
(a) Diracs
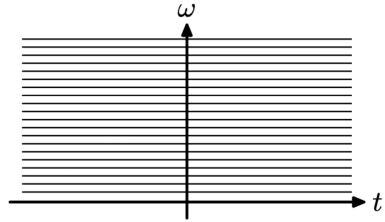
(b) Fourier
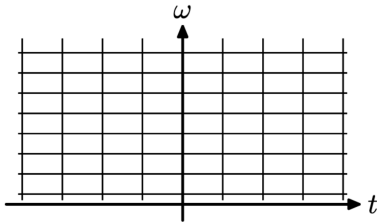
(c) STFT
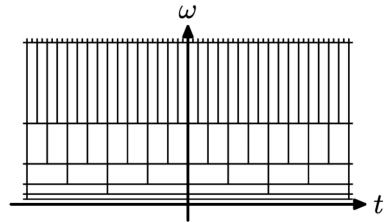
(d) Wavelets dyadics

Take home message

Heisenberg for wavelets
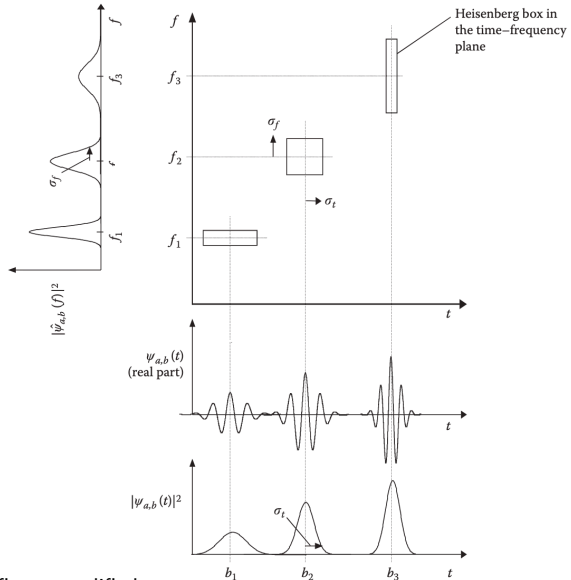
Credits: Paul S Addison's figure modified

Take home message

Scalogram construction
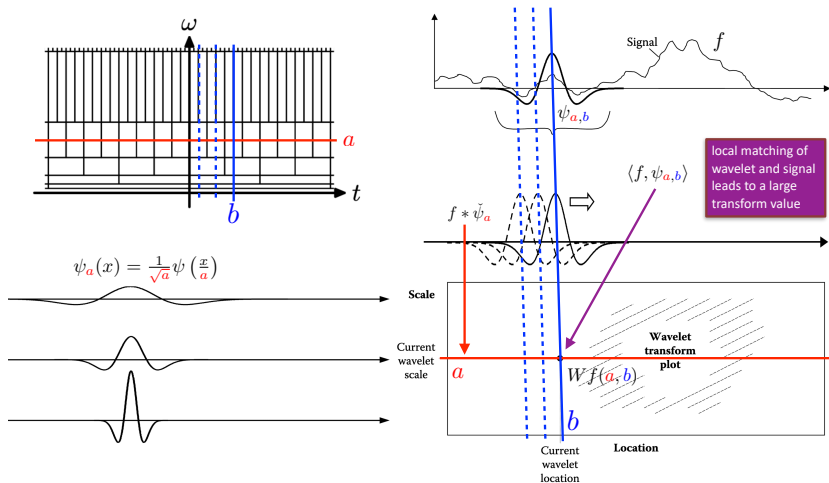
Credits: Paul S Addison's figure adapted

Wavelet zoom
a local characterization of functions

Local characterization of regularity via the derivatives

"Smoothness" depends on the differentiability class to which a function belongs to. Among these 4 continuous ($\mathcal{C}^0$) functions:

- $x \mapsto x^2$ is the only one differentiable everywhere and $\mathcal{C}^\infty$
- $x \mapsto |x|$ is not differentiable at $x = 0$ (corner)
- $x \mapsto \sqrt{|x|}$ (cusp) and $\mathbb{1}_{[0,T]}$ (jump) have kind of "infinite gradient" at the singularity point $x = 0$
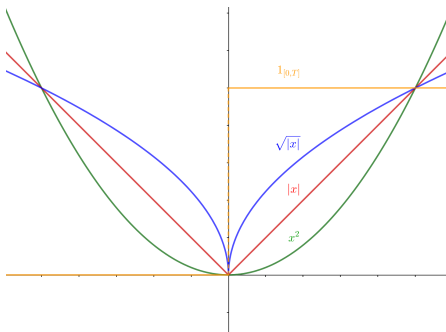
Prerequisite: Global regularity through Fourier coefficients

Lemma (Riemann-Lebesgue)

If f is L^1 then the Fourier transform of f satisfies

$$\widehat{f}(\omega) = \int f(x)\mathrm{e}^{-i\omega x} \xrightarrow[|\omega|\to\infty]{} 0$$

How fast the Fourier coefficients decrease?

For f p times continuously differentiable with bounded derivatives, since $\widehat{f}(\omega) = \frac{1}{i\omega}\widehat{\frac{\mathrm{d}}{\mathrm{d}x}f}(\omega)$ then by iterating we get $\widehat{f}(\omega) = \frac{1}{(i\omega)^p}\widehat{\frac{\mathrm{d}^p}{\mathrm{d}x^p}f}(\omega)$

$$|\widehat{f}(\omega)| \leqslant \frac{K}{|\omega|^p}$$

with $K = \sup \frac{\mathrm{d}^p}{\mathrm{d}x^p}f$

Prerequisite: Global regularity through Fourier coefficients

Conversely Fourier decay governs smoothness?

If $\widehat{f}$ is L^1 then $f \in L^\infty$ and f is continuous.

Proof:

$$|f(x)| \leqslant \frac{1}{2\pi} \int |e^{i\omega x} \widehat{f}(\omega)| \mathrm{d}\omega \leqslant \frac{1}{2\pi} \int |\widehat{f}(\omega)| \mathrm{d}\omega < +\infty$$

which proves boundedness. As for continuity, consider a sequence $y_n \to 0$ and

$$f(x - y_n) = \frac{1}{2\pi} \int e^{i\omega(x - y_n)} \widehat{f}(\omega) \, \mathrm{d}\omega$$

The integrand converges pointwise to $e^{i\omega x} \widehat{f}(\omega)$ and is uniformly bounded in modulus by the integrable function $\widehat{f}$. Hence Lebesgue's dominated convergence theorem applies and yields $f(x - y_n) \to f(x)$ that is continuity in x. $\qquad \square$

Prerequisite: Global regularity through Fourier coefficients

Theorem (Sufficiant condition for differentiability of f at order p)

A function f is bounded and p times continuously differentiable with bounded derivatives if

$$\int_{-\infty}^{\infty} |\widehat{f}(\omega)|(1 + |\omega|^p)\, \mathrm{d}\omega < +\infty$$

Proof: Knowing that $\widehat{f^{(k)}} : \omega \mapsto (i\omega)^k \widehat{f}(\omega)$, by the inversion formula

$$|f^{(k)}(t)| = \left| \int_{-\infty}^{\infty} \widehat{f^{(k)}}(\omega) \mathrm{e}^{i\omega t}\, \mathrm{d}\omega \right| \leqslant \int_{-\infty}^{\infty} |\widehat{f}(\omega)| \cdot |\omega|^k \mathrm{d}\omega < +\infty$$

for any $k \leqslant p$, so $f^{(k)}$ is continuous and bounded. $\qquad \square$

Corrolary. If it exists a constant K and $\epsilon > 0$ such that

$$|\widehat{f}(\omega)| \leqslant \frac{K}{1 + |\omega|^{p+1+\epsilon}}, \quad \text{then} \quad f \in \mathcal{C}^p$$

Credits: S. Mallat (Wavelet tour)

Prerequisite: Global regularity through Fourier coefficients

The decay of $|\widehat{f}(\omega)|$ depends on the worst singular behavior of f
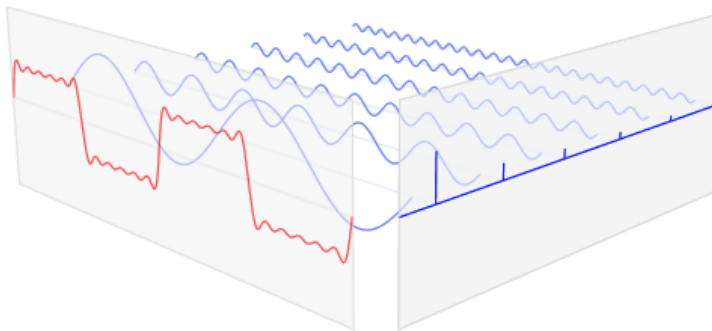
$$f(x) = \begin{cases} -1 \text{ if } -\pi \leqslant x < 0 \\ +1 \text{ if } 0 \leqslant x < \pi \end{cases} = \sum_{n=1}^{+\infty} \frac{4}{\pi(2n-1)} \sin((2n-1)x)$$

where f is <u>periodized</u>. For $f = \mathbb{1}_{[-T,T]} \Rightarrow |\widehat{f}(\omega)| = o(|\omega|^{-1})$

Credits: Wikipedia (https://en.wikipedia.org/wiki/Fourier_series)

Wavelet zoom: Lipschitz regularity

Definition (Lipschitz regularity of order α of a function f)

Let $\alpha \geq 0$ be the regularity parameter and $x_0 \in \mathbb{R}$.
f is pointwise Lipschitz–α at x_0, if there exist $C > 0$ and a polynomial P_n of degree $n = \lfloor \alpha \rfloor$, such that

$$\forall h \in \mathbb{R}, \quad |f(x_0 + h) - P_n(h)| \leq C|h|^\alpha \tag{1}$$

P_n is the Taylor expansion of f at x_0. (If $0 < \alpha < 1$, $P_0(h) = f(x_0)$)

- f is uniformly Lipschitz–α over $[a, b]$ if f satisfies (1) for all $x_0 \in [a, b]$, with a constant C independent of x_0.
- Extension to negative α (distributions): f uniformly Lipschitz–α over $]a, b[$ if its primitive is Lipschitz–$(\alpha + 1)$ over $]a, b[$.
- The Lipschitz regularity of f is the supremum of the α such that f is Lipschitz–α.

Lipschitz–α functions
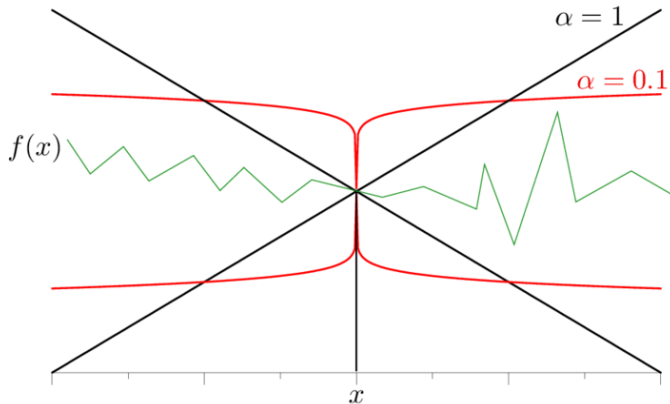
Figure: The schematic diagram of Lipschitz–α functions

Credits: Li-Wei Liu & Hong-Ki Hong

Some examples

- A Lipschitz–α function at x_0, with $0 < \alpha < 1$, is continuous, but a priori non differentiable.
- A $\mathcal{C}^1$ function in a neighborhood of x_0 is Lipschitz–1 at x_0.
- The Lipschitz regularity α with $n < \alpha < n+1$ allows to classify regularities between $\mathcal{C}^n$ and $\mathcal{C}^{n+1}$.
- A bounded function is Lipschitz–0. For example the Heavyside function $H(x) = 1$ if $x \geq 0$ and 0 if $x < 0$.
- The distribution δ is Lipschitz–(-1) (as the derivative of H).
- The function $x \mapsto |x - x_0|^{\alpha}$ $(0 < \alpha < 1)$ is Lipschitz–α
- The function $\sqrt{|\cos(2\pi x)|}$ is Lipschitz–$\frac{1}{2}$.
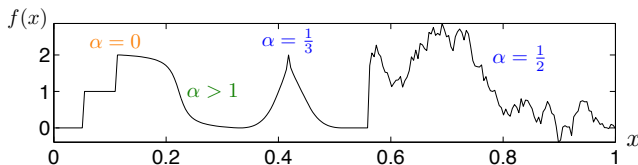
Some examples

A Holder function of exponant $\frac{1}{2}$
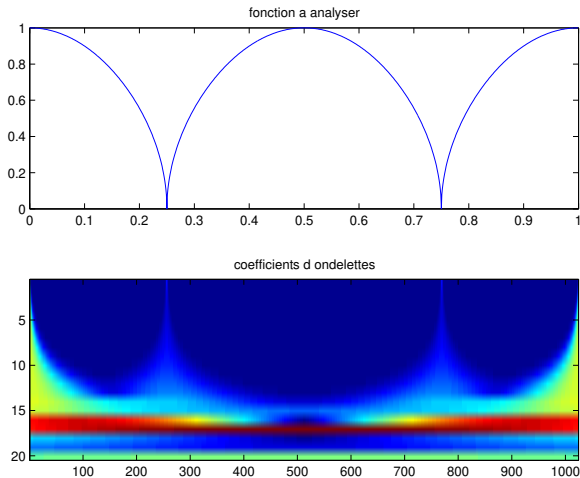
Figure: $f(x) = \sqrt{|\cos(2\pi x)|}$ and its CWT (modulus, Morlet wavelet, divided by $\sqrt{a}$)

Some examples
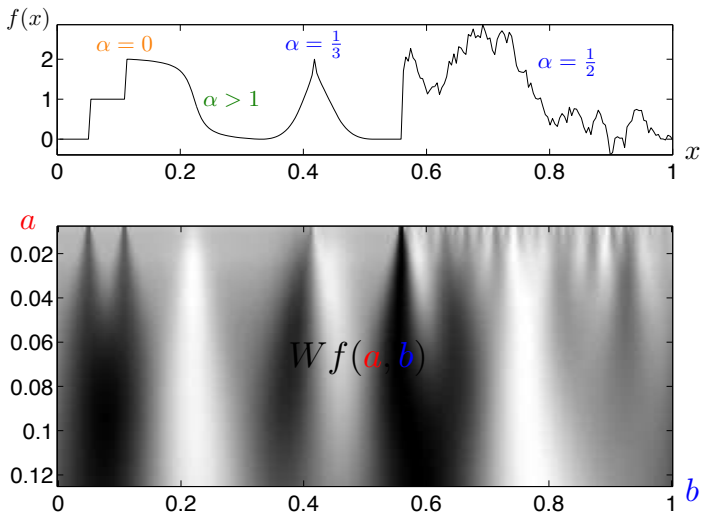
Figure: Wavelet transform $Wf(a, b)$ calculated with $\psi = -\theta'$ where θ is a Gaussian. Singularities create large amplitude coefficients in their cone influence.

Credits: S. Mallat (Wavelet tour)

Regularity measurements with wavelets

Let $\alpha \geq 0$ be fixed, ψ a wavelet with compact support $\subset [-L, L]$, and $N > \alpha$ vanishing moments:

$$\int x^n \psi(x) \, \mathrm{d}x = 0, \quad \text{for } 0 \leqslant n < N$$

Remark: a wavelet with N vanishing moments is orthogonal to polynomials of degree $N - 1$.

Polynomial Suppression. Let f Lipschitz-α at x_0, that is

$$f(x) = P_n(x - x_0) + \varepsilon(x - x_0) \quad \text{with} \quad |\varepsilon(x - x_0)| \leq |x - x_0|^{\alpha}$$

Since $\alpha < N$, the polynomial P_N has degree at most $N - 1$. With the change of variable $y = (x - b)/a$, we verify that

$$WP_n(a, b) = \int_{-\infty}^{+\infty} P_n(x) \frac{1}{\sqrt{a}} \psi\left(\frac{x - b}{a}\right) \mathrm{d}x = 0$$

Then,

$$Wf(a, b) = W\varepsilon(a, b)$$

Pointwise Lipschitz regularity and wavelet coefficients

Let $\alpha \geq 0$. One consider a wavelet ψ of regularity $\mathcal{C}^N$, with compact support $\operatorname{supp} \psi \subset [-L, L]$, and $N \geq \alpha$ vanishing moments.

Theorem (Jaffard, Estimation of the local regularity of f at point x_0)

If $f \in L^2(\mathbb{R})$ is Lipschitz–$\alpha \leq N$ at x_0, then $\exists A > 0$ such that

$$\forall (a, b) \in \mathbb{R} \times \mathbb{R}^+, \quad |Wf(a, b)| \leqslant A \, a^{\alpha + \frac{1}{2}} \left(1 + \left| \frac{b - x_0}{a} \right|^{\alpha} \right)$$

Conversely, if $\alpha < N$ is not an integer and there exist $A > 0$ and $\alpha' < \alpha$ such that

$$\forall (a, b) \in \mathbb{R} \times \mathbb{R}^+, \quad |Wf(a, b)| \leqslant A \, a^{\alpha + \frac{1}{2}} \left(1 + \left| \frac{b - x_0}{a} \right|^{\alpha'} \right)$$

then f is Lipschitz-α at x_0.

Proof of $\boxed{\Rightarrow}$

Since f is Lipschitz–α at x_0, there exists a polynomial P_N of degree $\lfloor \alpha \rfloor < N$ and $C > 0$ such that

$$|f(x) - P_N(x - x_0)| \leqslant C|x - x_0|^\alpha$$

Since ψ has N vanishing moments, we saw that $WP_n(a, b) = 0$, and thus

$$
\begin{aligned}
|Wf(a, b)| &= \left| \int_{-\infty}^{\infty} [f(x) - P_N(x - x_0)] \psi_{a,b}(x) \, \mathrm{d}x \right| \\
&\leqslant \int C|x - x_0|^\alpha \frac{1}{\sqrt{a}} \left| \psi\left(\frac{x - b}{a}\right) \right| \mathrm{d}x
\end{aligned}
$$

The change of variable $y = \frac{x - b}{a}$ gives

$$|Wf(a, b)| \leqslant \sqrt{a} \int_{-\infty}^{\infty} C|ay + b - x_0|^\alpha |\psi(y)| \mathrm{d}y$$

Proof of $\boxed{\Rightarrow}$

$$|Wf(a,b)| \leqslant \sqrt{a} \int_{-\infty}^{\infty} C|\underbrace{ay}_{t} + \underbrace{b-x_0}_{s}|^{\alpha}|\psi(y)|\mathrm{d}y$$

Lemma: $|t+s|^{\alpha} \leqslant 2^{\alpha}(|t|^{\alpha}+|s|^{\alpha})$

Proof: Let $m = \max(|t|,|s|)$ so that $|t+s| \leqslant |t|+|s| \leqslant 2m$. Then,

$$|t+s|^{\alpha} \leqslant (2m)^{\alpha} = 2^{\alpha}m^{\alpha} \leqslant 2^{\alpha}(|t|^{\alpha}+|s|^{\alpha}) .$$

By the lemma,

$$
\begin{aligned}
|Wf(a,b)| &\leqslant C2^{\alpha}\sqrt{a}\left(a^{\alpha}\int_{-\infty}^{\infty}|y|^{\alpha}|\psi(y)|\mathrm{d}y + |b-x_0|^{\alpha}\int_{-\infty}^{\infty}|\psi(y)|\mathrm{d}y\right) \\
&\leqslant \underbrace{KM2^{\alpha}}_{A} s^{\alpha+\frac{1}{2}}\left(1 + \left|\frac{b-x_0}{a}\right|^{\alpha}\right)
\end{aligned}
$$

with $M = \max\left(\int_{-\infty}^{\infty}|y|^{\alpha}|\psi(y)|, \int_{-\infty}^{\infty}|\psi(y)|\mathrm{d}y\right)$. $\qquad\square$

Cone of Influence

If $\text{supp } \psi = [-L, L]$, the cone of influence of x_0 in the time-scale space is the set of points such that $x_0 \in \text{supp } \psi_{a,b} = [b - La, b + La]$, that is

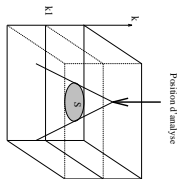$$\Gamma(x_0) = \{(b, a) \in \mathbb{R} \times \mathbb{R}_+^* \; : \; |b - x_0| < La\}$$

If f is Lipschitz-α at x_0, then $\exists A > 0$, such that for all $(b, a) \in \Gamma(x_0)$:

$$|Wf(a, b)| \leq A \, a^{\alpha + \frac{1}{2}}$$

and conversely for α non integer.

α is computed by the slope of the curve $\log a \to \log|Wf(a, b)|$

# Wavelet Transform Modulus Maxima

References

- S. Mallat, W.L. Hwang *Singularity detection and processing with wavelet*, IEEE Trans. Info. Theory, 38(2):617-643, Mars 1992

- S. Mallat, S.Zhong *Characterization of Signals from Multiscale Edges*, IEEE Trans. Patt. Anal. and Mach. Intell., 14(7):710-732, Juillet 1992

Wavelet construction from the derivatives of a Gaussian

Let $\theta(x) = \exp(-x^2/\sigma^2)$ the Gaussian Kernel and let considered

$$\psi^N(x) \equiv \theta^{(n)}(x) = \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^N e^{-\frac{x^2}{\sigma^2}}$$

The wavelet ψ^N has N vanishing moments.
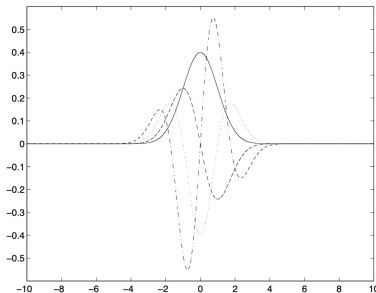
Figure: The Gaussian θ ($n = 0$) for $\sigma = 1$ and its two first derivatives: $n = 1$ is represented in $(- \cdot -)$ and $n = 2$ (the Mexican hat) in $(\cdots)$

Multiscale differential operator

A wavelet ψ has fast decay if

$$\forall m \in \mathbb{N}, \quad \exists C_m \quad \text{such that} \quad |\psi(x)| \leqslant \frac{C_m}{1 + |t|^m}, \quad \forall t \in \mathbb{R}$$

Theorem (Multiscale differential operator)

A wavelet ψ with fast decay has N vanishing moments if and only if there exists θ with a fast decay such that

$$\psi(x) = (-1)^N \frac{\mathrm{d}^N \theta}{\mathrm{d}x^N}(x)$$

As a consequence

$$W_N f(a, b) = a^N \frac{\mathrm{d}^N \theta}{\mathrm{d}b^N}(f * \check{\theta}_a)(b)$$

Moreover, ψ has no more vanishing moments iff $\int \psi \neq 0$.

Multiscale differential operator

Example

- The convolution $f * \breve{\theta}_a$ averages f over a domain proportional to a

- If the wavelet has only one vanishing moment: $\psi = -\theta'$ then $W_1(a, b) = a\frac{\mathrm{d}}{\mathrm{d}b}(f * \breve{\theta}_a)(b)$ has modulus maxima at sharp variation points of $f * \breve{\theta}_a$

- If the wavelet has two vanishing moment: $\psi = -\theta''$ then $W_2(a, b) = a\frac{\mathrm{d}^2}{\mathrm{d}b^2}(f * \breve{\theta}_a)(b)$ corresponds to locally maximum curvatures

f(t)

$f * \bar{\theta}_s(u)$ u

$W_1 f(u,s)$ u

$W_2 f(u,s)$ u
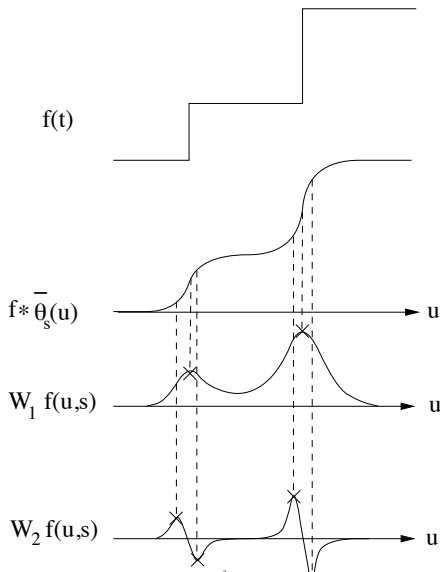
Wavelet Maxima Lines

- **Point of Modulus Maximum** are any point (b_0, a_0) in the time-scale plan such that the curve $b \mapsto |Wf(b, a_0)|$ is locally maximum at $b = b_0$. This implies that

$$\frac{\partial Wf(a_0, b_0)}{\partial b} = 0$$

- **Maxima lines** is any connected curve $a(b)$ in the scale-space plane (b, a) along which all points are modulus maxima.

Theorem (Hwang, Mallat)

Suppose that ψ is $\mathcal{C}^N$ with a compact support and $\psi = (-1)^N \theta^{(N)}$ with $\int \theta \neq 0$. Let $f \in L^1[b_0, b_1]$. If there exists $a_0 > 0$ such that $|Wf(a, b)|$ has no local maximum for $b \in [b_0, b_1]$ and $a < a_0$, then f is uniformly Lipschitz–N on $[b_0 + \epsilon, b_1 - \epsilon]$, for any $\epsilon > 0$.

Wavelet Maxima Lines

Remarks

- This theorem implies that f can be singular (not Lipschitz–1) at a point x_0 only if there is a sequence of wavelet maxima points $(b_k, a_k)_{k \in \mathbb{N}}$ that converges toward x_0 at fine scales:

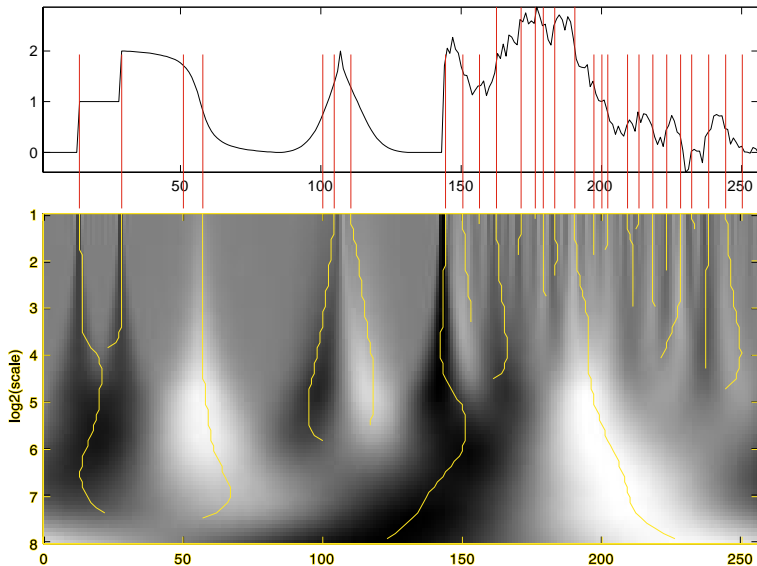$$\lim_{k \to +\infty} b_k = \boldsymbol{x}_0 \quad \text{and} \quad \lim_{k \to +\infty} a_k = 0$$

- These modulus maxima points may or may not be along the same maxima line. This result guarantees that all singularities are detected by following the wavelet transform modulus maxima at fine scales
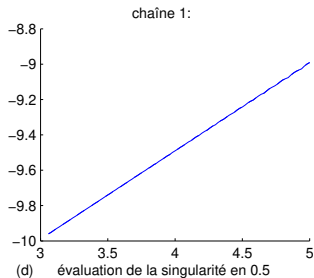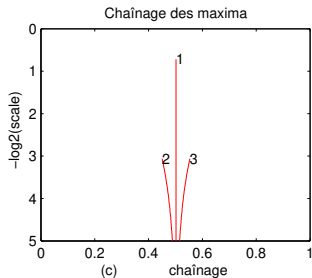
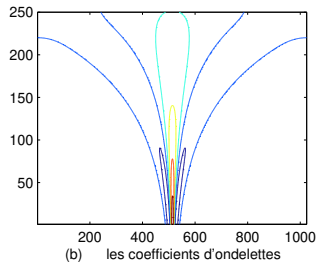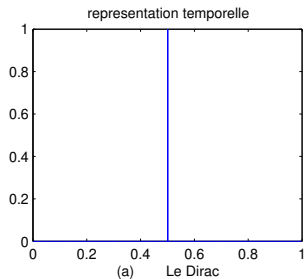## Theorem (Hummel, Poggio, Yuille)

Let $\psi = (-1)^N \theta^{(N)}$ where θ is Gaussian. For any $f \in L^2$, the modulus maxima of $Wf(a, b)$ belongs to connected curves that are never interrupted when the scale decreases

Wavelet Maxima Lines

Example

Example: a simple Dirac δ

representation temporelle

(a)   Le Dirac

(b)   les coefficients d'ondelettes

Chaînage des maxima

(c)   chaînage

−log2(scale)

chaîne 1:

(d)   évaluation de la singularité en 0.5

Example: 2 cusps $f(x) = |x - 0.25|^{\frac{1}{3}} + |x - 0.7|^{\frac{2}{3}}$

(a) La fonction

(b) les coefficients d'ondelettes

(c) chaînage

(d) évaluation de la singularité en 0.7 (trait continu) et 0.25 (trait pointillé)
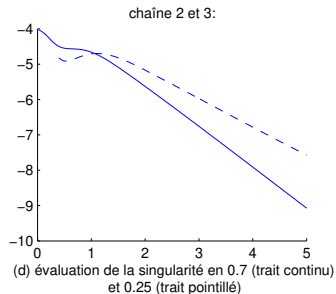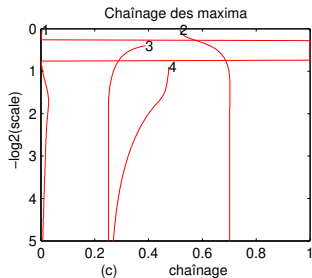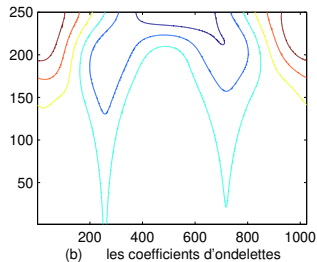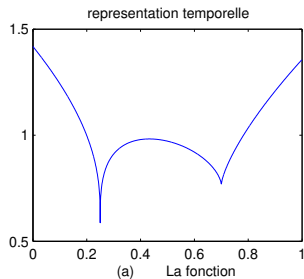
Example: $f(x) = |x - 0.25|^{\frac{1}{3}} + |x - 0.7|^{\frac{2}{3}} + \text{noise}$ (SNR=0.01)

(a) La fonction

(b) les coefficients d'ondelettes

(c) chaînage

Chaînage des maxima

chaîne 2 et 3:

(d) évaluation de la singularité en 0.7 (trait continu) et 0.25 (trait pointillé)
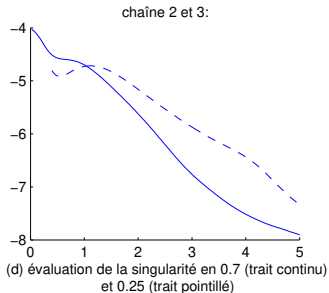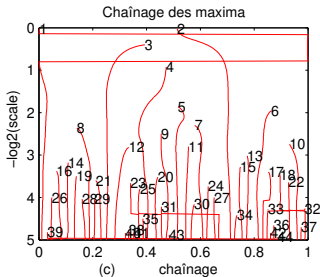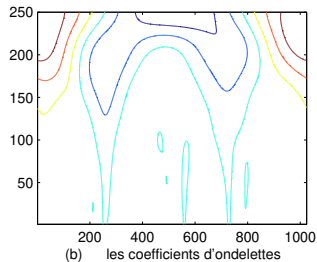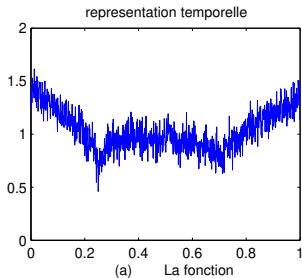
Practical estimation of α

f is uniformly Lipschitz–α in the neighborhood of x_0 iff there exists $A > 0$ such that each modulus maximum (b, a) in the cone satisfies

$$|Wf(a, b)| \leqslant A \, a^{\alpha + \frac{1}{2}}$$

which is equivalent to

$$\log_2 |Wf(a, b)| \leqslant \log_2 A + \left(\alpha + \frac{1}{2} \right) \log_2 a$$

$\Rightarrow$ The Lipschitz regularity at x_0 is the maximum slope of $\log_2 |Wf(a, b)|$ as a function of $\log_2 a$ along the maxima lines converging to x_0

Practical estimation of α

Example
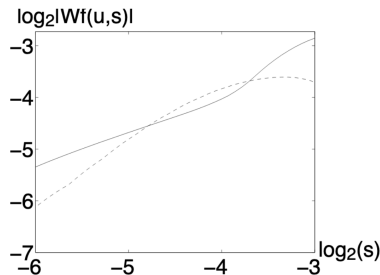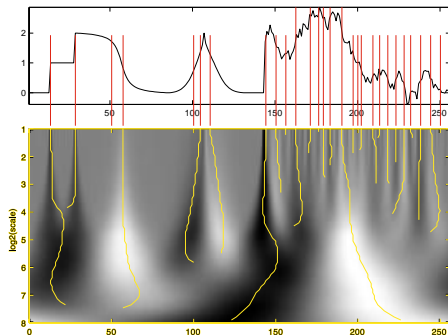
Figure: The full line gives the decay along the maxima line that converges to the first jump, and the dashed line to the first cusp.

Brownian motion
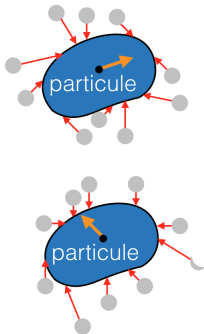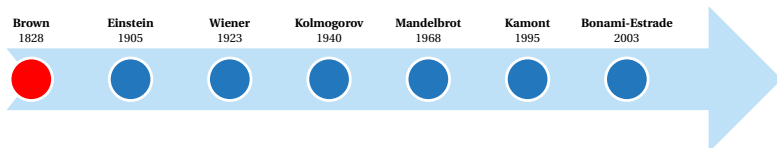
Properties

- Independants displacements
- Gaussian distribution
- Irregular trajectories

Brownian motion

| Brown 1828 | Einstein 1905 | Wiener 1923 | Kolmogorov 1940 | Mandelbrot 1968 | Kamont 1995 | Bonami-Estrade 2003 |

Independants displacements

Irregular trajectories

Gaussian distribution

Brownian motion

(a) (b)

$$\overline{(\Delta x)^2} \propto t$$

Brownian motion

$$X \; : \; T \times \Omega \; \longrightarrow \; E$$
$$(t, \omega) \; \longmapsto \; X(t, \omega)$$
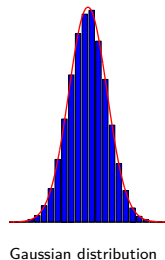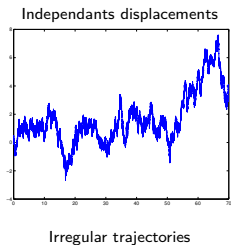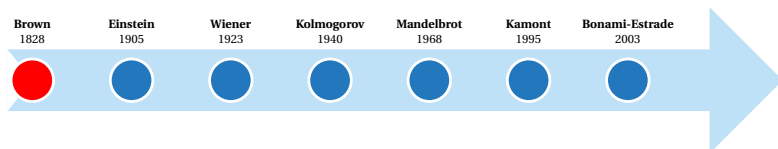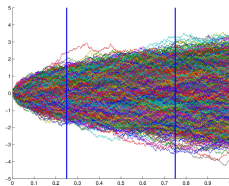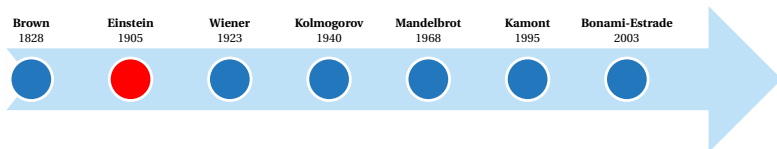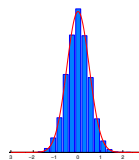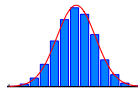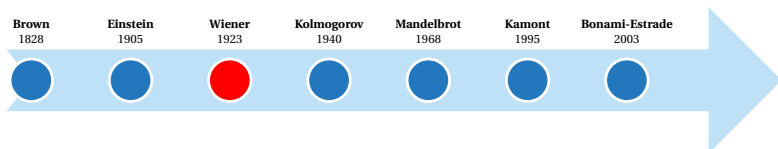
Brownian motion

- $(B_t)_t$ has independants increments, $B_0 = 0$ a.s.
- $B_{t_i} - B_{t_j} \sim \mathcal{N}(0, t_i - t_j)$
- $(B_t)_t$ has continuous sample paths a.s.

$$\overline{(\Delta x)^2} \propto t$$

Brownian motion

Brown 1828	Einstein 1905	Wiener 1923	Kolmogorov 1940	Mandelbrot 1968	Kamont 1995	Bonami-Estrade 2003
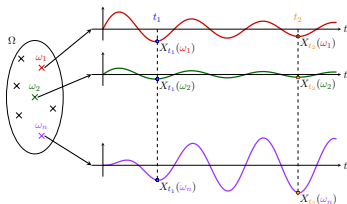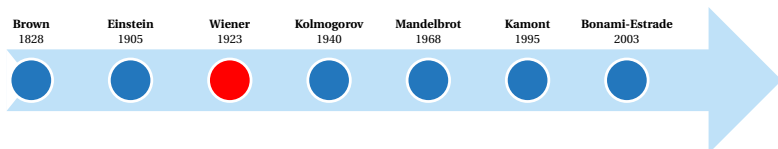
Isometry $\mathbf{W} : (L^2, \langle f, g \rangle_{L^2}) \to (\mathcal{G}, \mathbb{E}[XY])$

- $\mathbb{E}[\mathbf{W}(f)\mathbf{W}(g)] = \langle f, g \rangle_{L^2}, \quad \mathbf{W}(f) \sim \mathcal{N}(0, \|f\|_{L^2}^2)$

- $\forall t \in [0,1], \quad B_t \overset{\text{def}}{=} \mathbf{W}(\mathbb{1}_{[0,t]})$

- $\mathbb{E}[(B_t - B_s)^2] = \left\| \mathbb{1}_{[0,t]} - \mathbb{1}_{[0,s]} \right\|_{L^2}^2 = \int \mathbb{1}_{[s,t]} = t - s$

- $\mathbb{E}\left[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})\right] = \langle \mathbb{1}_{[t_{i-1},t_i]}, \mathbb{1}_{[t_{j-1},t_j]} \rangle_{L^2} = 0$

$$\text{Wiener stochastic integral} = \int f(x)\mathbf{W}(\mathrm{d}x)$$

Self-similarity
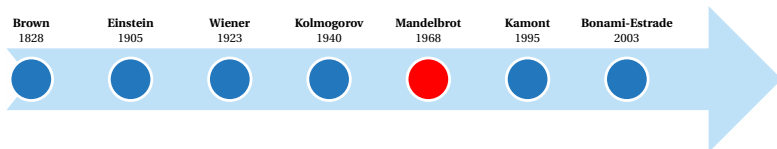
Self-similarity

$\{X(t)\}_{t \in T}$ self-similar of order H if

$$\forall \lambda \in \mathbb{R}, \ \{X(\lambda t)\}_{t \in T} \stackrel{(fdd)}{=} \lambda^H \{X(t)\}_{t \in T}$$
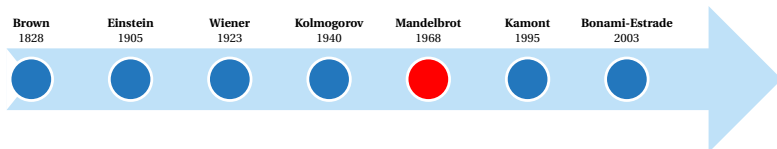
Self-similarity

Self-similarity

$\{X(t)\}_{t \in T}$ self-similar of order H if

$$\forall \lambda \in \mathbb{R}, \ \{X(\lambda t)\}_{t \in T} \overset{(fdd)}{=} \lambda^H \{X(t)\}_{t \in T}$$
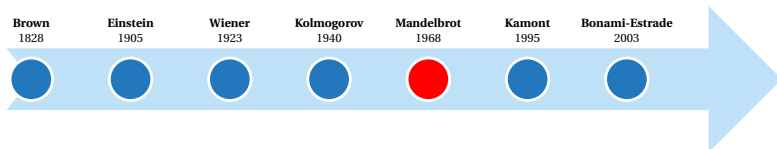
Self-similarity

Self-similarity

$\{X(t)\}_{t \in \mathcal{T}}$ self-similar of order H if

$$\forall \lambda \in \mathbb{R}, \ \{X(\lambda t)\}_{t \in \mathcal{T}} \overset{(fdd)}{=} \lambda^H \{X(t)\}_{t \in \mathcal{T}}$$
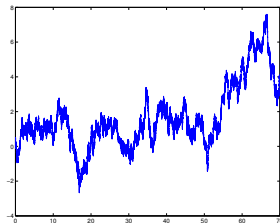
Fractional Brownian motion
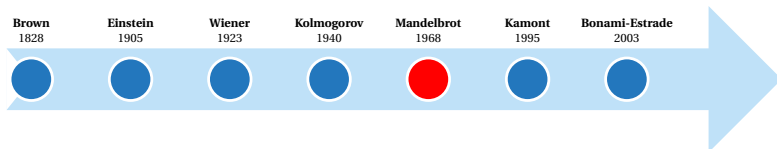
- $\mathbb{E}\left[(B^H(t) - B^H(s))^2\right] = |t - s|^{2H} \Rightarrow$ ~~independant increments~~
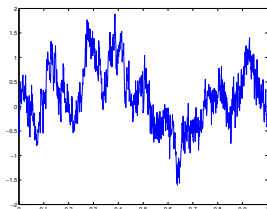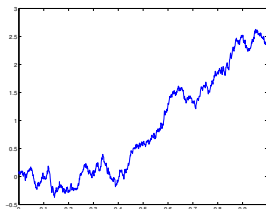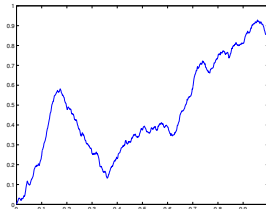
$H = 0.2$ $H = 0.5$ $H = 0.8$

Figure: Fractional Brownian motion B^H

Fractional Brownian motion
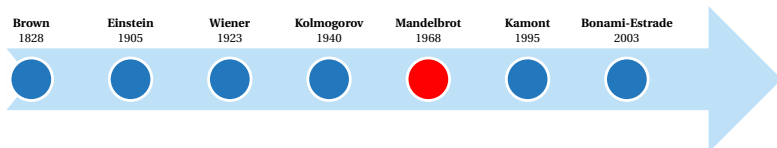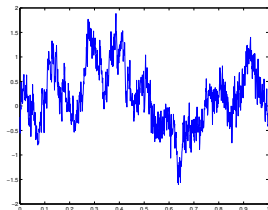
- $\mathbb{E}\left[(B^H(t) - B^H(s))^2\right] = |t - s|^{2H} \Rightarrow$ stationary increments

$H = 0.2$ $H = 0.5$ $H = 0.8$
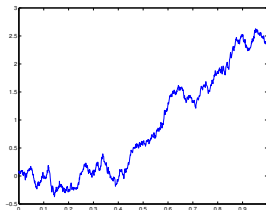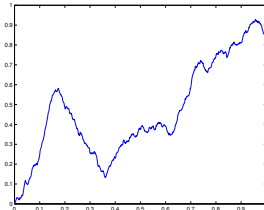
Fractional Brownian motion
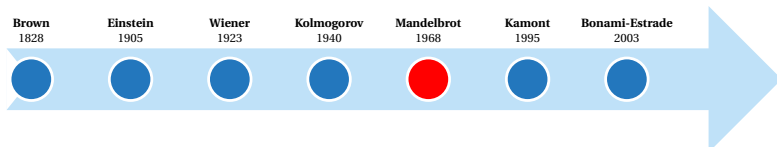
- $\mathbb{E}\left[(B^H(t) - B^H(s))^2\right] = |t - s|^{2H} \Rightarrow$ stationary increments
- $\mathbf{R}(t,s) = \mathrm{Cov}(B^H(t), B^H(s)) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H})$
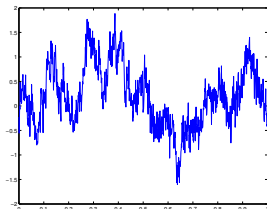
$H = 0.2 \qquad\qquad H = 0.5 \qquad\qquad H = 0.8$
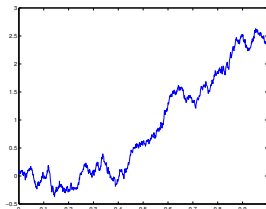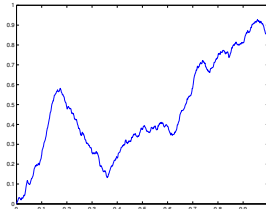
Fractional Brownian motion
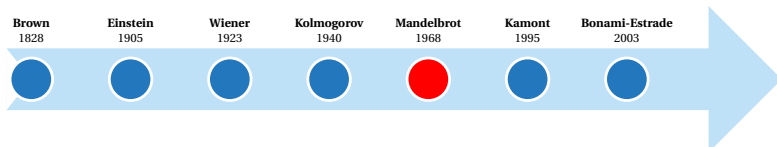
- $\mathbb{E}\left[(B^H(t) - B^H(s))^2\right] = |t - s|^{2H} \Rightarrow$ stationary increments
- $\mathbf{R}(t, s) = \mathrm{Cov}(B^H(t), B^H(s)) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H})$
- $B^H(t) = \frac{1}{c_H} \int_{\mathbb{R}} \frac{e^{jt\xi} - 1}{|\xi|^{H+1/2}} \widehat{\mathbf{W}}(\xi) \Rightarrow$ harmonizable formula

$H = 0.2$ $H = 0.5$ $H = 0.8$

Fractional Brownian field

Brown 1828 — Einstein 1905 — Wiener 1923 — Kolmogorov 1940 — Mandelbrot 1968 — Kamont 1995 — Bonami-Estrade 2003
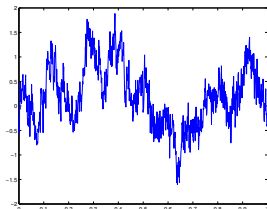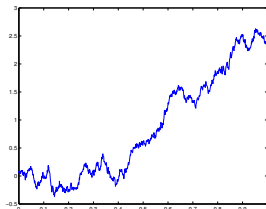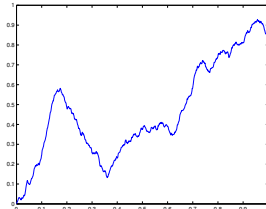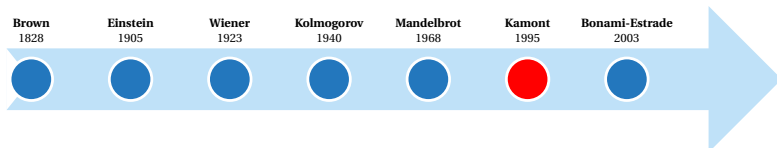
- $\mathbb{E}\left[(B^H(\boldsymbol{x}) - B^H(\boldsymbol{y}))^2\right] = \|\boldsymbol{x} - \boldsymbol{y}\|^{2H}, \; \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$
- $\mathbf{R}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\left(\|\boldsymbol{x}\|^{2H} + \|\boldsymbol{y}\|^{2H} - \|\boldsymbol{x} - \boldsymbol{y}\|^{2H}\right)$
- $B^H(\boldsymbol{x}) = \frac{1}{C_H}\int_{\mathbb{R}^2}\frac{e^{j\langle\boldsymbol{x},\boldsymbol{\xi}\rangle}-1}{\|\boldsymbol{\xi}\|^{H+1}}\widehat{\mathbf{W}}(d\boldsymbol{\xi})$

$H = 0.2$ — $H = 0.5$ — $H = 0.8$
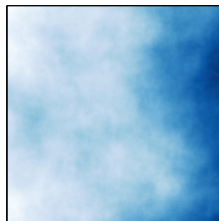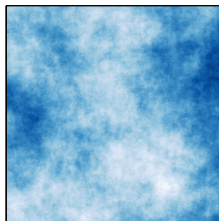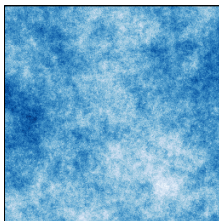
Wavelet-based estimation of the Hurst exponent

- Let us consider a discrete wavelet transform at scales $a = 2^{-j}$ and positions $b = k$

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)$$

which encodes series information in details

$$d_{j,k} = \langle B^H, \psi_{j,k}\rangle$$

- Compute wavelet variance

$$\mathrm{Var}(d_{j,\bullet}) = \frac{1}{n_j}\sum_{k=0}^{n_j-1}|d_{j,k}|^2$$

- Plot the $\log_2$ of variances versus scale j

$$\log_2(\mathrm{Var}(d_{j,\bullet})) = (2H+1)j + \mathsf{cste}$$

Wavelet Maxima Lines for Brownian motion
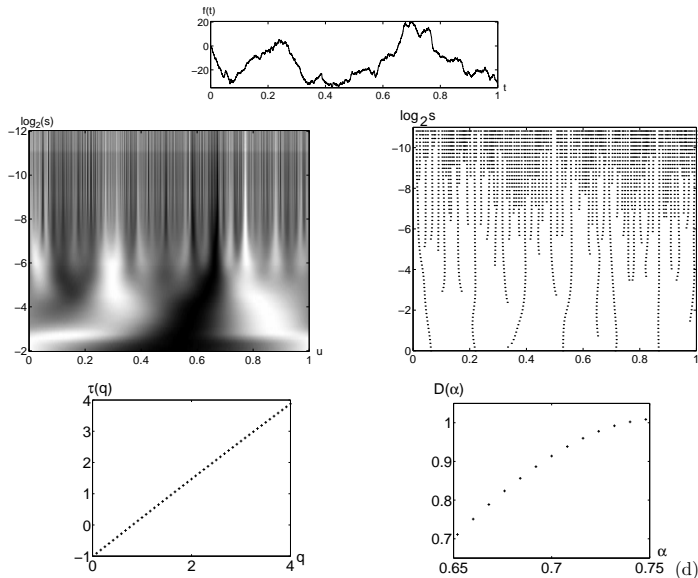
Credits: S. Mallat (Wavelet tour)

Take home message

- Vanishing moments up to order N make the wavelet ψ blind to polynomial of degree $\leqslant N$ (smooth part of the signal), leading to better detections of singularities

- If the function is Lipschitz–α, then the amplitude of the wavelet coefficients are going to decay very fast to zero when the scale goes to zero (all the more that α is high)

- A remarkable aspect is the reverse: if we know this property, then we can characterize the pointwise regularity of the function at any point

- All singularities are detected by following the wavelet transform modulus maxima at fine scale

- The Lipschitz regularity at every point can be retrieved by measuring the maximum slop of the decay of $\log_2 |Wf(a, b)|$

- The wavelet-based estimation of the Lipschitz regularity enables to recover the self-similarity exponent of fractals

The 2D Continuous Wavelet Transform

Bidimensional Continuous Wavelet Transform

1. 2D Wavelets
2. Directional Continuous Wavelet Transform, inversion formula
3. Isotropic Wavelet Transform
4. A wavelet for image analysis: the "Canny" wavelets

2D Fourier Transform

The bidimensional Fourier transform of a function f integrable on $\mathbb{R}^2$ is defined by:

$$\hat{f}(\boldsymbol{k}) = \iint\limits_{\mathbb{R}^2} f(\boldsymbol{x})\, e^{-2i\pi \boldsymbol{k}\cdot\boldsymbol{x}}\, d\boldsymbol{x}, \quad \forall \boldsymbol{k} \in \mathbb{R}^2$$

If $f \in L^2(\mathbb{R}^2)$, the inversion formula is given by:

$$f(\boldsymbol{x}) = \iint\limits_{\mathbb{R}^2} \hat{f}(\boldsymbol{k})\, e^{2i\pi \boldsymbol{k}\cdot\boldsymbol{x}}\, d\boldsymbol{k}$$

and the energy conservation writes:

$$\iint\limits_{\mathbb{R}^2} |f(\boldsymbol{x})|^2 d\boldsymbol{x} = \iint\limits_{\mathbb{R}^2} |\hat{f}(\boldsymbol{k})|^2 d\boldsymbol{k}$$

2D Fourier Transform
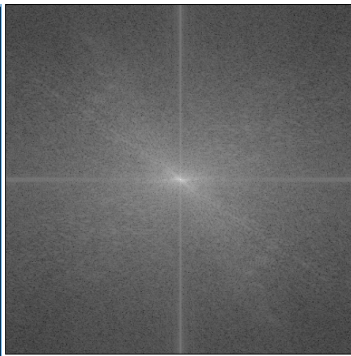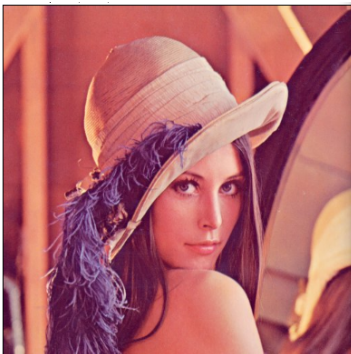
Definition of 2D wavelets

$\psi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ is a wavelet if it satisfies the admissibility condition:

$$c_\psi = \iint\limits_{\mathbb{R}^2} \frac{|\hat{\psi}(\boldsymbol{k})|^2}{\|\boldsymbol{k}\|^2} \, \mathrm{d}\boldsymbol{k} < +\infty$$

which implies (and is equivalent provided ψ has sufficient decay at infinity):

$$\iint\limits_{\mathbb{R}^2} \psi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = 0$$

In practice, one usually needs that ψ has *p* vanishing moments:

$$\iint\limits_{\mathbb{R}^2} x_1^{\alpha_1} x_2^{\alpha_2} \psi(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 = 0, \quad \forall \alpha_1, \alpha_2 \in \mathbb{N} \text{ s.t. } \alpha_1 + \alpha_2 \leq p - 1$$

Remark: this means that the Fourier transform of the wavelet should behave as $\|\boldsymbol{k}\|^p$ when $\boldsymbol{k} \to 0$ in Fourier domain.

2D Wavelet family

Let $\psi(\boldsymbol{x})$ be an admissible wavelet. The wavelet family is defined by dilation, rotation, and translation from ψ:

$$\psi_{(a,\boldsymbol{b},\theta)}(\boldsymbol{x}) = \frac{1}{a}\psi\left(\mathbf{R}_{-\theta}\left(\frac{\boldsymbol{x}-\boldsymbol{b}}{a}\right)\right)$$

with $\boldsymbol{b} \in \mathbb{R}^2$ the translation parameter, a the positive scale and $\mathbf{R}^{-\theta}$ the rotation of angle θ in $\mathbb{R}^2$, corresponding to matrix

$$\mathbf{R}_{-\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Example (Anisotropic Morlet Wavelet)

Let $\boldsymbol{u} = (\cos\alpha, \sin\alpha)$ the unitary vector of direction α.
The (complex) Morlet wavelet is:

$$\psi(\boldsymbol{x}) = \mathrm{e}^{-\pi\|\boldsymbol{x}\|^2}\mathrm{e}^{10i\pi\boldsymbol{x}\cdot\boldsymbol{u}}$$

Isotropic wavelets

Example (Iterated Laplacian of Gaussian)

For $n \geq 1$, the wavelet h_{2n} is defined by:

$$h_{2n}(\boldsymbol{x}) = (-1)^n \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right)^n e^{-\pi \|\boldsymbol{x}\|^2}$$

Its Fourier transform is given by:

$$\widehat{h}_{2n}(\boldsymbol{k}) = 4^n \pi^{2n} \|\boldsymbol{k}\|^{2n} e^{-\pi \|\boldsymbol{k}\|^2}$$

For $n = 2$, h_2 is the Laplacian of Gaussian, popular in computer vision, also called the *Mexican hat*.

Remark: the wavelet h_{2n} has exactly $2n$ vanishing moments. The maximum of its Fourier transform $\hat{h}_{2n}$ is achieved for $k_0 = \sqrt{2n}$.

Isotropic wavelets

Example (Iterated Laplacian of Gaussian)

- Wavelet $h_2(x, y)$ (Mexican hat) and 1D section:
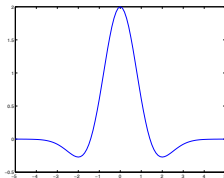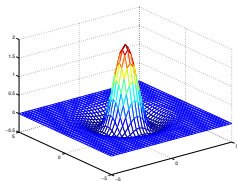
- 1D section in physical and Fourier space of the wavelet h_8
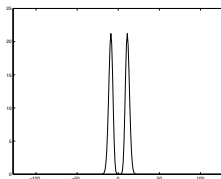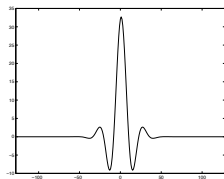
2D directional continuous wavelet transform

Let ψ be a 2D wavelet.

The directional wavelet transform of a given function $f \in L^2(\mathbb{R}^2)$ is defined by:

$$
\begin{aligned}
Wf(a, \boldsymbol{b}, \theta) &= \iint_{\mathbb{R}^2} f(\boldsymbol{x}) \, \overline{\psi_{(a,\boldsymbol{b},\theta)}(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x} \\
&= \frac{1}{a} \iint_{\mathbb{R}^2} f(\boldsymbol{x}) \, \psi\left(\mathbf{R}_{-\theta}\left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a}\right)\right) \mathrm{d}\boldsymbol{x}
\end{aligned}
$$

Applying Parseval formula, it writes:

$$
Wf(a, \boldsymbol{b}, \theta) = a \iint_{\mathbb{R}^2} \widehat{f}(\boldsymbol{k}) \, \overline{\widehat{\psi}(a\,\mathbf{R}_{-\theta}\boldsymbol{k})} \mathrm{e}^{2i\pi\boldsymbol{k}\cdot\boldsymbol{b}} \, \mathrm{d}\boldsymbol{k}
$$

Inversion formula

The function f can be reconstructed by:

$$f(\boldsymbol{x}) = \frac{1}{c_\psi} \int_0^{+\infty} \int_0^{2\pi} \iint_{\mathbb{R}^2} Wf(a, \boldsymbol{b}, \theta) \, \psi_{(a,\boldsymbol{b},\theta)}(\boldsymbol{x}) \, \frac{\mathrm{d}a}{a^3} \, \mathrm{d}\theta \, \mathrm{d}\boldsymbol{b}$$

with

$$c_\psi = \iint_{\mathbb{R}^2} \frac{|\widehat{\psi}(\boldsymbol{k})|^2}{\|\boldsymbol{k}\|^2} \, \mathrm{d}\boldsymbol{k}$$

The energy conservation writes:

$$\iint_{\mathbb{R}^2} |f(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = \frac{1}{c_\psi} \int_0^{+\infty} \int_0^{2\pi} \iint_{\mathbb{R}^2} |Wf(a, \boldsymbol{b}, \theta)|^2 \, \frac{\mathrm{d}a}{a^3} \, \mathrm{d}\theta \, \mathrm{d}\boldsymbol{b}$$

Inversion formula with a different wavelet

Let $f(\boldsymbol{x}) \in L^2(\mathbb{R}^2)$.

Wavelet decomposition of $f(\boldsymbol{x})$ with an *analysing wavelet* g:

$$W_g f(a, \boldsymbol{b}, \theta) = \iint\limits_{\mathbb{R}^2} f(\boldsymbol{x}) \frac{1}{a} \bar{g}\left(\mathbf{R}_{-\theta}\left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a}\right)\right) \mathrm{d}\boldsymbol{x}$$

Synthesis with a *reconstruction wavelet* h:

$$f(\boldsymbol{x}) = \frac{1}{c_{gh}} \int_0^{+\infty} \int_0^{2\pi} \iint\limits_{\mathbb{R}^2} W_g f(a, \boldsymbol{b}, \theta) \frac{1}{a} h\left(\mathbf{R}_{-\theta}\left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a}\right)\right) \frac{\mathrm{d}a}{a^3} \mathrm{d}\theta \, \mathrm{d}\boldsymbol{b}$$

Cross admissibility condition on functions $g, h \in L^2(\mathbb{R}^2)$:

$$c_{gh} = \iint\limits_{\mathbb{R}^2} \frac{\bar{\widehat{g}}(\boldsymbol{k})\widehat{h}(\boldsymbol{k})}{\|\boldsymbol{k}\|^2} \mathrm{d}\boldsymbol{k} < +\infty$$

Classical examples

Wavelet constructed from the Gaussian $G(\boldsymbol{x}) = \mathrm{e}^{-\pi \|\boldsymbol{x}\|^2}$

- Wavelet transform with an **isotropic wavelet**
  - $g(\boldsymbol{x}) = h(\boldsymbol{x}) = \Delta G(\boldsymbol{x})$ *(Mexican hat)*
  - $g(\boldsymbol{x}) = G(\boldsymbol{x})$ (*g* is not a *wavelet*) and $h(\boldsymbol{x}) = \Delta G(\boldsymbol{x})$
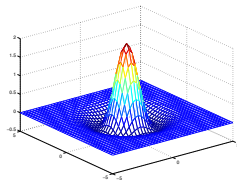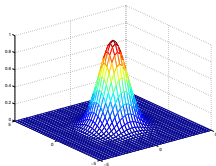
- Wavelet transform with a **vector wavelet** $g(\boldsymbol{x}) = \nabla G(\boldsymbol{x})$
  (Canny multi-scale detector)

Isotropic Wavelet Transform

When the wavelet is real, isotropic (i.e rotation invariant $\psi(\boldsymbol{x}) = h(\|\boldsymbol{x}\|)$), the wavelet transform of f comes down:

$$Wf(a, \boldsymbol{b}) = \frac{1}{a} \iint\limits_{\mathbb{R}^2} f(\boldsymbol{x}) \, \psi\left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a}\right) \mathrm{d}\boldsymbol{x}$$

$\Rightarrow$ the integral on θ disappears

From Parseval equality, it writes:

$$Wf(a, \boldsymbol{b}) = a \iint\limits_{\mathbb{R}^2} \widehat{f}(\boldsymbol{k}) \, \overline{\widehat{\psi}(a\boldsymbol{k})} \mathrm{e}^{2i\pi \boldsymbol{k} \cdot \boldsymbol{b}} \, \mathrm{d}\boldsymbol{k}$$

$\Rightarrow$ the wavelet transform acts as a filter on the Fourier transform of f around the frequency $\frac{\boldsymbol{k}_0}{a}$.

Isotropic Wavelet Transform
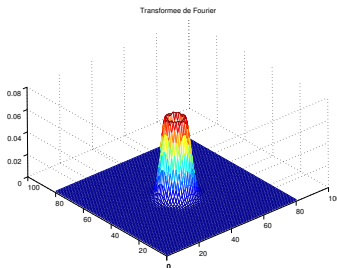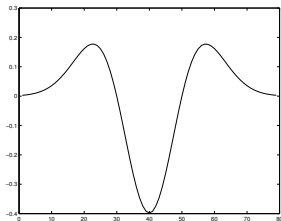
If ψ is admissible, one has the energy conservation:

$$\iint\limits_{\mathbb{R}^2} |f(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = \frac{1}{c_\psi} \int_0^{+\infty} \iint\limits_{\mathbb{R}^2} |Wf(a, \boldsymbol{b})|^2 \frac{\mathrm{d}a}{a^3} \, \mathrm{d}\boldsymbol{b}$$

and the synthesis formula:

$$f(\boldsymbol{x}) = \frac{1}{c_\psi} \int_0^{+\infty} \iint\limits_{\mathbb{R}^2} Wf(a, \boldsymbol{b}) \psi_{a, \boldsymbol{b}}(\boldsymbol{x}) \frac{\mathrm{d}a}{a^3} \, \mathrm{d}\boldsymbol{b}$$

The Canny multiscale detector for image processing

Let Θ be a smoothing kernel such that:

- $\iint_{\mathbb{R}^2} \Theta = 1$
- $\Theta \geq 0$
- Θ isotropic or $\Theta(x,y) = \Theta_1(x)\Theta_2(y)$

Example (Gaussian)

$\Theta(\boldsymbol{x}) = G(\boldsymbol{x}) = \mathrm{e}^{-\pi\|\boldsymbol{x}\|^2}$ a smoothing kernel isotropic and tensorial

Directional Wavelets

$$\boldsymbol{\Psi}(x) = \nabla\Theta(x) = (\psi^1, \psi^2)$$

$$\psi^1 = -\frac{\partial\Theta}{\partial x_1} \ \text{ and } \ \psi^2 = -\frac{\partial\Theta}{\partial x_2}$$

Wavelets in the direction $\varphi \Rightarrow \psi^\varphi = \cos\varphi\frac{\partial G}{\partial x_1} + \sin\varphi\frac{\partial G}{\partial x_2} = \vec{\varphi}\cdot\nabla G$

The Canny multiscale detector for image processing

Decomposition: computation of the vector wavelet transform

$$\boldsymbol{W}f(a, \boldsymbol{b}) = (W^1 f(a, \boldsymbol{b}), W^2 f(a, \boldsymbol{b}))$$

- $W^1 f(a, \boldsymbol{b}) = \iint\limits_{\mathbb{R}^2} f(\boldsymbol{x}) \frac{1}{a} \psi^1 \left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a} \right) \mathrm{d}\boldsymbol{x} \; \rightarrow$ vertical singularities

- $W^2 f(a, \boldsymbol{b}) = \iint\limits_{\mathbb{R}^2} f(\boldsymbol{x}) \frac{1}{a} \psi^2 \left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a} \right) \mathrm{d}\boldsymbol{x} \; \rightarrow$ horizontal singularities

Interpretation:

$$\boldsymbol{W}f(a, \boldsymbol{b}) = a \nabla \left(f * \frac{1}{a} \breve{\Theta} \left(\frac{\boldsymbol{x}}{a} \right) \right) (\boldsymbol{b})$$

$\boldsymbol{W}f$ represents the gradient of the image, smoothed by Θ at scale a

Proof: Let define $L(\boldsymbol{x}) = -\frac{\boldsymbol{x}}{a}$, $\check{\Theta}_a = \frac{1}{a}\Theta \circ L$, $\check{\psi}_a^k(\boldsymbol{x}) = \frac{1}{a}\psi^k\left(-\frac{\boldsymbol{x}}{a}\right)$.

$$W^k f(a, \boldsymbol{b}) = \iint\limits_{\mathbb{R}^2} f(\boldsymbol{x})\frac{1}{a}\psi^k\left(\frac{\boldsymbol{x}-\boldsymbol{b}}{a}\right)\mathrm{d}\boldsymbol{x} = (f * \check{\psi}_a^k)(\boldsymbol{b})$$

By the chain rule $\frac{\partial}{\partial x_k}(\Theta \circ L)(\boldsymbol{x}) = \frac{\partial \Theta}{\partial x_k}(L(\boldsymbol{x}))\frac{\partial L}{\partial x_k}(\boldsymbol{x})$ hence

$$\check{\psi}_a^k(\boldsymbol{x}) = -\frac{1}{a}\frac{\partial \Theta}{\partial x_k}\left(-\frac{\boldsymbol{x}}{a}\right) = a\frac{\partial}{\partial x_k}\left(\frac{1}{a}\Theta \circ L\right)(\boldsymbol{x}) = a\frac{\partial \check{\Theta}_a}{\partial x_k}$$

$$
\begin{aligned}
(f * \check{\psi}_a^k)(\boldsymbol{b}) &= a\iint f(\boldsymbol{x})\frac{\partial \check{\Theta}_a}{\partial x_k}(\boldsymbol{b}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \frac{\partial}{\partial b_k}\iint f(\boldsymbol{x})\check{\Theta}_a(\boldsymbol{b}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&= \frac{\partial}{\partial b_k}(f * \check{\Theta}_a)(\boldsymbol{b})
\end{aligned}
$$

$$\begin{pmatrix} W^1 f(a, \boldsymbol{b}) \\ W^2 f(a, \boldsymbol{b}) \end{pmatrix} = a\begin{pmatrix} \frac{\partial}{\partial b_1}(f * \check{\Theta}_a)(\boldsymbol{b}) \\ \frac{\partial}{\partial b_2}(f * \check{\Theta}_a)(\boldsymbol{b}) \end{pmatrix} = a\nabla(f * \check{\Theta}_a)(\boldsymbol{b}) \quad \square$$

Multiscale detector and directional wavelet transform

If Θ is isotropic, one has $\frac{\partial \Theta}{\partial x_1}(\mathbf{R}^{-\theta}\boldsymbol{x}) = \cos\theta \frac{\partial \Theta}{\partial x_1}(\boldsymbol{x}) + \sin\theta \frac{\partial \Theta}{\partial x_2}(\boldsymbol{x})$

Then,

$$W_{\psi^1}f(a, \boldsymbol{b}, \varphi) = \vec{\varphi} \cdot \boldsymbol{W}f(a, \boldsymbol{b}) \to \text{singularities in the direction } \vec{\varphi}^{\perp}$$

where $W_{\psi^1}f$ is the directional wavelet transform of f with ψ^1 as analyzing wavelet.

One can write:

$$\begin{pmatrix} W_{\psi^1}f(a, \boldsymbol{b}, \varphi) \\ W_{\psi^2}f(a, \boldsymbol{b}, \varphi) \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} W^1 f(a, \boldsymbol{b}) \\ W^2 f(a, \boldsymbol{b}) \end{pmatrix}$$

In vector formulation:

$$\boldsymbol{W}_{\nabla\Theta} = \mathbf{R}_{-\varphi}\boldsymbol{W}$$

⤳ will provide a reconstruction formula for the multiscale detector!

Inversion of the multiscale detector (Le Cadet, PhD 2004)

$$f(\boldsymbol{x}) = \frac{\pi}{C_{\psi^1}} \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} \boldsymbol{W}f(\boldsymbol{b}, a) \cdot \boldsymbol{\Psi}_{a,\boldsymbol{b}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{b}$$

with $C_{\psi^1} = \pi^2$ for $\Theta = G$.

Proof: The reconstruction formula of the directional wavelet transform with wavelet ψ^1 gives:

$$f(\boldsymbol{x}) = \frac{1}{C_{\psi^1}} \int_0^{2\pi} \int_0^{+\infty} \iint_{\mathbb{R}^2} W_{\psi^1}f(a, \boldsymbol{b}, \theta) \frac{1}{a} \psi^1 \left(\mathbf{R}^{-\theta} \left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a} \right) \right) \mathrm{d}\theta \frac{\mathrm{d}a}{a^3} \, \mathrm{d}\boldsymbol{b}$$

Replacing $W_{\psi^1}f(a, \boldsymbol{b}, \theta)$ by its definition:

$$f(\boldsymbol{x}) = \frac{1}{C_{\psi^1}} \int_0^{2\pi} \int_0^{+\infty} \iint_{\mathbb{R}^2} \Big[\cos\theta \, W^1 f(a, \boldsymbol{b})$$
$$+ \sin\theta \, W^2 f(a, \boldsymbol{b}) \Big] \frac{1}{a} \frac{\partial \Theta}{\partial x_1} \left(\mathbf{R}^{-\theta} \left(\frac{\boldsymbol{x} - \boldsymbol{b}}{a} \right) \right) \mathrm{d}\theta \frac{\mathrm{d}a}{a^3} \, \mathrm{d}\boldsymbol{b}$$

If Θ is isotropic, one has:

$$\frac{\partial \Theta}{\partial x_1}(\mathbf{R}^{-\theta}\boldsymbol{x}) = \cos\theta \, \frac{\partial \Theta}{\partial x_1}(\boldsymbol{x}) + \sin\theta \, \frac{\partial \Theta}{\partial x_2}(\boldsymbol{x})$$

then

$$
\begin{aligned}
f(\boldsymbol{x})C_{\psi^1} \;=\; & \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} \mathrm{d}\boldsymbol{b} \int_0^{2\pi} \mathrm{d}\theta \cos^2\theta \; W^1 f(a, \boldsymbol{b}) \, \frac{1}{a}\frac{\partial \Theta}{\partial x_1}\Big(\frac{\boldsymbol{x}-\boldsymbol{b}}{a}\Big) \\[2mm]
+ \; & \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} \mathrm{d}\boldsymbol{b} \int_0^{2\pi} \mathrm{d}\theta \sin^2\theta \; W^2 f(a, \boldsymbol{b}) \, \frac{1}{a}\frac{\partial \Theta}{\partial x_2}\Big(\frac{\boldsymbol{x}-\boldsymbol{b}}{a}\Big) \\[2mm]
+ \; & \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} \mathrm{d}\boldsymbol{b} \int_0^{2\pi} \mathrm{d}\theta \cos\theta\sin\theta \; W^1 f(a, \boldsymbol{b}) \, \frac{1}{a}\frac{\partial \Theta}{\partial x_2}\Big(\frac{\boldsymbol{x}-\boldsymbol{b}}{a}\Big) \\[2mm]
+ \; & \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} \mathrm{d}\boldsymbol{b} \int_0^{2\pi} \mathrm{d}\theta \cos\theta\sin\theta \; W^2 f(a, \boldsymbol{b}) \, \frac{1}{a}\frac{\partial \Theta}{\partial x_1}\Big(\frac{\boldsymbol{x}-\boldsymbol{b}}{a}\Big)
\end{aligned}
$$

Since $\int_0^{2\pi} \cos^2\theta \, \mathrm{d}\theta = \int_0^{2\pi} \sin^2\theta \, \mathrm{d}\theta = \pi$ and $\int_0^{2\pi} \cos\theta\sin\theta \, \mathrm{d}\theta = 0$ then

$$f(\boldsymbol{x}) = \frac{\pi}{C_{\psi^1}} \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} \mathrm{d}\boldsymbol{b} \left[W^1 f(a, \boldsymbol{b})\psi^1_{a,\boldsymbol{b}}(\boldsymbol{x}) + W^2 f(a, \boldsymbol{b})\psi^2_{a,\boldsymbol{b}}(\boldsymbol{x}) \right] \;\square$$

Energy conservation formula

The vector $\boldsymbol{W}f(a,\boldsymbol{b})$ should be represented in modulus-orientation:

$$
\begin{aligned}
Mf(a,\boldsymbol{b}) &= \|\boldsymbol{W}f(a,\boldsymbol{b})\| \quad \text{Modulus} \\
Af(a,\boldsymbol{b}) &= \operatorname{Arg}\left(\boldsymbol{W}f(a,\boldsymbol{b})\right) \quad \text{Orientation}
\end{aligned}
$$

The energy conservation (with an isotropic kernel Θ) writes:

$$
\iint_{\mathbb{R}^2} |f(\boldsymbol{x})|^2 \mathrm{d}\boldsymbol{x} = \frac{\pi}{C_\psi} \int_{a>0} \frac{\mathrm{d}a}{a^3} \iint_{\mathbb{R}^2} (Mf(a,\boldsymbol{b}))^2 \, \mathrm{d}\boldsymbol{b}
$$

Example (Application: edge detection in 2D images)

1. Edge points at scale a are points where $\boldsymbol{b} \mapsto Mf(a,\boldsymbol{b})$ is locally maximum in the direction $Af(a,\boldsymbol{b})$.

2. Estimation of the maxima lines linking edge points through scales a. The *tops* of these maxima lines ($a \to 0$) finally constitute the edge points of the image.

3. Computation of the Lipschitz regularity at any edge point.

Application: edge detection in 2D images

Detection and classification of edges of a regular image, regular outside regular singularity lines.
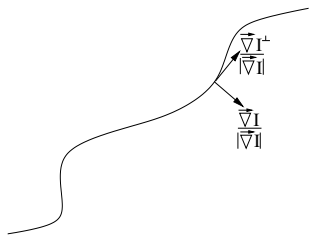
The edge is characterized by a singularity in the intensity, in the direction of the gradient $\vec{\nabla}I$. Along the edge, i.e. in the orthogonal direction of the gradient, the regularity is maximal

Remark: In pratice one will consider $\nabla(I * \frac{1}{a^2} G(\frac{x}{a}))$, which correspond to wavelet coefficients of I with a wavelet *Gaussian gradient*.

Application: edge detection in 2D images

Edge Model (Canny 86)

A point (x_0, y_0) of an image is an edge point if at this point the gradient modulus of the intensity, smoothed by a kernel θ_a, $|\nabla(I * \theta_a)|$, is locally maximum in the direction of the gradient $\nabla(I * \theta_a)$.
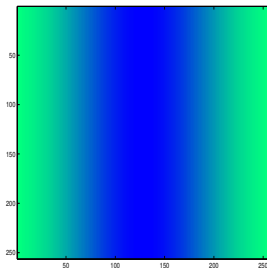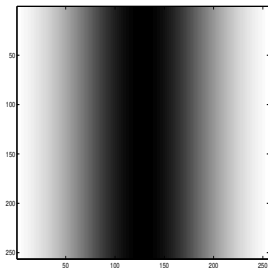
Variation of the intensity of a Gaussian distribution; where are the edges?

Application: edge detection in 2D images

New Edge Model (Mallat-Zhong, Mallat-Hwang 92, Le Cadet 2004)

f image smoothed by a kernel θ_a of scale a with $0 < a < a_{max}$: $g_a = f * \theta_a$. If there exists a connected curve through scales, along which all points are local maxima in the gradient direction ∇g_a, the limit (x_0, y_0) of this curve at small scales is an edge point.

a=1

Point de contour non significatif

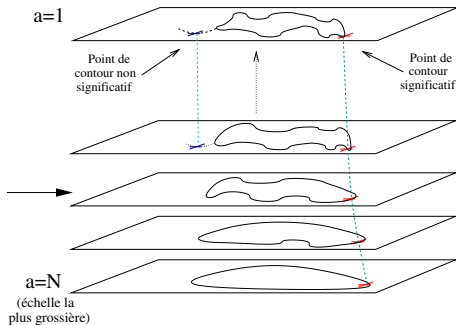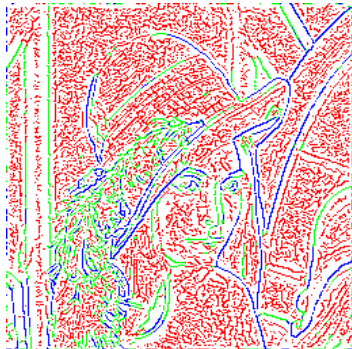Point de contour significatif

a=N
(échelle la plus grossière)

Figure: 2D maxima lines

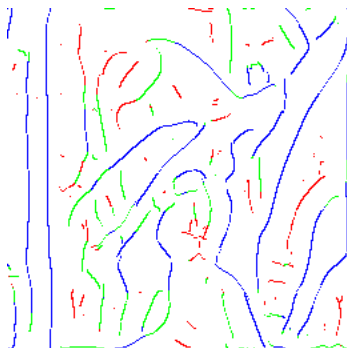# Application: modulus of the wavelet transform local max.

Fine scale

Application: modulus of the wavelet transform local max.

Intermediate scale
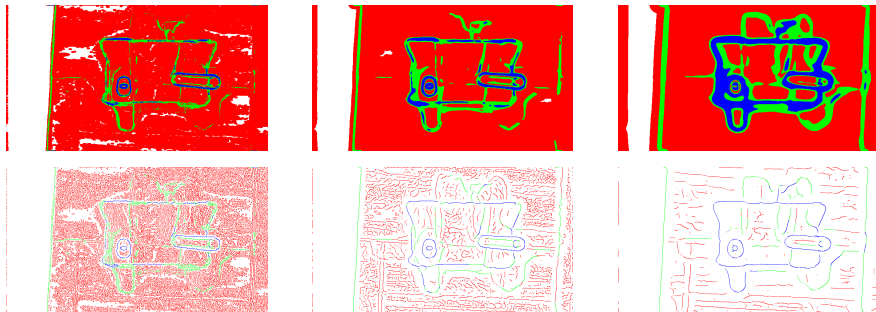
Application: modulus of the wavelet transform local max.

Large scale

Application: modulus of the wavelet transform local max.
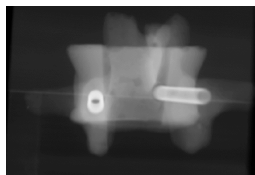
Figure: Edge points (top), wavelet coefficients maps at fixed scale (bottom) of a X-Ray image

Practice of the maxima line construction in 2D

1. Map of modulus maxima (in the gradient direction) at each scale.
2. Two modulus maxima between two successive scales are linked if they are neighbors in the gradient direction.
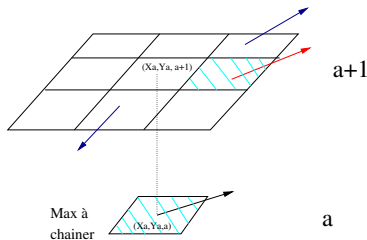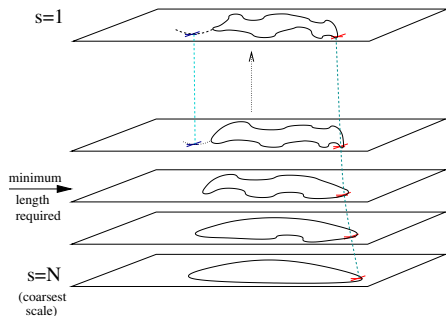
Practice of the maxima line construction in 2D

1. Let $Mf(x_0, y_0, a_{dep})$ be a modulus maximum at scale a_{dep}.
2. One consider, the 9 modulus $Mf(x_0(\pm 1), y_0(\pm 1), a_{dep+1})$.
3. One links with the maximum modulus that has the angle $Af(x_1, y_1, a_{dep+1})$ closest to $Af(x_0, y_0, a_{dep})$.
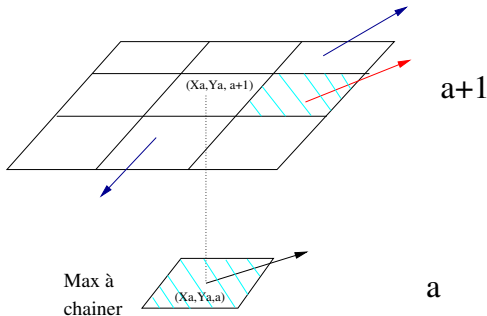
The dyadic wavelet transform

The scale varies along the dyadic sequence $\{2^j\}_{j\in\mathbb{Z}}$. Let $1 \leq k \leq 2$

$$\psi^k(\boldsymbol{x}) = -\frac{\partial\theta}{\partial x_k}, \quad \psi^k_{2^j}(\boldsymbol{x}) = \frac{1}{2^j}\psi^k\left(\frac{\boldsymbol{x}}{2^j}\right), \quad \check{\psi}^k_{2^j}(\boldsymbol{x}) = \psi^k_{2^j}(-\boldsymbol{x})$$

The dyadic wavelet transform at $\boldsymbol{b} = (b_1, b_2)$ is:

$$W^k(2^j, \boldsymbol{b}) = \langle f, \psi^k_{2^j}(\cdot - \boldsymbol{b})\rangle = f * \check{\psi}^k_{2^j}(\boldsymbol{b})$$

Let $\theta_{2^j}(\boldsymbol{x}) = 2^{-j}\theta(2^{-j}\boldsymbol{x})$ and $\check{\theta}_{2^j}(\boldsymbol{x}) = \theta_{2^j}(-\boldsymbol{x})$. The wavelet transform components are proportional to the gradient of f smoothed by $\check{\theta}_{2^j}$:

$$\begin{pmatrix} W^1 f(2^j, \boldsymbol{b}) \\ W^2 f(2^j, \boldsymbol{b}) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial b_1}(f * \check{\theta}_{2^j})(\boldsymbol{b}) \\ \frac{\partial}{\partial b_2}(f * \check{\theta}_{2^j})(\boldsymbol{b}) \end{pmatrix} = 2^j \nabla(f * \check{\theta}_{2^j})(\boldsymbol{b})$$

The dyadic wavelet transform

- The modulus of this gradient vector is proportional to the wavelet transform modulus:
$$Mf(2^j, \boldsymbol{b}) = \sqrt{|W^1 f(2^j, \boldsymbol{b})|^2 + |W^2 f(2^j, \boldsymbol{b})|^2}$$

- The angle $Af(2^j, \boldsymbol{b})$ of the wavelet transform vector:
$$Af(2^j, \boldsymbol{b}) = \begin{cases} \alpha(\boldsymbol{b}) & \text{if } W^1 f(2^j, \boldsymbol{b}) \geqslant 0 \\ \pi + \alpha(\boldsymbol{b}) & \text{if } W^2 f(2^j, \boldsymbol{b}) \geqslant 0 \end{cases} ,$$

$$\alpha(\boldsymbol{b}) = \tan^{-1}\left[\frac{W^2 f(2^j, \boldsymbol{b})}{W^1 f(2^j, \boldsymbol{b})}\right], \, \boldsymbol{n}_j(\boldsymbol{b}) = (\cos Af(2^j, \boldsymbol{b}), \, \sin Af(2^j, \boldsymbol{b}))$$

- An edge point $\boldsymbol{b}_0$ at the scale 2^j: $Mf(2^j, \boldsymbol{b})$ is locally maximum at $\boldsymbol{b} = \boldsymbol{b}_0$ when $\boldsymbol{b} = \boldsymbol{b}_0 + \lambda \boldsymbol{n}_j(\boldsymbol{b}_0)$ and $|\lambda|$ small enough.

- The level sets of $g(\boldsymbol{x})$ are the curves $\boldsymbol{x}(s)$ where $g(\boldsymbol{x}(s))$ is constant. If $\boldsymbol{\tau} \perp \boldsymbol{x}(s)$ then
$$\frac{\partial \boldsymbol{x}(s)}{\partial s} = \nabla g \cdot \boldsymbol{\tau} = 0$$

The dyadic wavelet transform

- The level set property applied to $g = f * \check{\theta}_{2^j}$ proves that a maximum point $\boldsymbol{b}_0$ the vector $\boldsymbol{n}_j(\boldsymbol{b}_0)$ of angle $Af(2^j, \boldsymbol{b}_0)$ is perpendicular to the level set of $f * \check{\theta}_{2^j}$ going through $\boldsymbol{b}_0$.

- If the intensity profile remains constant along an edge, then the inflection points (maxima points) are along a level set. The intensity profile of an edge may not be constant but its variations are often negligible over a neighborhood of size 2^j for a small scale 2^j. The tangent of the maxima curve is then nearly perpendicular to $\boldsymbol{n}_j(\boldsymbol{b}_0)$
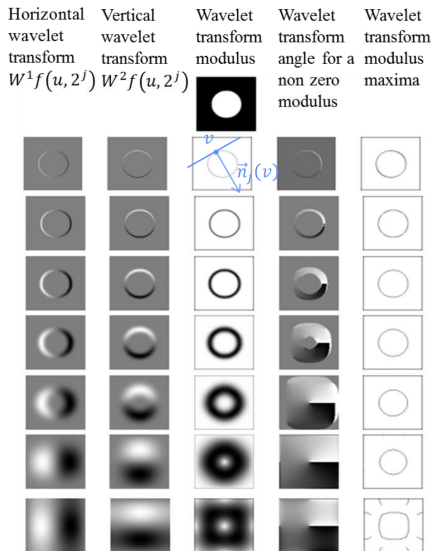
Reconstruction of edge curves
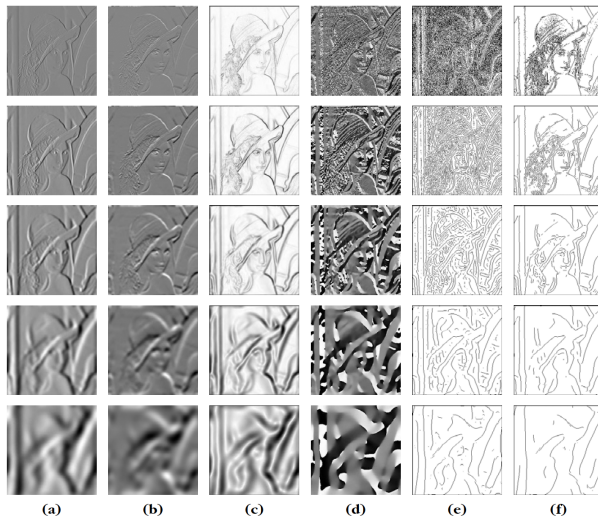

FIGURE 6.11

Multiscale edges of the Lena image shown in Figure 6.12. **(a)** $\{W^1 f(u, 2^j)\}_{-7 \leqslant j \leqslant -3}$. **(b)** $\{W^2 f(u, 2^j)\}_{-7 \leqslant j \leqslant -3}$. **(c)** $\{Mf(u, 2^j)\}_{-7 \leqslant j \leqslant -3}$. **(d)** $\{Af(u, 2^j)\}_{-7 \leqslant j \leqslant -3}$. **(e)** Modulus maxima support. **(f)** Support of maxima with modulus values above a threshold.

Credits: Mallat

Application: characterization of the singularities

Regularity of edge curves

Let $0 \le \alpha < 1$. $f(x,y)$ Lipschitz-α at (x_0, y_0) if $\exists A$ s.t $\forall \boldsymbol{h} = (h_1, h_2)$,

$$|f(x_0 + h_1, y_0 + h_2) - f(x_0, y_0)| \le A\|\boldsymbol{h}\|^\alpha$$
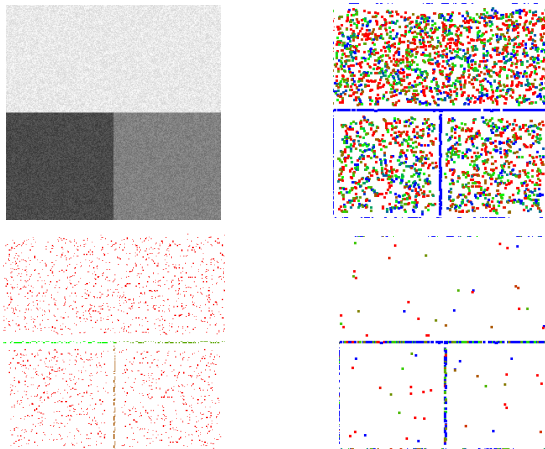

Contour   Contour
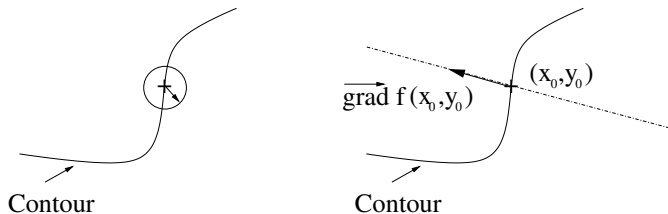
On a curve of discontinuity, the estimation of the regularity reduces to the one dimensional case. f is uniformly Lipschitz-α inside Ω iff

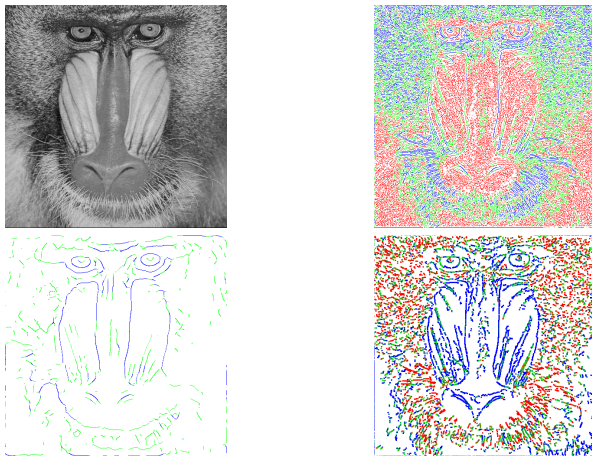$$\forall (x,y) \in \Omega, \ \forall j, \quad |Mf(x,y,2^j)| \le A\, 2^{j(\alpha+1)}$$

Computation of the Lipschitz regularity

The Lipschitz regularity is evaluated at each edge point, by computing the slope of $\log Mf(x_c, y_c, a) = g(\log a)$

Figure: Three noisy domain: maps of modulus, Lipschitz regularity, denoising

Examples

Figure: Mandrill original image (top left), large scale edge points (bottom left) and fine scale edge points (top right) and local regularities computed on maxima lines (bottom right)

Examples
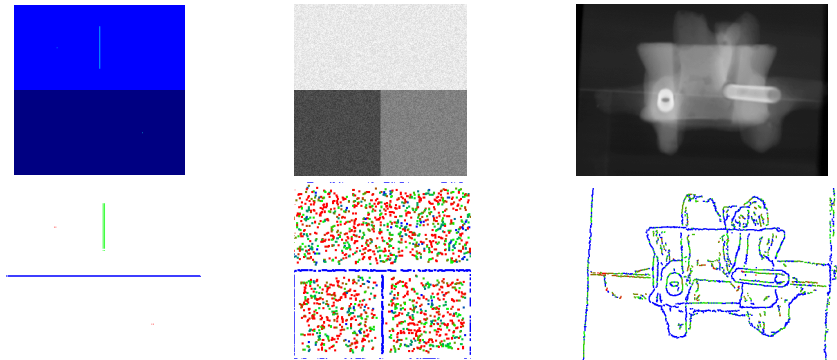
Figure: Top: original images; Bottom: edge points (the colors represent the regularity parameter)
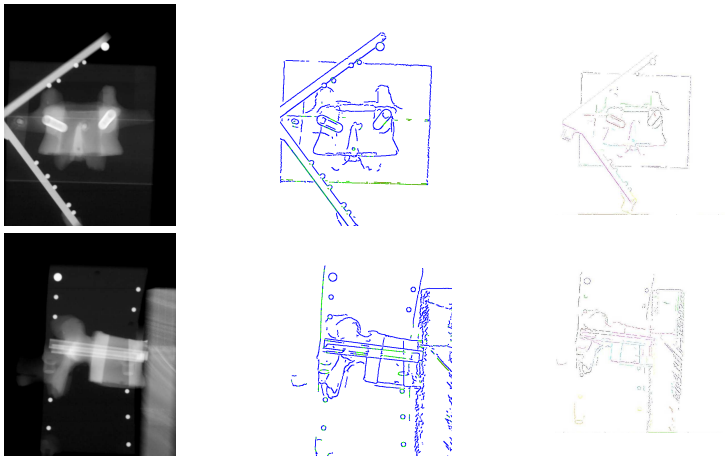
Examples

Figure: Edge detection on two X-rays of vertebra

Reconstruction from edges

Image approximations can be computed by projecting the image on the space generated by wavelets on the modulus maxima support. Let Λ be the set of all modulus maxima points $(2^j, \boldsymbol{b})$, $\boldsymbol{n}$ is the unit vector in the direction $Af(2^j, \boldsymbol{b})$ and

$$\psi^3_{2^j, \boldsymbol{b}}(\boldsymbol{x}) = 2^{2j} \frac{\partial^2 \theta_{2^j}(\boldsymbol{x} - \boldsymbol{b})}{\partial \boldsymbol{n}^2}$$

Since the wavelet modulus $Mf(2^j, \boldsymbol{b})$ has a local maximum at $\boldsymbol{b}$ in the direction of $\boldsymbol{n}$ then $\langle f, \psi^3_{2^j, \boldsymbol{b}} \rangle = 0$.

A modulus maxima approximation f_Λ can be computed as an **orthonormal projection** of f on the space generated by the family of maxima wavelets $\{\psi^k_{2^j, \boldsymbol{b}}\}_{(2^j, \boldsymbol{b}) \in \Lambda, 1 \le k \le 3}$:

$$f_\Lambda = \mathbf{L}^{-1}(\mathbf{L}f), \quad \mathbf{L}y = \sum_{(2^j, \boldsymbol{b}) \in \Lambda} \sum_{k=1}^{2} \langle y, \psi^k_{2^j, \boldsymbol{b}} \rangle \psi^k_{2^j, \boldsymbol{b}}$$

Credits: Mallat (see chapter 5 on frames and especially section 5.1.3 on dual synthesis)

Reconstruction from edges

(a)   (b)

(c)

FIGURE 6.12
(a) Original Lena image. (b) Image reconstructed from the wavelet maxima displayed in Figure 6.11(e) and larger-scale maxima. (c) Image reconstructed from the thresholded wavelet maxima displayed in Figure 6.11(f) and larger-scale maxima.

Credits: Mallat
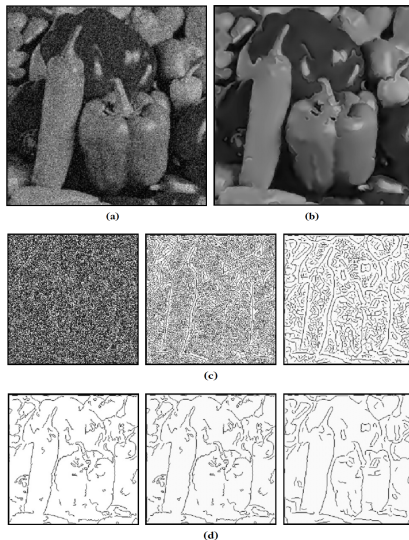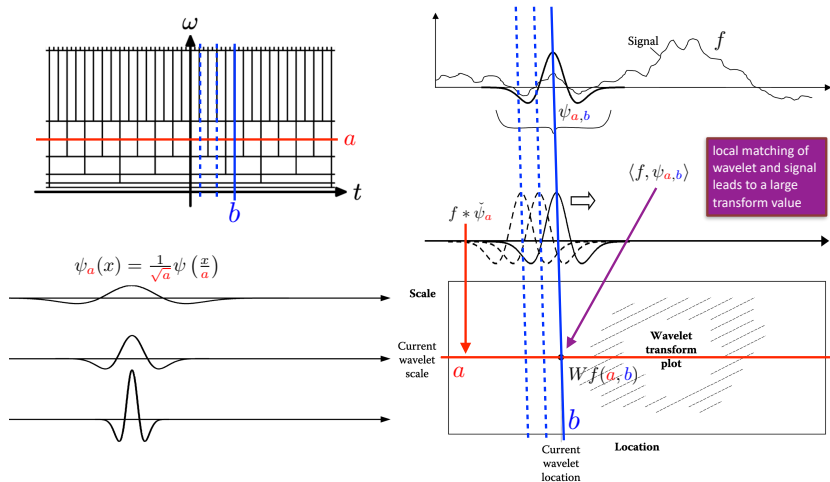
Denoising by multiscale edge thresholding

FIGURE 6.13

(a) Noisy peppers image. (b) Peppers image restored from the thresholding maxima chains shown in (d). The images in row (c) show the wavelet maxima support of the noisy image—the scale increases from left to right, from 2^{-7} to 2^{-5}. The images in row (d) give the maxima support computed with a thresholding selection of the maxima chains.

Credits: Mallat

The 1D Discrete Wavelet Transform

From the CWT to the DWT

$$\psi_a(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x}{a}\right)$$

local matching of wavelet and signal leads to a large transform value

Scaling function

When $Wf(a, b)$ is known only for $a < a_0$, to recover f we need a complement of information that corresponds to $Wf(a, b)$ for $a > a_0$. This is obtained by introducing a scaling function ϕ that is an aggregation of wavelets at scales larger than 1:

$$|\widehat{\phi}(\omega)|^2 = \int_1^{+\infty} |\widehat{\psi}(a\omega)|^2 \frac{\mathrm{d}a}{a} = \int_\omega^{+\infty} \frac{|\widehat{\psi}(\xi)|^2}{|\xi|}\, \mathrm{d}\xi$$

and the complex phase of $\widehat{\phi}(\omega)$ can be arbitrarily chosen. One can verify that $\|\phi\| = 1$, and from admissibility condition that $\lim_{\omega \to 0} |\widehat{\phi}(\omega)|^2 = C_\psi$. The scaling function therefore can be interpreted as the impulse response of a **low-pass filter**. Let us denote $\phi_a(x) = a^{-1/2}\phi(x/a)$ and $\breve{\phi}_a(x) = \phi_a^*(-x)$. The low-frequency approximation of f at scale a is $Lf(a, b) = f * \breve{\phi}_a(b)$ and it can be shown that:

$$f(x) = \frac{1}{C_\psi a_0} Lf(a_0, \cdot) * \phi_{a_0}(x) + \frac{1}{C_\psi} \int_0^{a_0} Wf(a, \cdot) * \psi_a(x) \frac{\mathrm{d}a}{a^2}$$

From the CWT to the DWT

- We need to discretize the CWT for numerical applications
- It requires to choose a sampling grid, that is a **discrete lattice**

$$\Gamma = \{a_j, b_{j,k}, j, k \in \mathbb{Z}\}$$

- Noting $\psi_{j,k} = \psi_{a_j, b_{j,k}}$ and $\tilde{\psi}_{j,k}$ explicitly derived from $\psi_{j,k}$ we want:
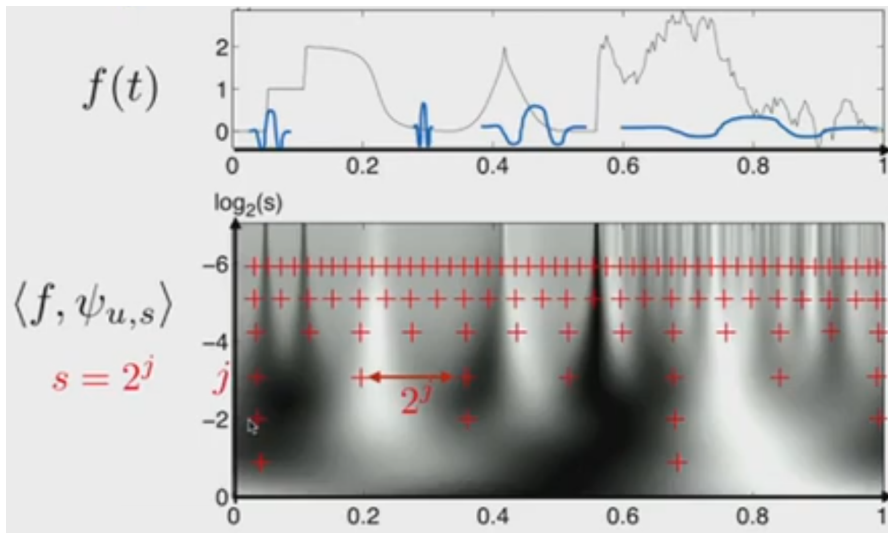
$$f = \sum_{j,k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \tilde{\psi}_{j,k}$$

- The dyadic grid corresponds to the choice $a_j = 2^{-j}$ and $b_{j,k} = k2^{-j}$

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad j, k \in \mathbb{Z}$$

$\Rightarrow$ mostly leads to frames not bases.

From the CWT to the DWT

Credits: S. Mallat

From the CWT to the DWT

Definition (Frame)

$\{\psi_{j,k}\}$ is is a frame in the Hilbert space $\mathcal{H}$ if there exists $B \geqslant A > 0$ such that

$$A\|f\|^2 \leqslant \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_{j,k}\rangle|^2 \leqslant B\|f\|^2$$

- A, B are the frame bounds
- $A = B \neq 1$ is a tight frame
- $A = B = 1$ and $\|\psi_{j,k}\| = 1$ is an orthonormal basis

$\Rightarrow$ Given a wavelet ψ we need to find lattice Γ such that $\{\psi_{j,k}\}$ is a "good frame" that is $\frac{A}{B} \approx 1$.

(Example)

Mexican hat and Morlet wavelets give good nontight frames for a dyadic lattice, adapted to geometry.

From frames to bases

Questions

- Can we reconstruct any function of Hilbert space from the discrete subset of wavelet coefficients?

- Is there a basis of orthogonal wavelets on $L^2(\mathbb{R})$?

- How can we construct such wavelets? With specific properties: regular, with compact support, ...

- Is there a fast algorithm to compute them?

The effervescence

- **Meyer** made the link with the Calderon's identity

$$f(x) = \int_0^{+\infty} \int_{\mathbb{R}} Wf(a, b)\psi_{a,b}(x) \, db \frac{da}{a^2}$$

- **Meyer**, **Grossmann**, **Daubechies** (1985): construction of $L^2(\mathbb{R})$ bases:

$$f(x) = \sum_{j,k} a_{j,k} 2^{j/2} \psi(2^j x - k)$$

- **Meyer**, **Malat** (1986): Fast Wavelet Transform (FWT)

Yves Meyer   Ingrid Daubechies   Stéphane Mallat

From Fourier series to Wavelet series

$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k} \psi(2^j x - k)$$
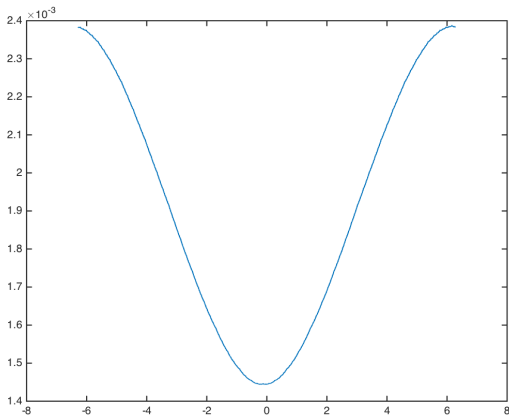
Figure: For $J = 0$ the approximation contains $N = 1$ terms

From Fourier series to Wavelet series

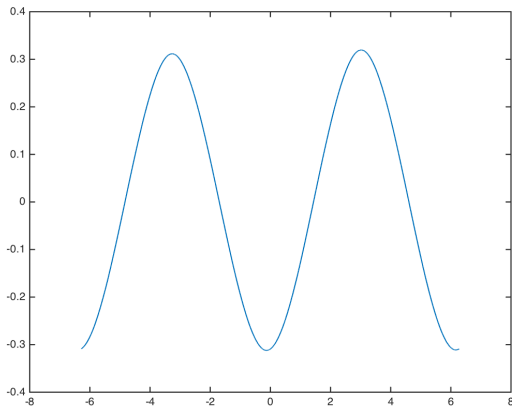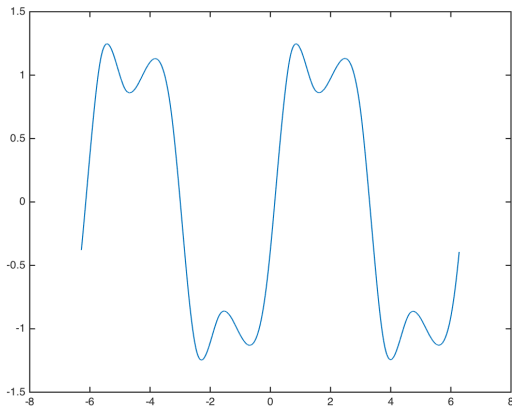$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k} \psi(2^j x - k)$$

Figure: For $J = 0$ the approximation contains $N = 1 + 2$ terms

From Fourier series to Wavelet series

$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k} \psi(2^j x - k)$$

Figure: For $J = 0$ the approximation contains $N = 1 + 2 + 4$ terms

From Fourier series to Wavelet series

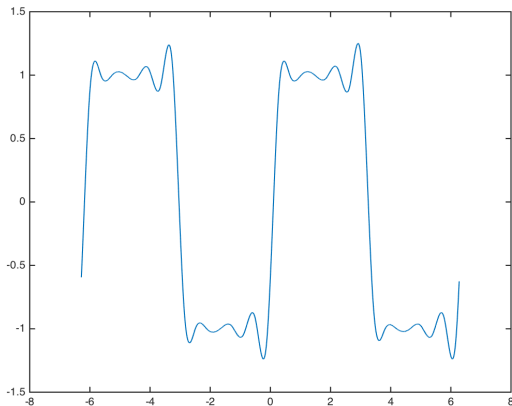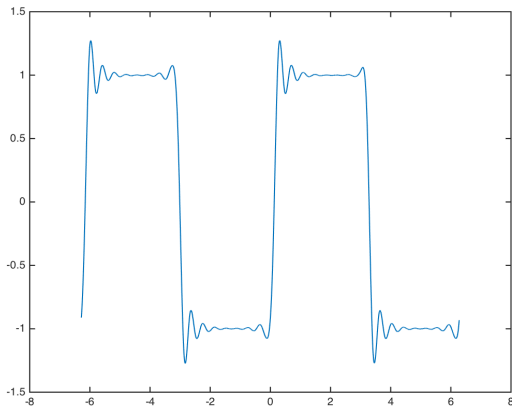$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k} \psi(2^j x - k)$$

Figure: For $J = 0$ the approximation contains $N = 1 + 2 + 4 + 8$ terms

From Fourier series to Wavelet series

$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k}\psi(2^j x - k)$$

Figure: For $J = 0$ the approximation contains $N = 1 + 2 + 4 + 8 + 16$ terms

From Fourier series to Wavelet series

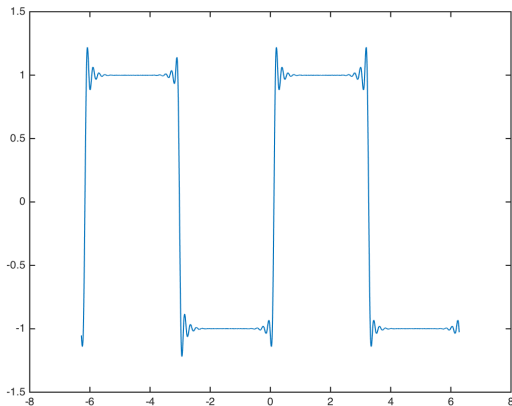$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k} \psi(2^j x - k)$$

Figure: For $J = 0$ the approximation contains $N = 1 + 2 + 4 + 8 + 16 + 32$ terms

From Fourier series to Wavelet series

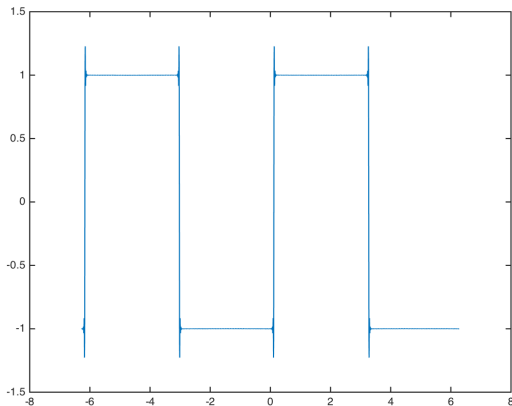$$f(x) = \sum_{i=0}^{J} \sum_{k=0}^{2^j-1} c_{j,k} \psi(2^j x - k)$$

Figure: For $J = 0$ the approximation contains $N = 1 + \ldots + 512 = 1023$ terms

From Fourier series to Wavelet series

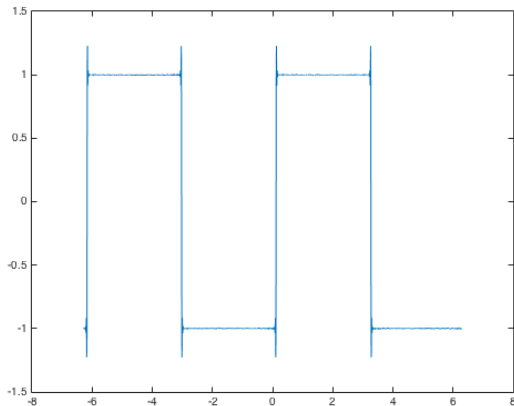$$f(x) \approx \sum_{|c_{j,k}| > 10^{-2}} c_{j,k} \psi(2^j x - k)$$

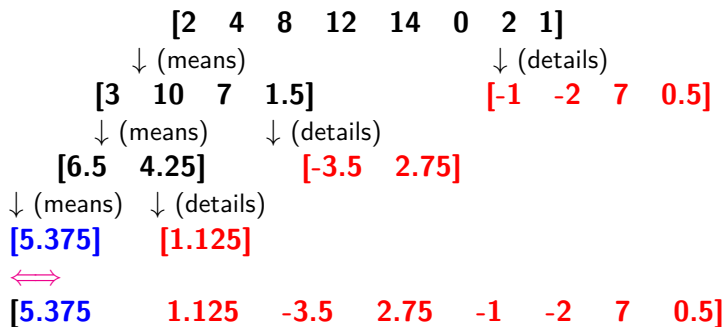Figure: The approximation contains $N = 207$ terms

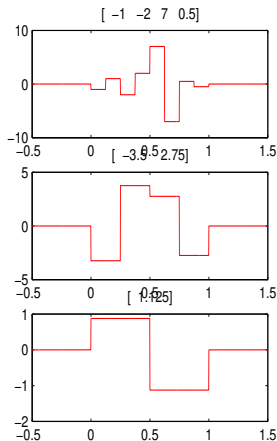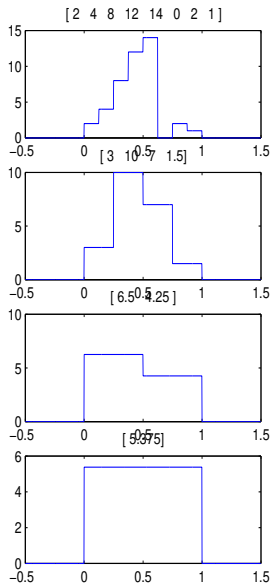# The four musketeers of wavelets

Figure: Stéphane Mallat, Yves Meyer, Ingrid Daubechies & Emmanuel Candès

1. The Haar Basis

Decomposition algorithm

[2   4   8   12   14   0   2   1]

     ↓ (means)   ↓ (details)

[3   10   7   1.5]   **[-1   -2   7   0.5]**

    ↓ (means)   ↓ (details)

[6.5   4.25]   **[-3.5   2.75]**

↓ (means)   ↓ (details)

[5.375]   **[1.125]**

⟺
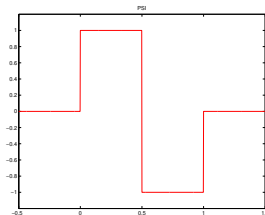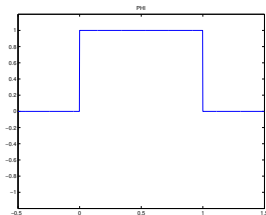
[5.375   1.125   -3.5   2.75   -1   -2   7   0.5]

Decomposition algorithm

Credits: V. Perrier

The Haar basis of $L^2(0,1)$

Let $\varphi = 1$ on $[0,1]$ and $\psi(x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}[\\ -1 & \text{if } x \in [\frac{1}{2}, 1[\end{cases}$

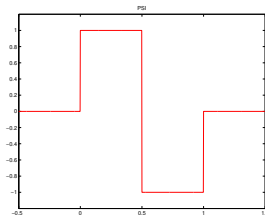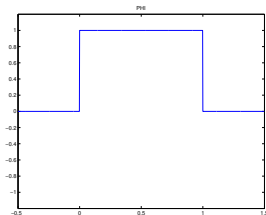For $j \geq 0$ and $0 \leq k \leq 2^j - 1$, one set: $\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k)$ then

$$\psi_{j,k}(x) = \begin{cases} 2^{\frac{j}{2}} & \text{if } x \in [k2^{-j}, (k+\frac{1}{2})2^{-j}[\\ -(2^{\frac{j}{2}}) & \text{if } x \in [(k+\frac{1}{2})2^{-j}, (k+1)2^{-j}[\end{cases}$$

The family $\{\varphi, \psi_{j,k}\}$ is an **orthonormal basis** of $L^2(0,1)$, called Haar basis.

The Haar basis of $L^2(0,1)$

Let $\varphi = 1$ on $[0,1]$ and $\psi(x) = \begin{cases} 1 & \text{if } x \in [0, \frac{1}{2}[\\ -1 & \text{if } x \in [\frac{1}{2}, 1[\end{cases}$
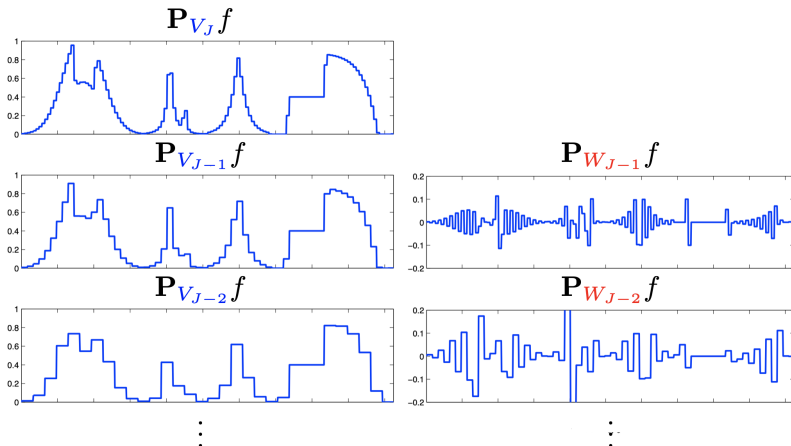

For $j \geq 0$ and $0 \leq k \leq 2^j - 1$, one set: $\varphi_{j,k}(x) = 2^{\frac{j}{2}} \varphi(2^j x - k)$ then

$$\varphi_{j,k}(x) = \begin{cases} 2^{\frac{j}{2}} & \text{if } x \in [k2^{-j}, (k+\frac{1}{2})2^{-j}[\\ 0 & \text{otherwise} \end{cases}$$

Compression: $\varphi_{j,k} = \dfrac{\varphi_{j+1,2k} + \varphi_{j+1,2k+1}}{\sqrt{2}}$, $\psi_{j,k} = \dfrac{\varphi_{j+1,2k} - \varphi_{j+1,2k+1}}{\sqrt{2}}$

The Haar basis of $L^2(0,1)$

- Projection on approx. space: $\mathbf{P}_{V_j} f = \sum_k \langle f, \varphi_{j,k} \rangle \varphi_{j,k} = \sum_k c_{j,k} \varphi_{j,k}$
- Projection on details space: $\mathbf{P}_{W_j} f = \sum_k \langle f, \psi_{j,k} \rangle \psi_{j,k} = \sum_k d_{j,k} \psi_{j,k}$
- Projection on orthogonal spaces: $\mathbf{P}_{V_{j+1}} f = \mathbf{P}_{V_j} f + \mathbf{P}_{W_j} f$

Credits: G. Peyré

The Haar basis of $L^2(0,1)$

- V_j: vector space of constant functions on $\left\{ \left[\frac{k}{2^j}, \frac{k+1}{2^j} \right] \right\}_{k=0,..,2^j-1}$
- The family $\psi_{j,k}(t)$ defines a o.n.b of W_j (dim $2^j - 1$) such that

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

- $f^{j+1}(x) = \mathbf{P}_{V_{j+1}} f(x) = \sum_{k=0}^{2^{j+1}-1} c_{j+1,k} \varphi_{j+1,k}(x)$

- $f^{j+1}(x) = \mathbf{P}_{V_j} f(x) + \mathbf{P}_{W_j} f(x) = \sum_{k=0}^{2^j-1} c_{j,k} \varphi_{j,k}(x) + \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x)$

- $c_{j,k} = \dfrac{c_{j+1,2k} + c_{j+1,2k+1}}{\sqrt{2}}$, $d_{j,k} = \dfrac{c_{j+1,2k} - c_{j+1,2k+1}}{\sqrt{2}}$
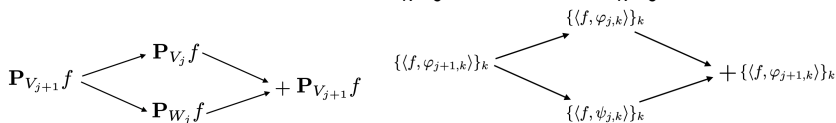
The Haar basis of $L^2(0,1)$

- **Decompression**:

$$\varphi_{j+1,2k} = \frac{\varphi_{j,k} + \psi_{j,k}}{\sqrt{2}}, \quad \varphi_{j+1,2k+1} = \frac{\varphi_{j,k} - \psi_{j,k}}{\sqrt{2}}$$
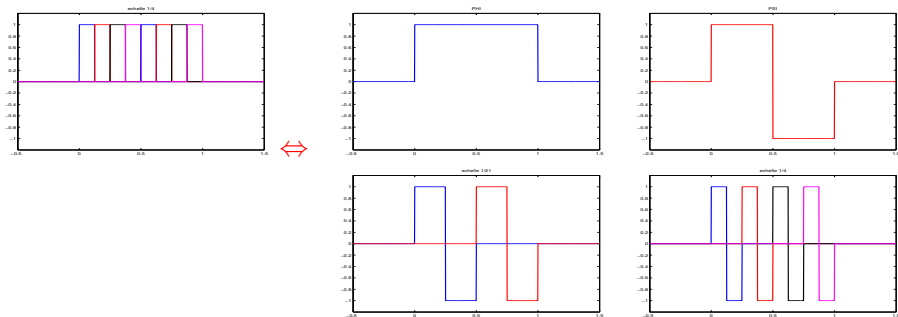
- $f^{j+1}(x) = \mathbf{P}_{V_{j+1}} f(x) = \sum_{k=0}^{2^{j+1}-1} c_{j+1,k} \varphi_{j+1,k}(x)$

- $f^{j+1}(x) = \mathbf{P}_{V_j} f(x) + \mathbf{P}_{W_j} f(x) = \sum_{k=0}^{2^j-1} c_{j,k} \varphi_{j,k}(x) + \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x)$

$$\mathbf{P}_{V_{j+1}} f \nearrow \overset{\mathbf{P}_{V_j} f}{\underset{\mathbf{P}_{W_j} f}{\searrow}} + \mathbf{P}_{V_{j+1}} f \qquad \{\langle f, \varphi_{j+1,k}\rangle\}_k \nearrow \overset{\{\langle f, \varphi_{j,k}\rangle\}_k}{\underset{\{\langle f, \psi_{j,k}\rangle\}_k}{\searrow}} + \{\langle f, \varphi_{j+1,k}\rangle\}_k$$

- $c_{j+1,2k} = \dfrac{c_{j,k} + d_{j,k}}{\sqrt{2}}, \ c_{j+1,2k+1} = \dfrac{c_{j,k} - d_{j,k}}{\sqrt{2}}$

Haar Basis Functions

Two equivalent bases of the piecewise constant function space on [0,1], associated to the subdivision $k/8, k = 0, \ldots, 7$

Credits: V. Perrier

Advantage of the decomposition

The Haar decomposition of a function $f \in L^2(0,1)$ finally writes:

$$f = c_0 + \sum_{j=0}^{+\infty} \sum_{k=0}^{2^j-1} d_{j,k}\,\psi_{j,k}$$

with

$$c_0 = \langle f, \varphi \rangle = \int_0^1 f(x)\,\mathrm{d}x, \quad d_{j,k} = \langle f, \psi_{j,k} \rangle = \int_0^1 f(x)\,\psi_{j,k}(x)\,\mathrm{d}x$$

Local smoothness characterization

(i) if $f \in C^1(I_{j,k})$ then $|d_{j,k}| \le C\,2^{-3j/2}$

(ii) if $f \in C^\alpha(x_0)$ i.e. $|f(x) - f(x_0)| \le k|x - x_0|^\alpha$ $(0 < \alpha < 1)$ then

$$|d_{j,k}| \le C\,2^{-j(\alpha+1/2)}$$

$\Rightarrow$ Useful property for compression!

Proof of (i)

For fixed $j \geq 0$ and $k \in \{0, \dots 2^j - 1\}$, let $I_{j,k} :=]k2^{-j}, (k+1)2^{-j}[$.

$$\text{Supp}\{\psi_{j,k}\} = [k2^{-j}, (k+1)2^{-j}] = \bar{I}_{j,k}$$

The Haar coefficient on $\psi_{j,k}$ of a function f is given by:

$$d_{j,k} = \int_{I_{j,k}} f\psi_{j,k}$$

If $f \in C^1(I_{j,k})$ then for all $x \in I_{j,k}$:

$$f(x) = f\left(x - \left(k + \frac{1}{2}\right)2^{-j}\right) + \left(x - \left(k + \frac{1}{2}\right)2^{-j}\right)f'(\theta_x), \quad \theta_x \in I_{j,k}$$
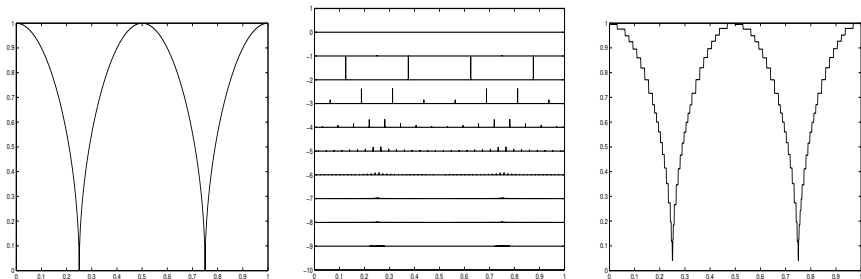
Then,

$$d_{j,k} = \int_{I_{j,k}} \left(x - \left(k + \frac{1}{2}\right)2^{-j}\right)f'(\theta_x)\psi_{j,k}(x)\,\mathrm{d}x$$

since $\int \psi_{j,k} = 0$, hence

$$|d_{j,k}| \leq \sup_{I_{j,k}}|f'|\int_{I_{j,k}}|2^{-j-1}|2^{j/2}\,\mathrm{d}x \leq \frac{1}{2}\sup_{I_{j,k}}|f'|\,2^{-3j/2}$$

Example: $f(x) = \sqrt{|\cos 2\pi x|}$

Left figure: *function f sampled on $1024 = 2^{10}$ values.*

Middle figure: *Haar coefficient map*
(abscissa : $k2^{-j} \in [0,1]$, ordinates: $-j$, $j = 1, \ldots 9$).

Right figure: *Reconstructed function from the 80 largest coefficients (> 0.06)*
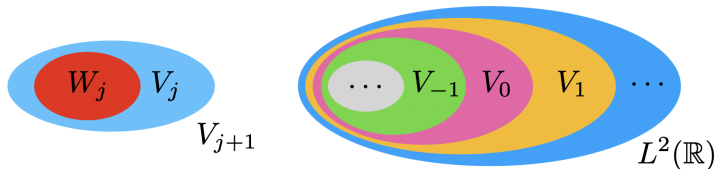*(compression = 92.2 %, L^2-relative error = 6.10^{-3}).*

Credits: V. Perrier

2. Regular wavelet bases

Multiresolution Analysis (MRA)

A multiresolution analysis of $L^2(\mathbb{R})$ is a sequence of closed subspaces $(V_j)_{j\in\mathbb{Z}}$ s.t.:

1. $\forall j \in \mathbb{Z},\ V_j \subset V_{j+1} \subset \cdots \to L^2(\mathbb{R})$,
2. $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ and $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$,
3. $f(x) \in V_j \iff f(2x) \in V_{j+1}$,
4. $f(x) \in V_0 \iff \forall n \in \mathbb{Z},\ f(x-n) \in V_0$,
5. $\exists \varphi \in V_0$ s.t $\{\varphi(x-n) : n \in \mathbb{Z}\}$ is an **orthonormal basis** of V_0.

φ is called the scaling function of the multiresolution analysis.

Multiresolution Analyses – Examples

The spaces V_j are dilation invariant, then:

$$V_j = \text{Vec}\,\{\varphi_{j,k} := 2^{\frac{j}{2}}\varphi(2^j x - k)\,;\ k \in \mathbb{Z}\}$$

Haar:

$$V_0 = \{\text{Piecewiese constant functions on } [k, k+1[,\ \forall k \in \mathbb{Z}\}$$

Splines of degree 1:

$$V_0 = \{\text{Continuous functions on } \mathbb{R},\ \text{affines on } [k, k+1[,\ \forall k \in \mathbb{Z}\}$$

Splines of degree n:

$$V_0 = \{\mathcal{C}^{n-1} \text{ functions on } \mathbb{R},\ \text{piecewise polynomial of deg } n \text{ on } [k, k+1[\}$$

Shannon:

$$V_0 = \{f \in L^2(\mathbb{R})\,;\ \text{supp}\,\widehat{f} \subset [1, 2]\}$$

MRA – Two-scale equation for the scaling function

$V_0 \subset V_1 = \mathrm{span}\{\varphi_{1,k} := \sqrt{2}\varphi(2x - n) \; ; \; n \in \mathbb{Z}\}$, then $\varphi \in V_0$ writes:

$$\varphi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \, \varphi(2x - n) \quad \text{with} \quad h_n = \sqrt{2} \int_{\mathbb{R}} \varphi(x)\varphi(2x - n) \, \mathrm{d}x$$

Applying the Fourier Transform:

$$\widehat{\varphi}(\xi) = m_0\left(\frac{\xi}{2}\right) \widehat{\varphi}\left(\frac{\xi}{2}\right) \quad \text{with} \quad m_0(\xi) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n \mathrm{e}^{-2i\pi n\xi}$$

Assume that $\varphi \in L^1(\mathbb{R})$ and $\int_{\mathbb{R}} \varphi = 1$, then:

$$\hat{\varphi}(\xi) = \prod_{j=0}^{\infty} m_0\left(\frac{\xi}{2^j}\right)$$

(h_n) is a low pass filter and m_0 is its **transfer function**.

MRA – Construction of the wavelets

$V_j \subset V_{j+1}$, let W_j be the orthogonal complement space of V_j in V_{j+1}:

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

One searches for a function ψ s.t. $\{\psi(x - n) : n \in \mathbb{Z}\}$ is an **orthonormal basis** of W_0. Since $\psi \in W_0 \subset V_1$, one searches for g_n such that

$$\psi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \, \varphi(2x - n)$$

This is equivalent in Fourier domain to:

$$\widehat{\psi}(\xi) = m_1 \left(\frac{\xi}{2} \right) \widehat{\varphi} \left(\frac{\xi}{2} \right) \quad \text{with} \quad m_1(\xi) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} g_n \mathrm{e}^{-2i\pi n \xi}$$

$\Rightarrow$ What are the assumptions on filters (h_n) and (g_n) in order to construct a scaling function φ and a wavelet ψ generating a MRA?

Detail filter (necessary) constraints for h

$\boxed{\Rightarrow}$ If $\{\varphi_{j,n}\}$ is an orthonormal basis of V_j then:

1. From the two-scale equation it comes
$$\widehat{h}(0) = \sqrt{2} \quad (C_1)$$

2. $\{\varphi(\cdot - n)\}_n$ orthogonal is equivalent to:
$$\forall n \in \mathbb{N}, \quad \varphi \star \check{\varphi}(n) = \delta[n] \iff \sum_k |\widehat{\varphi}(\xi + 2k\pi)|^2 = 1$$

since sampling a function periodizes its Fourier transform.
Inserting $\widehat{\varphi}(\xi) = 2^{-1/2}\widehat{h}(\xi/2)\widehat{\varphi}(\xi/2)$ and separating even and odd integers terms (with $\widehat{h}$ is 2π-periodic) yields:

$$\left|\widehat{h}\left(\frac{\xi}{2}\right)\right|^2 \sum_{p=-\infty}^{+\infty} \left|\widehat{\varphi}\left(\frac{\xi}{2} + 2p\pi\right)\right|^2 + \left|\widehat{h}\left(\frac{\xi}{2} + \pi\right)\right|^2 \sum_{p=-\infty}^{+\infty} \left|\widehat{\varphi}\left(\frac{\xi}{2} + \pi + 2p\pi\right)\right|^2 = 2$$

Putting $\xi' = \xi/2$ and $\xi' = \xi/2 + \pi$ in the two sums yields:
$$|\widehat{h}(\xi')|^2 + |\widehat{h}(\xi' + \pi)|^2 = 2 \quad (C_2)$$

Detail filter (sufficient) constraints for h

$\boxed{\Leftarrow}$ Conversely, the following theorem gives sufficient conditions on $\widehat{h}$ to guarantee that this infinite product is the Fourier transform of a scaling function:

Theorem (Mallat, Meyer)

Ff $\widehat{h}(\xi)$ is 2π-periodic and continuously differentiable in a neighborhood of $\xi = 0$, if it satisfies $(C_1), (C_2)$ and

$$\inf_{\xi \in [-\pi/2, \pi/2]} |\widehat{h}(\xi)| > 0 \quad (C_3)$$

then

$$\widehat{\varphi}(\xi) = \prod_{p=1}^{+\infty} \frac{\widehat{h}(2^{-p}\xi)}{\sqrt{2}}$$

Detail filter (necessary) constraints for g

$\boxed{\Rightarrow}$ If $\{\psi_{j,n}\}$ is an orthonormal basis of W_j then:

❶ $\{\psi(\cdot - n)\}_n$ orthogonal is equivalent to:
$$\forall n \in \mathbb{N}, \quad \psi \star \breve{\psi}(n) = \delta[n] \Longleftrightarrow \sum_k |\widehat{\psi}(\xi + 2k\pi)|^2 = 1$$

Inserting $\widehat{\psi}(\xi) = 2^{-1/2}\widehat{g}(\xi/2)\widehat{\varphi}(\xi/2)$ and separating even and odd integers terms (with $\widehat{g}$ 2π-periodic) also yields:
$$|\widehat{g}(\xi)|^2 + |\widehat{g}(\xi + \pi)|^2 = 2 \quad (C_4)$$

❷ $\{\psi(\cdot - n)\}_n$ orthogonal to $\{\varphi(\cdot - n)\}_n$ is equivalent to:
$$\forall n \in \mathbb{N}, \quad \psi \star \breve{\varphi}(n) = 0 \Longleftrightarrow \sum_k \widehat{\psi}(\xi + 2k\pi)\widehat{\varphi}^*(\xi + 2k\pi) = 0$$

which leads to:
$$\widehat{g}(\xi)\widehat{h}(\xi)^* + \widehat{g}(\xi + \pi)\widehat{h}(\xi + \pi)^* = 0 \quad (C_5)$$

Detail filter (sufficient) constraints for g

$\boxed{\Leftarrow}$ Conversely, the following theorem gives sufficient conditions on $\widehat{h}$ and $\widehat{g}$ to guarantee that the constructed wavelets $\{\psi(\cdot - n)\}_n$ give an orthonormal basis of W_j:

Theorem (Mallat, Meyer)

Under conditions $(C_1) - (C_2) - (C_3)$

$$\{\psi(\cdot - n)\}_n \text{ orthonormal basis of } W_j \Longleftrightarrow (C_4) + (C_5)$$
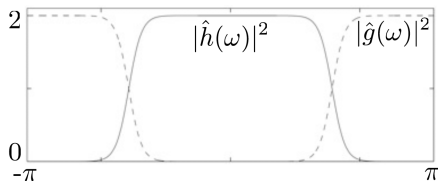
Quadrature mirror filters:

$$\widehat{g}(\xi) = \mathrm{e}^{-i2\pi\xi}\widehat{h}(\xi + \pi) \Longleftrightarrow g[n] = (-1)^{1-n}h[1-n]$$

MRA – Wavelet decomposition

$$L^2(\mathbb{R}) = V_0 \bigoplus_{j=0}^{+\infty} W_j = \bigoplus_{j=-\infty}^{+\infty} W_j$$

$$W_j = \text{Vec}\,\{\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)\;;\;k \in \mathbb{Z}\}$$

Let $f \in L^2(\mathbb{R})$. Its wavelet decomposition writes:

$$f(x) = \sum_{k \in \mathbb{Z}} c_k \varphi(x-k) + \sum_{j=0}^{+\infty} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x) = \sum_{j=-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} d_{j,k}\,\psi_{j,k}(x)$$

with $c_k = \langle f, \varphi(\cdot - k) \rangle$ and $d_{j,k} = \langle f, \psi_{j,k} \rangle$.

Property of wavelet bases

- Vanishing Moments. A wavelet ψ satisfies:

$$\int_{\mathbb{R}} \psi = 0$$

One usually impose N Vanishing Moments:

$$\int_{\mathbb{R}} x^n \psi = 0, \quad \forall n = 0, \dots N - 1$$

- Characterization of the local smoothness of f. For $n \leq N$, $\alpha < N$,
  (i) if $f \in C^n(V_{x_0})$ then $|d_{j,k}| \leq C\, 2^{-j(n+1/2)}$ (for $k2^{-j}$ "neighbor" of x_0)

  (ii) if $f \in C^\alpha(x_0)$ i.e. $|f^{[\alpha]}(x) - f^{[\alpha]}(x_0)| \leq k|x - x_0|^{\alpha - [\alpha]}$ then

  $$|d_{j,k}| \leq C\, 2^{-j(\alpha+1/2)} \quad \text{(for } k2^{-j} \text{ "neighbor" of } x_0)$$

  $\Rightarrow$ Important property in view of compression

Examples of wavelets constructed from filters

Debauchies family

1. $\widehat{h}(0) = \sqrt{2}$
2. $|\widehat{h}(\xi)|^2 + |\widehat{h}(\xi + \pi)|^2 = 2$
3. p vanishing moments $\Leftrightarrow \forall k < p, \ \dfrac{\mathrm{d}^k \widehat{h}}{\mathrm{d}\xi^k}(\pi) = 0$

$\Rightarrow$ orthogonal wavelets with minimal support $2p - 1$

- $p = 1$ (Haar): $h = [0.7071, 0.7071]$
- $p = 2$: $h = [0.4830, 0.8365, 0.2241, -0.1294]$
- $p = 3$: $h = [0, 0.3327, 0.8069, 0.4599, -0.1350, -0.0854, 0.0352]$
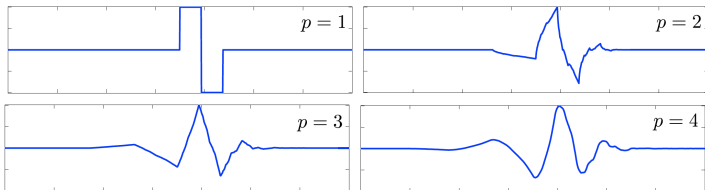
Credits: G. Peyré

Examples of scaling functions and wavelets
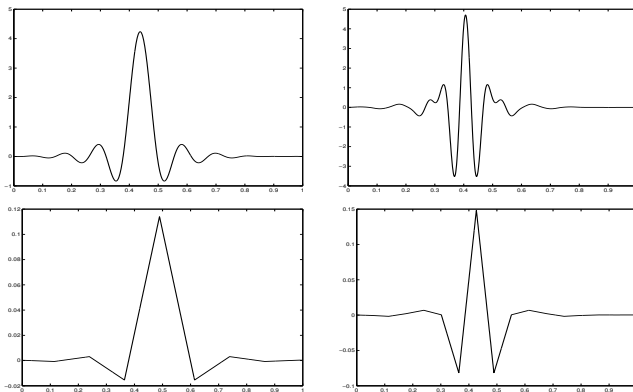
Scaling function (left) and *wavelet* (right):
1st line: Meyer functions (C^∞ and infinite number of vanishing moments).
2d line: Splines of degree 1 (2 vanishing moments).

Examples of scaling functions and wavelets
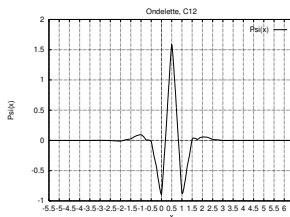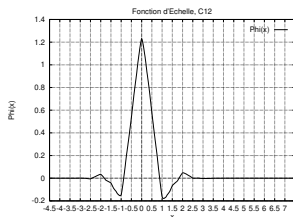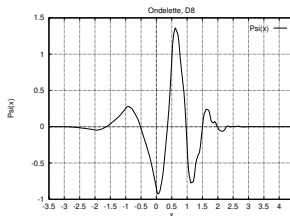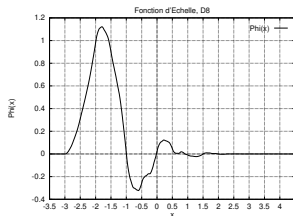
Scaling function (left) and wavelet (right) compactly supported:
1st line: D8 (4 vanishing moments).
2d line: Coifman C12 (4 vanishing moments).

Fast Wavelet transform (FWT)

Let f be a discrete 1D signal discret 1D of length $N = 2^J$.
Step 0 of the algorithm: computing the coefficients $c_J = (c_{J,k})$

$$c_{J,k} \approx 2^{-\frac{J}{2}} f(k 2^{-J}), \quad k \in \mathbb{Z},$$

(using $\int \varphi = 1$). One consider the function f_J of V_J:

$$f_J = \sum_{k \in \mathbb{Z}} c_{Jk} \varphi_{Jk}$$

Decomposition: $V_J = V_0 \oplus W_0 \oplus \cdots \oplus W_{J-1}$
For $j = J, \ldots, 1$ one uses $V_j = V_{j-1} \oplus W_{j-1}$, and then $\forall k \in \mathbb{Z}$:

$$c_{j-1,k} = \sum_{n \in \mathbb{Z}} c_{j,n} h_{n-2k}$$

$$d_{j-1,k} = \sum_{n \in \mathbb{Z}} c_{j,n} g_{n-2k}$$

Fast Wavelet transform (FWT)

Proof: From the two-scale equation

$$\varphi(x) = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \, \varphi(2x - n)$$

by replacing $x \leftarrow 2^j x - k$ and multiplying by $2^{j/2}$ we get:

$$
\begin{aligned}
2^{j/2}\varphi(2^j x - k) &= 2^{j/2}\sqrt{2}\sum_{n \in \mathbb{Z}} h_n \varphi(2^{j+1}x - (n + 2k)) \\
\varphi_{j,k}(x) &\overset{n \leftarrow n-2k}{=} \sum_{n \in \mathbb{Z}} h_{n-2k}\varphi_{j+1,n} \\
c_{j,k} &\overset{\langle f, \cdot \rangle}{=} \sum_{n \in \mathbb{Z}} h_{n-2k} c_{j+1,n}
\end{aligned}
$$

Example (Haar wavelets)

- Low-pass filter: $h = [\cdots, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \cdots]$
- High-pass filter: $g = [\cdots, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, \cdots]$

Fast Wavelet transform (FWT)

Noting $c_j = (c_{j,k})_{k \in \mathbb{Z}}$:

> **convolution - decimation:**
>
> $$c_{j-1}[k] = (c_j \star \check{h})[2k], \quad \forall k \in \mathbb{Z}$$
> $$d_{j-1}[k] = (c_j \star \check{g})[2k], \quad \forall k \in \mathbb{Z}$$

with $\check{h}[n] = h[-n]$ and $\check{g}[n] = g[-n]$.
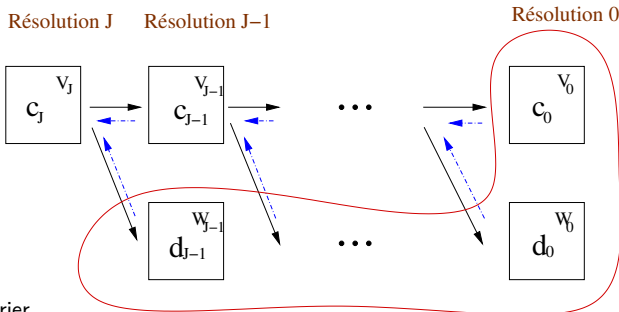
Credits: V. Perrier

Fast Wavelet transform (FWT)

Noting $c_j = (c_{j,k})_{k \in \mathbb{Z}}$:

> convolution - decimation:
>
> $c_{j-1}[k] = (c_j \star \check{h})[2k], \quad \forall k \in \mathbb{Z}$
> $d_{j-1}[k] = (c_j \star \check{g})[2k], \quad \forall k \in \mathbb{Z}$

with $\check{h}[n] = h[-n]$ and $\check{g}[n] = g[-n]$.

Credits: V. Perrier

Fast Wavelet transform (FWT)

Recomposition: From the wavelet coefficients and the scaling coefficients at scale 0: $[c_{0k}, \{d_{jk}\}_{j=0\cdots J-1, k\in\mathbb{Z}}]$, one wants to retrieve the scaling coefficients at finest scale J: $c_J = [(c_{Jk})_{k\in\mathbb{Z}}]$.

One uses $V_{j-1} \oplus W_{j-1} = V_j$, for $j = 0, \ldots, J-1$:

$$c_{jk} = \sum_{n\in\mathbb{Z}} c_{j-1,n} h_{k-2n} + \sum_{n\in\mathbb{Z}} d_{j-1,n} g_{k-2n}, \quad \forall k \in \mathbb{Z}$$

Fast Wavelet transform (FWT)

Recomposition: From the wavelet coefficients and the scaling coefficients at scale 0: $[c_{0k}, \{d_{jk}\}_{j=0\cdots J-1, k\in\mathbb{Z}}]$, one wants to retrieve the scaling coefficients at finest scale J: $c_J = [(c_{Jk})_{k\in\mathbb{Z}}]$.

It writes, in vector form, noting:

$$\widetilde{x_n} = (x \uparrow 2)[n] = \left\{ \begin{array}{ll} x_p & \text{if } n = 2p \\ 0 & \text{if } n = 2p+1 \end{array} \right.$$
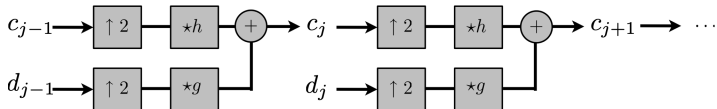
$$\boxed{c_j[k] = (\widetilde{c_{j-1}} \star h)[k] + (\widetilde{d_{j-1}} \star g)[k]}$$

Fast Wavelet Transform algorithm
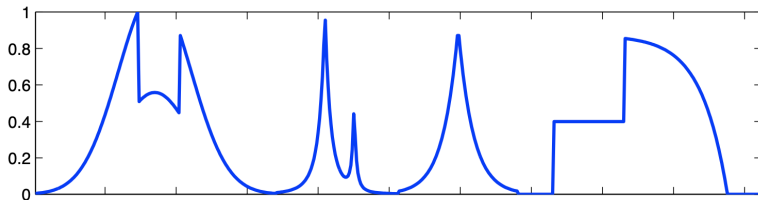
Algorithm (FWT)

Initialization: $c_J = f$, $N = 2^J$

For $j = J, \ldots, 0$
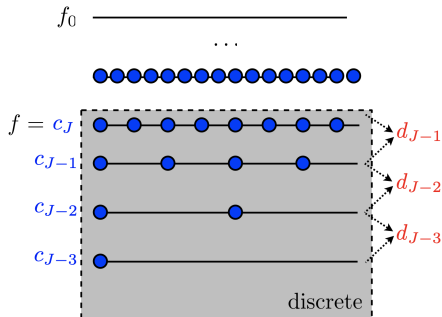
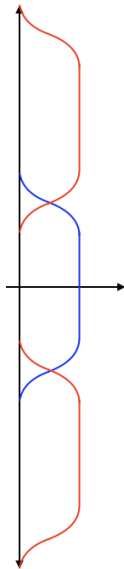$\quad c_{j-1} = (c_j \star \check{h}) \downarrow 2$

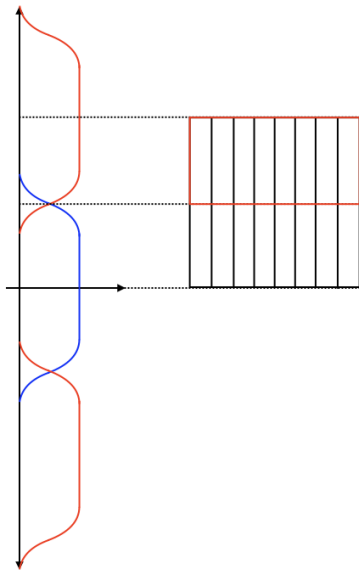$\quad d_{j-1} = (c_j \star \check{g}) \downarrow 2$
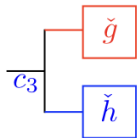
Credits: G. Peyré

Fast Wavelet Transform algorithm
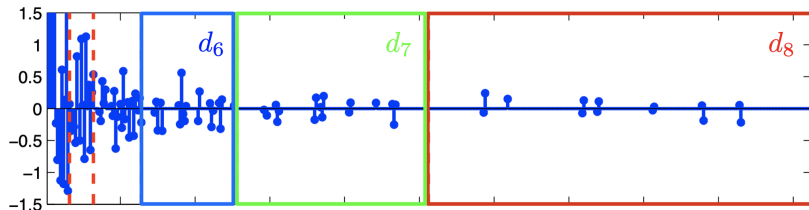
Algorithm (FWT)

Initialization: $c_J = f$, $N = 2^J$

For $j = J, \ldots, 0$

$\quad c_{j-1} = (c_j \star \check{h}) \downarrow 2$

$\quad d_{j-1} = (c_j \star \check{g}) \downarrow 2$
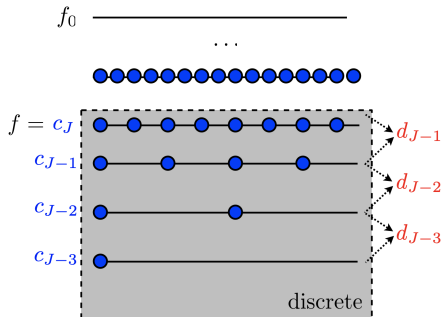
Credits: G. Peyré

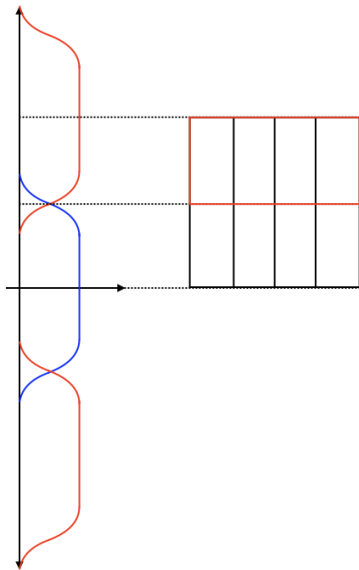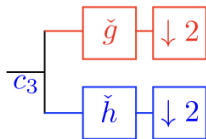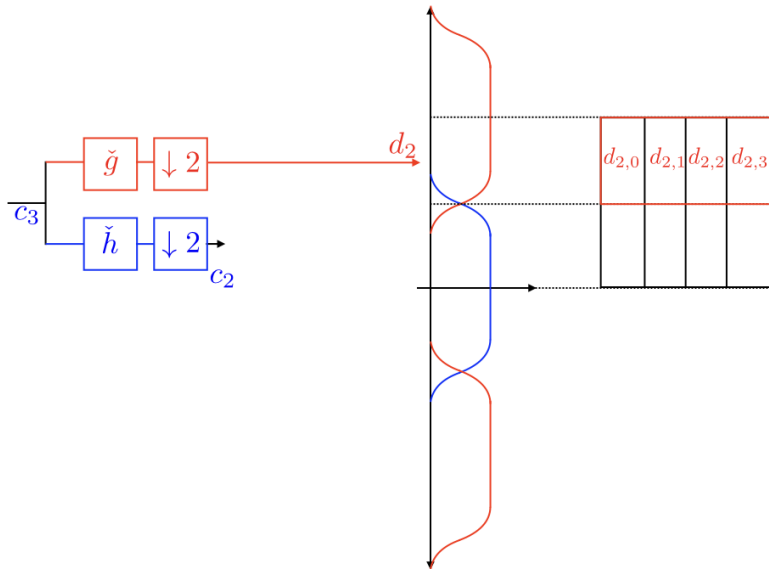# Fast Wavelet Transform algorithm

$\overrightarrow{c_3}$
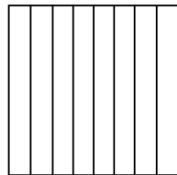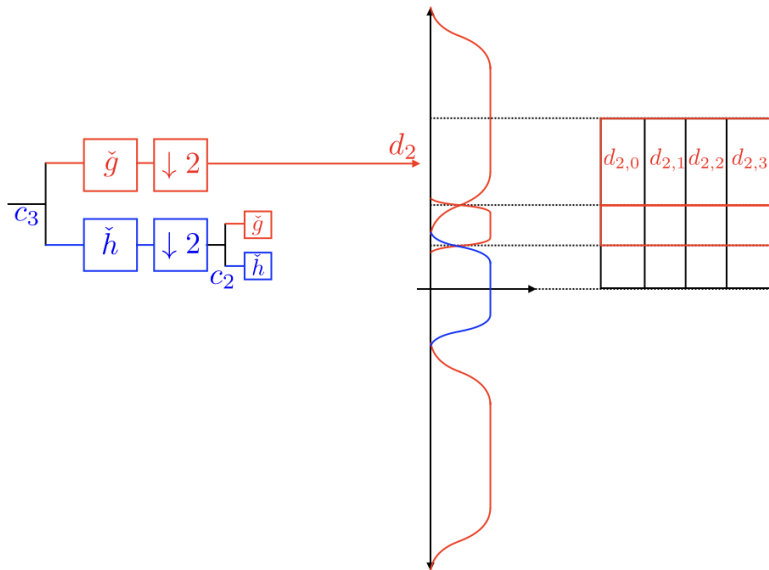

c_3

Fast Wavelet Transform algorithm

Fast Wavelet Transform algorithm
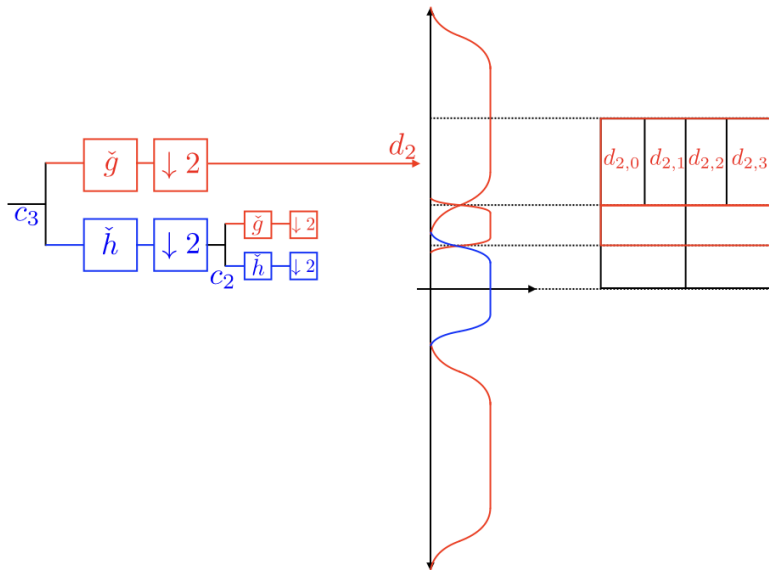
Fast Wavelet Transform algorithm

Fast Wavelet Transform algorithm

Fast Wavelet Transform algorithm

Fast Wavelet Transform algorithm

Fast Wavelet Transform algorithm

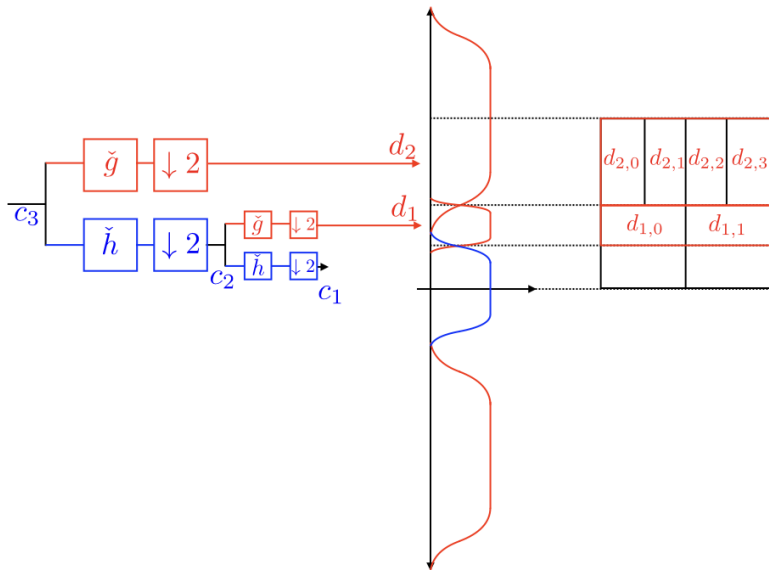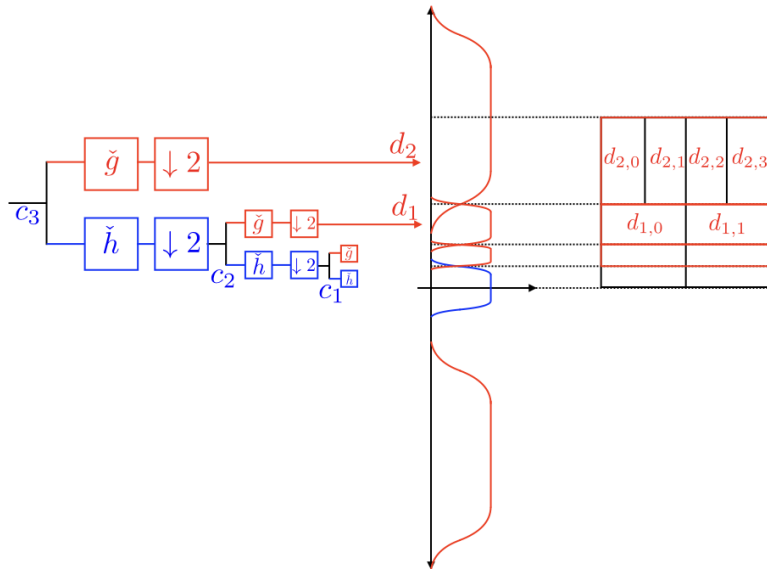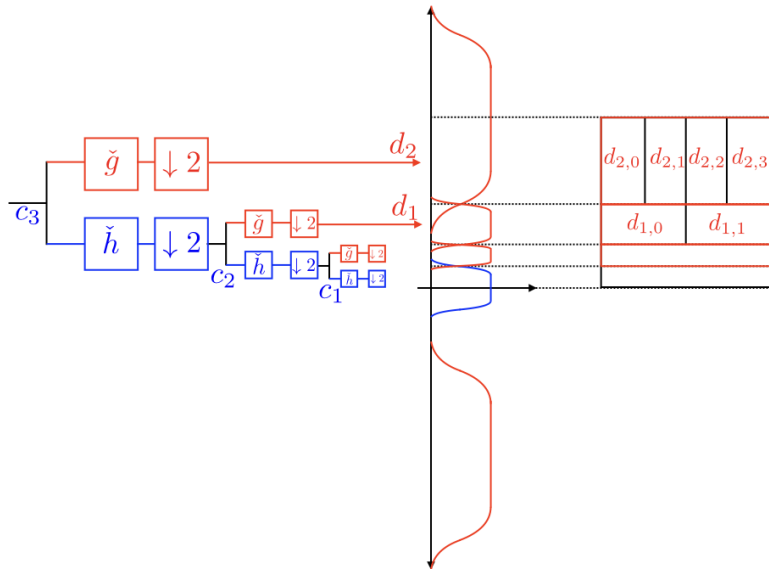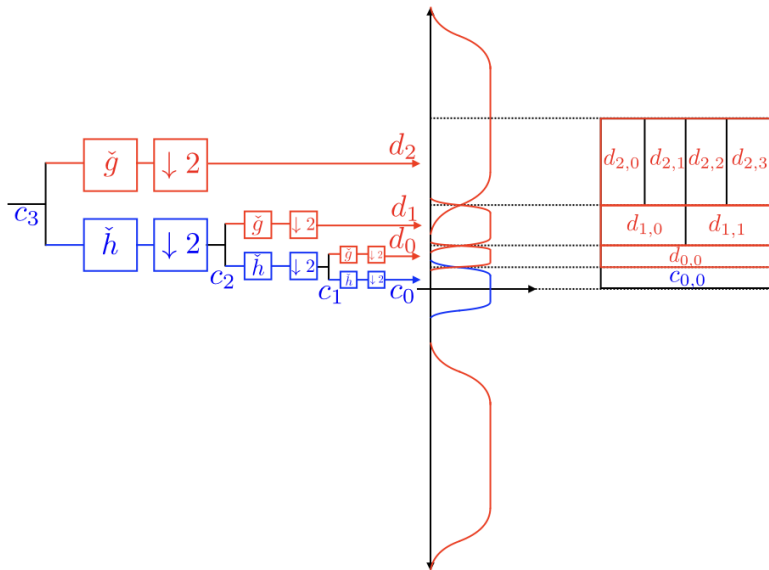# Fast Wavelet Transform algorithm

Fast Wavelet Transform algorithm

Example from Mallat
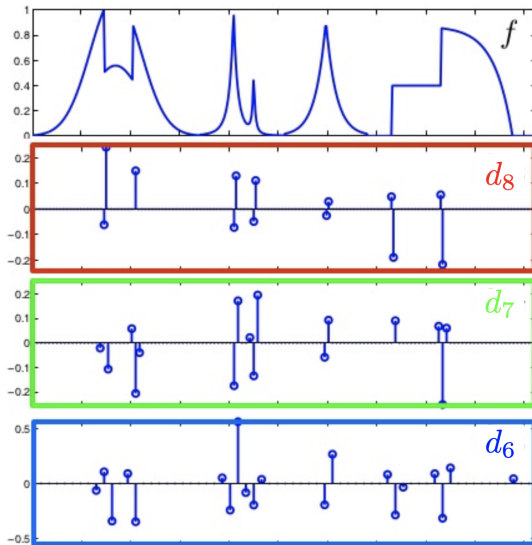
Credits: G. Peyré

Example: $f(x) = \sqrt{|\cos 2\pi x|}$
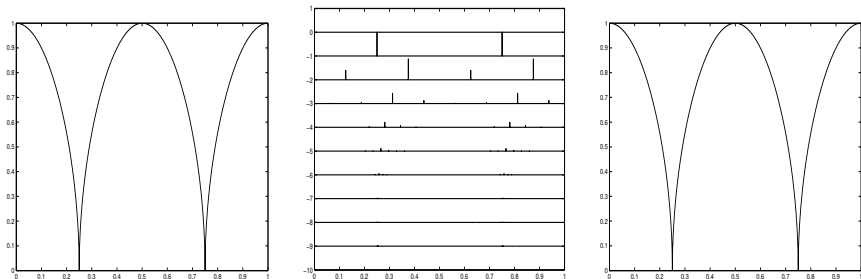
Left figure : *function f discretized on $1024 = 2^{10}$ values.*

Middle figure: *wavelet coefficient map D8*
(abscissa: $k2^{-j} \in [0, 1]$, ordinate: $-j$, $j = 1, \dots 9$).

Right figure: *reconstructed function with the 80 highest coefficients ($> 10^{-3}$)*
(compression = 92.2 %, relative error $L^2 = 3.10^{-7}$).

The 2D Discrete Wavelet Transform

The 2D Haar Basis

From $\varphi(x)$ and $\psi(x)$ one can define the bidimensional functions:

$$\Phi(x,y) = \varphi(x)\varphi(y), \quad \Psi^1(x,y) = \psi(x)\varphi(y)$$
$$\Psi^2(x,y) = \varphi(x)\psi(y), \quad \Psi^3(x,y) = \psi(x)\psi(y)$$

The values of Ψ^1, Ψ^2 and Ψ^3 on $[0,1] \times [0,1]$ are:

For $i = 1, 2, 3$, $j \geq 0$ et $\boldsymbol{k} = (k_x, k_y)$, $0 \leq k_x, k_y \leq 2^j - 1$:

$$\Psi^i_{j,\boldsymbol{k}}(x,y) = 2^{\frac{j}{2}}\Psi^i(2^j x - k_x, 2^j y - k_y)$$

The family $\left\{\Phi, \Psi^1_{j,\boldsymbol{k}}, \Psi^2_{j,\boldsymbol{k}}, \Psi^3_{j,\boldsymbol{k}}\right\}$ is an **orthonormal basis** of $L^2\left([0,1]^2\right)$.

2D Haar expansion
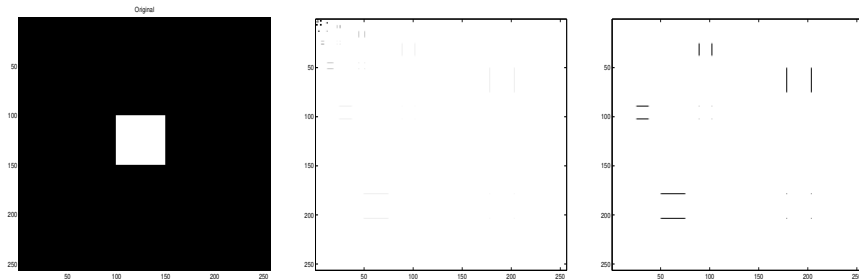
$$f = C_0 + \sum_{j=0}^{+\infty} \sum_{k_x, k_y = 0}^{2^j - 1} \left(D_{j,k}^1 \Psi_{j,k}^1 + D_{j,k}^2 \Psi_{j,k}^2 + D_{j,k}^3 \Psi_{j,k}^3 \right)$$

with $C_0 = \iint_{[0,1]^2} f$ and $D_{j,k}^i = \iint_{[0,1]^2} f \; \Psi_{j,k}^i$

Example: square

Left figure: *Original image.*
Middle figure: *entire Haar coefficient map.*
Right Figure: *coefficients of the two finest scales.*

Image decomposition on the Haar basis

Left figure: *Original image (256² pixels) and its Haar coefficients:*
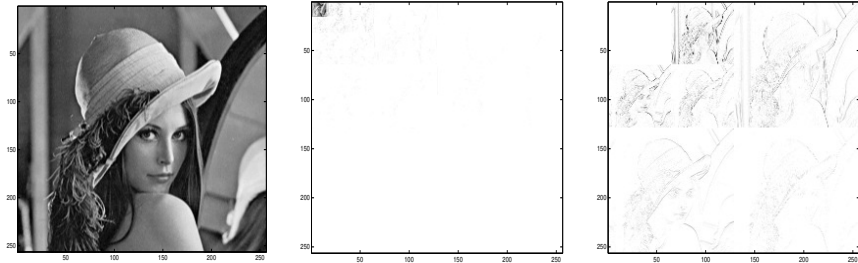Middle figure: *entire Haar coefficient map.*
Right Figure: *coefficients of the two finest scales.*

Image compression with the Haar basis
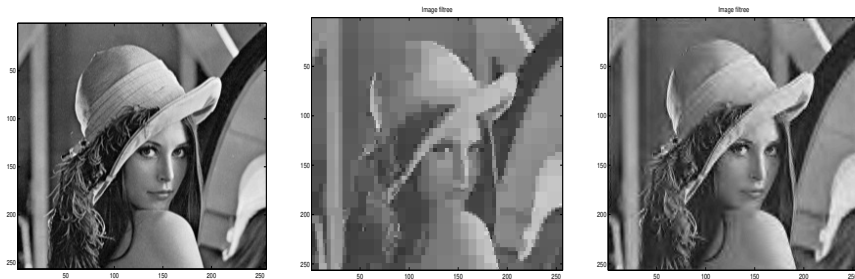
Original image and compressed images:

Middle figure: *keeping the 1024 largest coefficients (compression 98,4%).*

Right Figure: *keeping the 3467 largest coefficients (compression 94,7%).*

Image decomposition

The wavelet bases 2D are constructed by tensor product of 1D bases. Let φ and ψ be the scaling function and wavelet of a MRA. Two constructions are available:

(1) Pure tensor product wavelet bases (anisotropic) :

$$\Psi^{j,j'}_{k,k'}(x,y) = \psi_{j,k}(x)\psi_{j',k'}(y), \;\; j,j' \in \mathbb{Z}, k, k' \in \mathbb{Z}$$

The associated 2D wavelet transform uses 1D FWT on lines followed by 1D FWT on columns of the image.

(2) Wavelets arising from tensor product 2D MRA of $L^2(\mathbb{R}^2)$:

- **Approximation space** (scaling functions) $\mathcal{V}_j = V_j \otimes V_j$
- **Detail spaces** (wavelets) $\mathcal{W}_j$ defined by $\mathcal{V}_{j+1} = \mathcal{V}_j \bigoplus \mathcal{W}_j$

$$
\begin{aligned}
\mathcal{V}_{j+1} &= V_{j+1} \otimes V_{j+1} \\
&= (V_j \oplus W_j) \otimes (V_j \oplus W_j) \\
&= (V_j \otimes V_j) \oplus (W_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j)
\end{aligned}
$$

Separable scaling function and wavelets
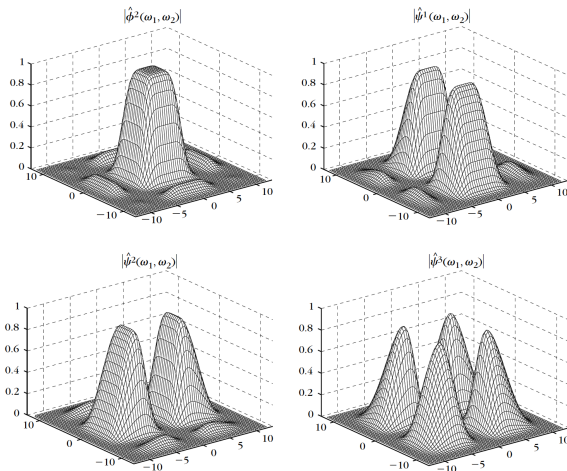
Figure: Fourier transforms of a separable scaling function and of three separable wavelets calculated from a one-dimensional Daubechies 4 wavelet

2D multi-resolutions

$$\Phi(x,y) = \varphi(x)\varphi(y), \quad \Psi^1(x,y) = \psi(x)\varphi(y)$$
$$\Psi^2(x,y) = \varphi(x)\psi(y), \quad \Psi^3(x,y) = \psi(x)\psi(y)$$

$$\mathcal{V}_{j+1} = \underbrace{(V_j \otimes V_j)}_{\text{coarse approx}} \oplus \underbrace{(W_j \otimes V_j)}_{\text{horiz. details}} \oplus \underbrace{(V_j \otimes W_j)}_{\text{vert. details}} \oplus \underbrace{(W_j \otimes W_j)}_{\text{diag. details}}$$

$$\overbrace{\mathcal{V}_j = \text{Span}\{\Phi_{j,\boldsymbol{k}}\}_{\boldsymbol{k}}}^{} \quad \overbrace{\mathcal{W}_j^1 = \text{Span}\{\Psi_{j,\boldsymbol{k}}^1\}_{\boldsymbol{k}}}^{} \quad \overbrace{\mathcal{W}_j^2 = \text{Span}\{\Psi_{j,\boldsymbol{k}}^2\}_{\boldsymbol{k}}}^{} \quad \overbrace{\mathcal{W}_j^3 = \text{Span}\{\Psi_{j,\boldsymbol{k}}^3\}_{\boldsymbol{k}}}^{}$$

Image $C_J[\boldsymbol{k}] = \langle f, \Phi_{J,\boldsymbol{k}} \rangle$, wavelet coefficients: $D_j^i[\boldsymbol{k}] = \langle f, \Psi_{j,\boldsymbol{k}}^i \rangle$

Figure: (Haar approximation) $\mathbf{P}_{\mathcal{V}_j} f = \sum_{\boldsymbol{k}} \langle f, \Phi_{j,\boldsymbol{k}} \rangle \Phi_{j,\boldsymbol{k}} = \sum_{\boldsymbol{k}} C_{j,\boldsymbol{k}} \Phi_{j,\boldsymbol{k}}$
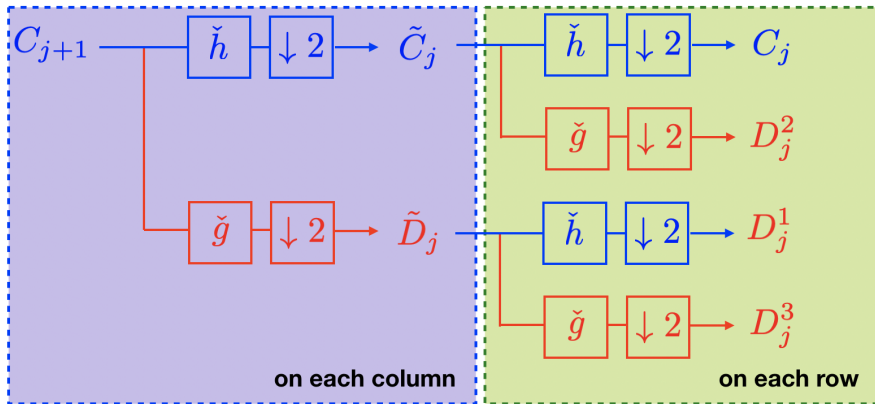
Credits: G. Peyré

Image decomposition

$$\mathcal{W}_j = \text{Vec}\{\psi_{j,k}(x)\varphi_{j,k'}(y)\,;\,\varphi_{j,k}(x)\psi_{j,k'}(y)\,;\,\psi_{j,k}(x)\psi_{j,k'}(y) \mid (k,k') \in \mathbb{Z}^2\}$$
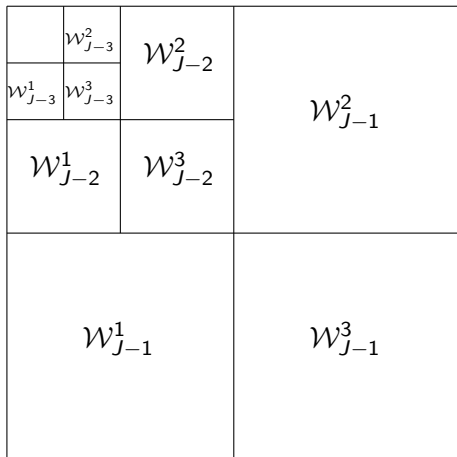
Fast 2D Wavelet Transform

Fast 2D Wavelet Transform

$$C_J$$

$$\widehat{f}(\omega_x, \omega_y)$$

Fast 2D Wavelet Transform
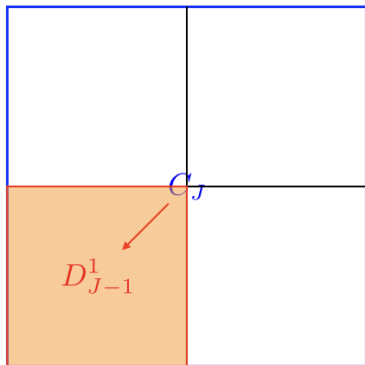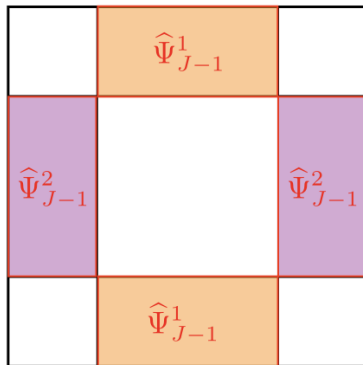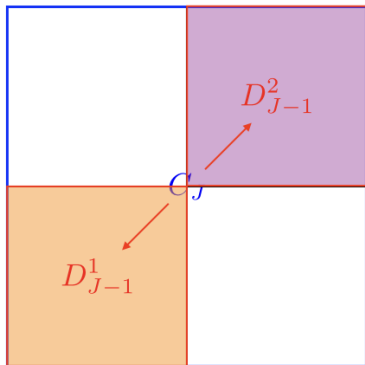
Fast 2D Wavelet Transform
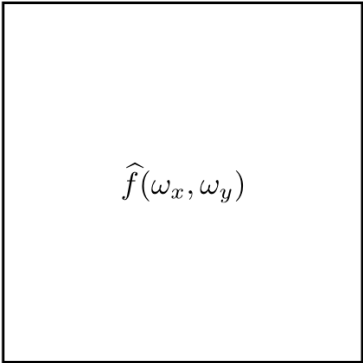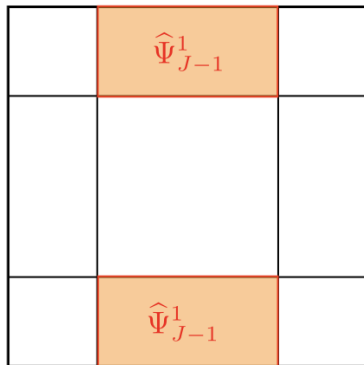
Fast 2D Wavelet Transform
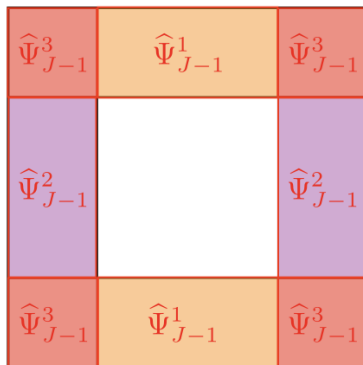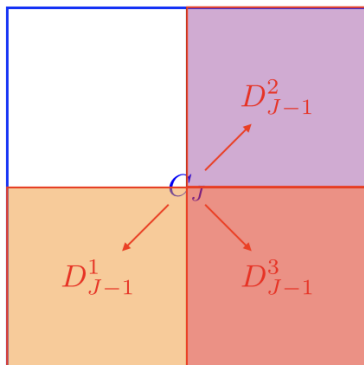
Fast 2D Wavelet Transform
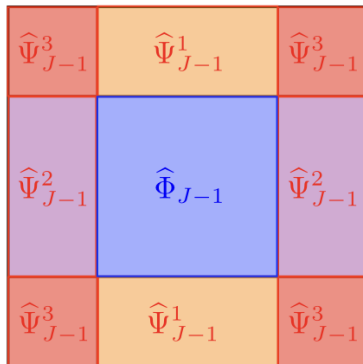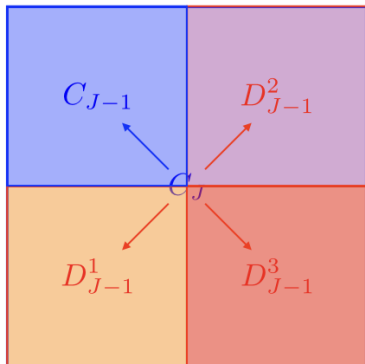
Fast 2D Wavelet Transform
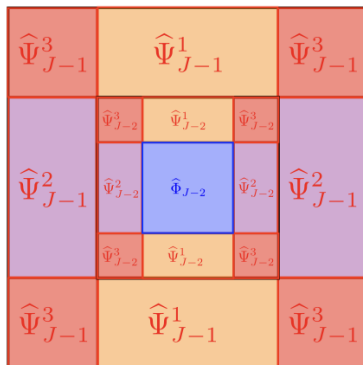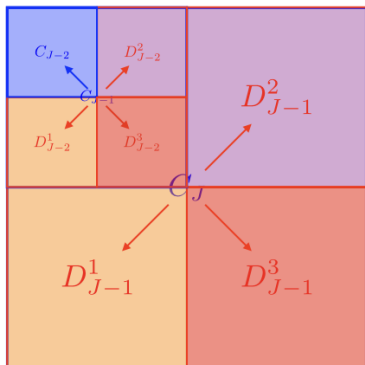
Fast 2D Wavelet Transform
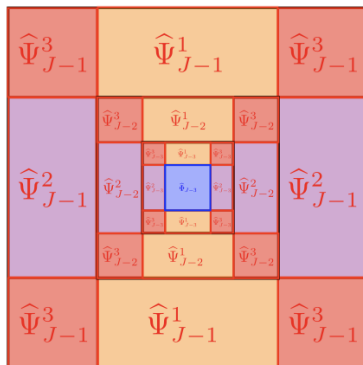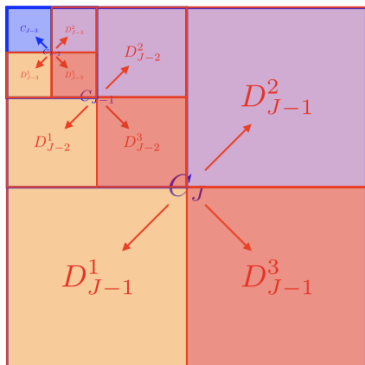
Lena multi-resolution
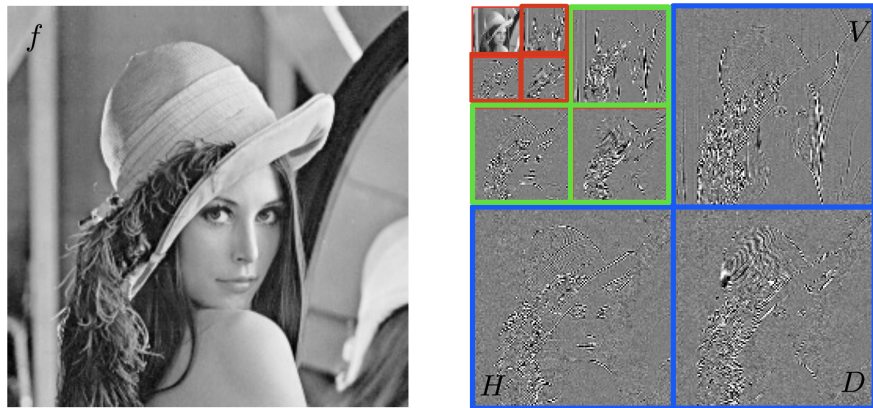
Figure: (Left) original Lena C_O. (Right) coarse approximation C_6 (small Lena on the top left corner in red) and the discrete wavelets coefficients D_j^i for $i \in \{1, 2, 3\}$ and at scales $j = 8$, $j = 7$ and $j = 6$

Credits: G. Peyré

Other examples
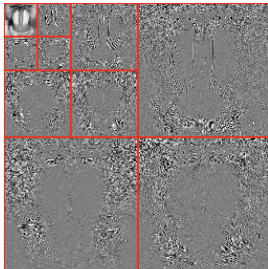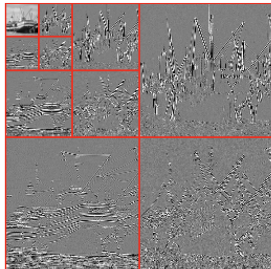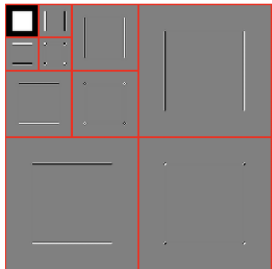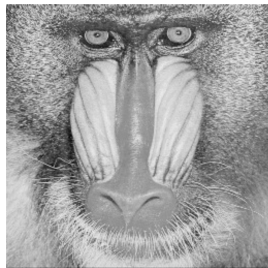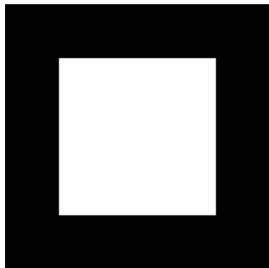
Credits: G. Peyré

Example: Lena compression
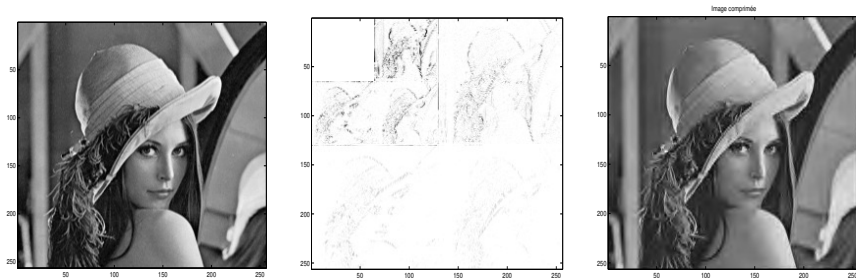
Left figure: *Image* $256 \times 256 = 65536$ *pixels, 256 grey levels.*
Middle figure: *wavelet coefficients of the 2 finest scales*
Right figure: *Reconstructed image from its 4000 largest wavelet coefficients (4 vanishing moments).*
Compression factor $= (65536 - 4000)*100/65536 =$ **93,9 %**

Comparison Haar basis vs. regular wavelets
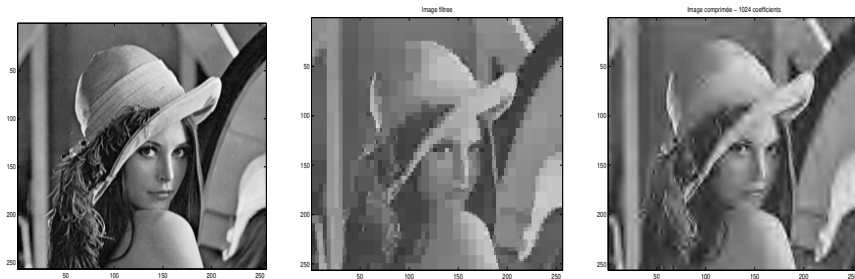
Left figure: *Image* $256 \times 256 = 65536$ *valeurs, 256 niveaux de gris.*
Middle figure: *reconstructed image from its 1024 largest wavelet coefficients in the Haar basis.*
Right figure: *reconstructed image from its 1024 largest wavelet coefficients in a wavelet bases with 4 vanishing moments.*
Compression factor $= (65536 - 1024)*100/65536 = $ **98,44 %**

Fast 2D Inverse Wavelet Transform
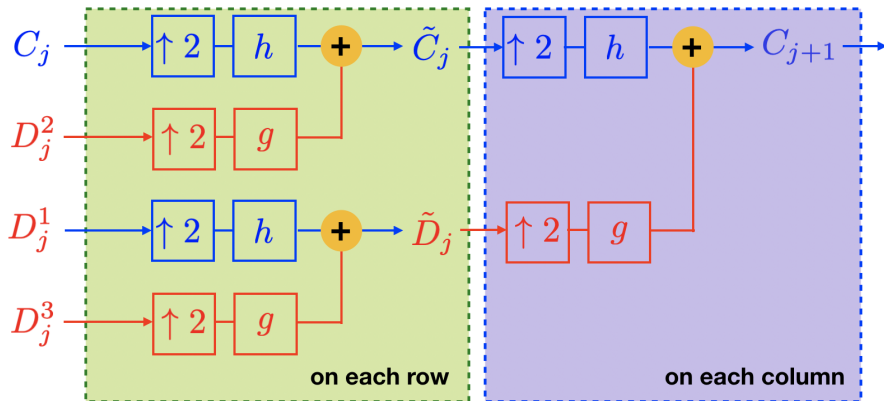
JPEG-2K Compression
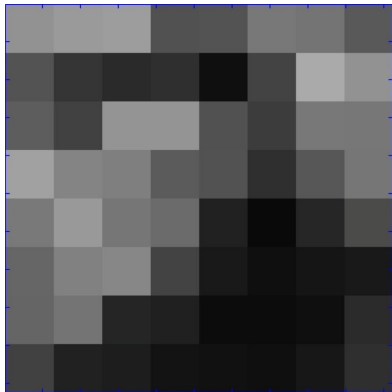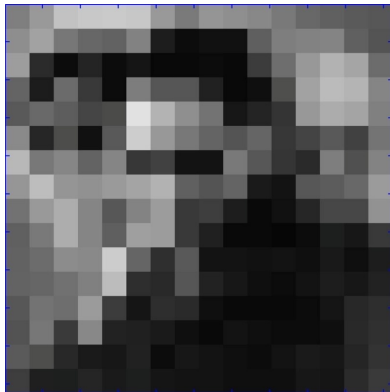
JPEG-2K Compression

JPEG-2K Compression

JPEG-2K Compression

JPEG-2K Compression

JPEG-2K Compression
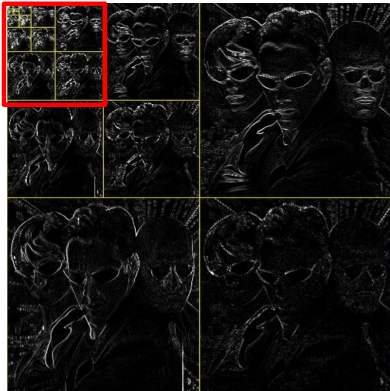
JPEG-2K Compression

JPEG-2K Compression

JPEG-2K Compression
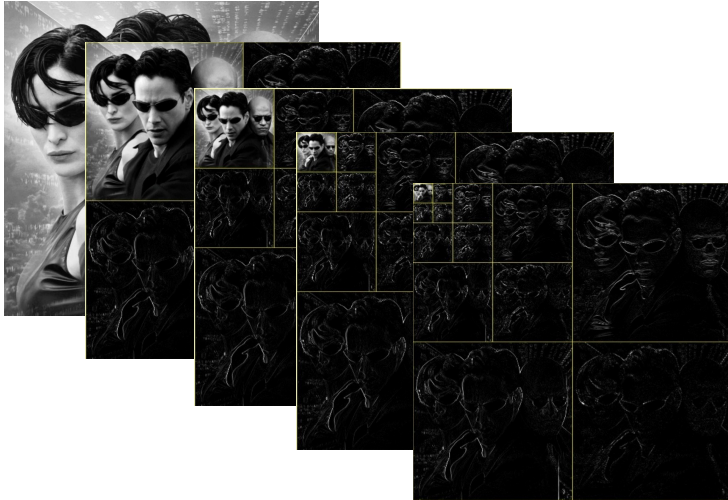
CWT-2D sampled vs. FWT-2D
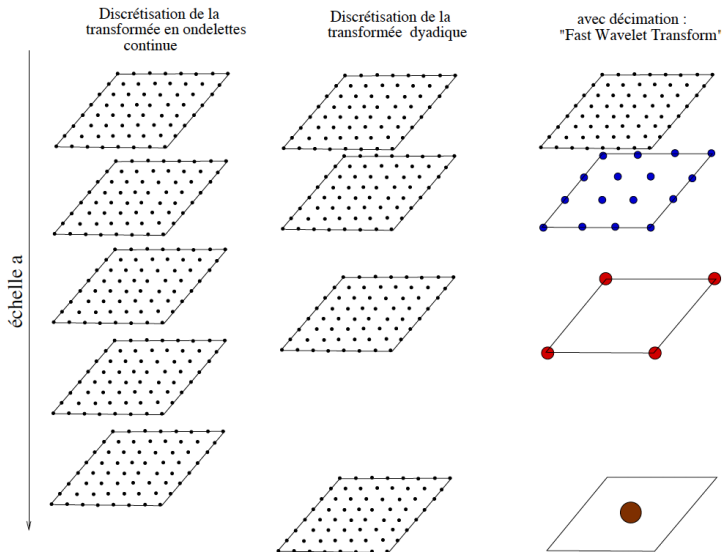
Credits: Le Cadet

Take home message

Question: CWT (discretized) or DWT?

Answer: depends on the application

- **CWT** for feature detection (no *a priori* choice for a, b): more flexible, more robust to noise, but only frames in general.
- **DWT** for large amount of data, data compression: bases, faster, but more rigid (need generalization)

Generalizations

- Biorthogonal wavelets
- Wavelet packets
- Continuous wavelet packets (integrated wavelets)
- Redundant WT (on a rectangular lattice)
- Second generation wavelets (lifting scheme)
- ...

Linear and nonlinear approximations in wavelet bases

Linear approximation

- **Orthogonal projection over the space V_J**

$$
\begin{aligned}
\mathbf{P}_J \ : \ L^2(\mathbb{R}) &\longrightarrow V_J \\
f \ &\longmapsto \sum_{k\in\mathbb{Z}}\langle f,\varphi_{J,k}\rangle\varphi_{J,k} = \sum_{j\leq J-1}\sum_{k\in\mathbb{Z}}\langle f,\psi_{j,k}\rangle\psi_{j,k}
\end{aligned}
$$

- **Strang-Fix condition** of order N ($x^n \in V_0$):

$$
\forall n = 0,\dots,N-1, \quad x^n = \sum_{k\in\mathbb{Z}} a_k^n \ \varphi(x-k)
$$

  i.e ψ has N vanishing moments.

- **Projection error**:

$$
\|f - \mathbf{P}_J f\|_{L^2}^2 = \sum_{j=J}^{+\infty}\sum_{k\in\mathbb{Z}}|\langle f,\psi_{j,k}\rangle|^2
$$

Linear approximation

Theorem

If $f \in H^s(\mathbb{R})$ with $s \leq N$, with

$$\|f - \mathbf{P}_J f\|_{L^2} \leq C \ 2^{-Js} \ \|f\|_{H^s}$$

(*cf.* finite elements with $h = 2^{-J}$) then the following Sobolev-norm equivalence holds:

$$
\begin{aligned}
\|f\|_{H^s}^2 \quad &\sim \quad \sum_{j=-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} 2^{2Js} |\langle f, \psi_{j,k} \rangle|^2 \\
&\sim \quad \|\mathbf{P}_0 f\|_{L^2} + \sum_{j=0}^{+\infty} \sum_{k \in \mathbb{Z}} 2^{2Js} |\langle f, \psi_{j,k} \rangle|^2
\end{aligned}
$$

Nonlinear approximation

Let $N \in \mathbb{N}$. Let $u \in L^2(\mathbb{R})$, and its wavelet decomposition:

$$u = u_0 + \sum_{j=0}^{+\infty} \sum_{k=-\infty}^{+\infty} d_{j,k} \psi_{j,k}$$

One sorts the wavelet coefficients $d_{j,k}$ in decreasing order:

$$|d_{j_1,k_1}| > |d_{j_2,k_2}| > \cdots > |d_{j_{N-1},k_{N-1}}| > \ldots$$

and one introduces the best N-terms non-linear approximation

$$\Sigma_N(u) = u_0 + \sum_{i=1}^{N} d_{j_i,k_i} \psi_{j_i,k_i}$$

If $u \in B_q^{s,q}$ with $\frac{1}{q} = \frac{1}{2} + s$, which is equivalent to:

$$\|u\|_{B_q^{s,q}}^q \sim \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |d_{j,k}|^q < +\infty$$

Nonlinear approximation

The non-linear approximation error is defined as:

$$\|u - \Sigma_N(u)\|_{L^2}^2 = \sum |d_{j_i, k_i}|^2$$

Theorem

$$\|u - \Sigma_N(u)\|_{L^2} \leq C \left(\frac{1}{N}\right)^s \|u\|_{B_q^{s,q}}$$

(in dimension d, s should be replaced by $\frac{s}{d}$).

Proof:

$$n|d_{j_n, k_n}|^q \leq \sum_{i=0}^{n-1} |d_{j_i, k_i}|^q \leq \sum_{i=0}^{+\infty} |d_{j_i, k_i}|^q = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} |d_{j,k}|^q \leq C \|u\|_{B_q^{s,q}}^q$$

Then

$$|d_{j_n, k_n}| \leq C n^{-1/q} \|u\|_{B_q^{s,q}}$$

$$\|u - \Sigma_N(u)\|_{L^2} \leq C\|u\|_{B_q^{s,q}} \Big(\sum_{n \geq N+1} n^{-\frac{2}{q}} \Big)^{1/2} \leq C N^{\frac{1}{2} - \frac{1}{q}} \|u\|_{B_q^{s,q}} = C N^{-s} \|u\|_{B_q^{s,q}}$$

Nonlinear approximation

Remarks

- For $N = 2^J$ one obtains the same convergence rate for the linear and nonlinear approximation. But $B_q^{s,q}$ is a space which contains more functions than the space H^s, for instance discontinuous functions for arbitrary large values of s, whenever functions of H^s are necessarily continuous if $s > d/2$ (d space dimension).

- One has also the characterization: if $u \in B_q^{s,q}$

$$\mathrm{Card}\left\{\lambda : |d_\lambda| \geq \varepsilon\right\} \leq C\varepsilon^{-q}$$

Compression factor of a turbulent 2D vorticity field
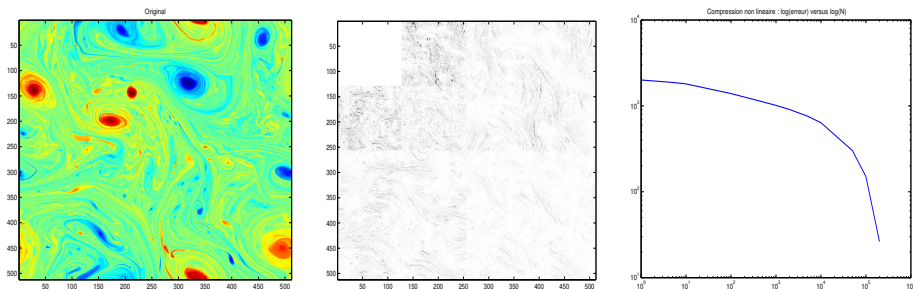
Figure: Analysis of a 2D turbulent field: vorticity field, its wavelet coefficients, and nonlinear approximation error, in terms of the number of retained coefficients

Denoising in orthonornal wavelet bases

References: articles of Donoho and Johnstone

Noised data:

$$X[n] = f[n] + W[n], \quad n = 0, \dots, N-1$$

- X: measured data
- f: (unknown) signal of size N, corrupted by noise
- W: Gaussian *white* noise, with zero mean and variance σ^2

The aim is to provide an estimator $\tilde{F} = \mathcal{D}(X)$ of f minimizing the risk (mean square error):

$$r(\mathcal{D}, f) = \mathbb{E}\left\{\|f - \tilde{F}\|^2\right\} = \sum_{n=0}^{N-1} \mathbb{E}\left\{|f[n] - \tilde{F}[n]|^2\right\}$$

Nonlinear estimators in bases

Let $\mathcal{B} = \{g_k \in \mathbb{R}^N, k = 0, \ldots, N-1\}$ be an orthonormal basis of $\mathbb{R}^N$. One decomposes the noisy signal in $\mathcal{B}$:

$$X[n] = \sum_{k=0}^{N-1} \langle X, g_k \rangle g_k[n]$$

and the inner products satisfy:

$$\langle X, g_k \rangle = \langle f, g_k \rangle + \langle W, g_k \rangle$$

Remarks

- $(\langle W, g_k \rangle)_k$ are independent Gaussian variables of variance σ^2 (since $\mathcal{B}$ is orthonormal).
- $\mathbb{E}\left\{\langle X, g_k \rangle\right\} = |\langle f, g_k \rangle|^2 + \sigma^2$

Diagonal operators

A *diagonal operator* $\mathcal{D}$ in the basis $\mathcal{B}$ leads to an estimator of the form:

$$\tilde{F} = \mathcal{D}X = \sum_{k=0}^{N-1} d_k \left(\langle X, g_k \rangle \right) g_k$$

where the d_k are attenuation functions of the noisy coefficients.

Ideal estimator (*i.e.* which minimizes the risk $r(\mathcal{D}, f)$)

$$\tilde{F} = \mathcal{D}X = \sum_{k=0}^{N-1} \langle X, g_k \rangle \, \theta(k) \, g_k$$

with

$$\theta(k) = \left\{ \begin{array}{lll} 1 & \text{if} & |\langle f, g_k \rangle| \geq \sigma \\ 0 & \text{if} & |\langle f, g_k \rangle| < \sigma \end{array} \right.$$

In this case, the operator $\mathcal{D}$ is nonlinear.

Thresholding estimators

A thresholding estimator in the basis $\mathcal{B}$ corresponds to a *diagonal operator* $\mathcal{D}$:

$$F = \mathcal{D}X = \sum_{k=0}^{N-1} d_k \left(\langle X, g_k \rangle \right) g_k$$

where the d_k are thresholding functions (let T be a threshold):

$$d_k(x) = \rho_T(x) = \left\{ \begin{array}{ll} x & \text{if } |x| > T \quad \text{("hard" thresholding)} \\ 0 & \text{if } |x| \leq T \end{array} \right.$$

or

$$d_k(x) = \rho_T(x) = \left\{ \begin{array}{ll} x - T & \text{if } x \geq T \quad \text{("soft" thresholding)} \\ x + T & \text{if } x \leq -T \\ 0 & \text{if } |x| \leq T \end{array} \right.$$

- **Question**: choice of T to approach the risk of the ideal estimator?
- **Answer**: the choice $T = \sigma \sqrt{2 \log_e N}$ leads to a risk slightly larger (Theorem of Donoho-Jonstone).

Wavelet thresholding

Consider a (periodic) wavelet basis:

$$\mathcal{B} = \{\varphi, \psi_{j,k} \; ; \; 0 \leq j \leq J-1, \; k = 0 : 2^j - 1\} \; (N = 2^J = \text{size of the data})$$

The thresholding estimator writes:

$$\tilde{F} = \rho_T(\langle X, \varphi \rangle) \, \varphi + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \rho_T \left(\langle X, \psi_{j,k} \rangle \right) \psi_{j,k}$$

Estimation of the noise variance σ^2:

If f is piecewise regular, a robust estimator is given by the *median* of the wavelet coefficients at the finest scale:

- $\{\langle X, \psi_{j,k} \rangle\}_{k=0:2^{J-1}-1} : 2^{J-1} = \frac{N}{2}$ wavelet coefficients of the noisy data at the finest scale.
- If $\langle f, \psi_{j,k} \rangle$ is small (f is regular on the support of $\psi_{J-1,k}$), one has: $\langle X, \psi_{j,k} \rangle \approx \langle W, \psi_{j,k} \rangle$.

Wavelet thresholding

- If $\langle f, \psi_{j,k} \rangle$ is large, it corresponds to a singularity of f, but for a piecewise regular functions with isolated singularity, only few coefficients $\langle X, \psi_{j,k} \rangle$ are affected at the finest scale.
- Then $\langle X, \psi_{j,k} \rangle$ is a random variable of variance σ^2.

The noise standard deviation σ is estimated by the formula (exact for $P = 2^{J-1}$ independent Gaussian variables, of zero mean, and variance σ^2):

$$\sigma \approx \frac{M_X}{0,6745}$$

where M_X is the median of the coefficients $\{\langle X, \psi_{j,k} \rangle\}_{k=0:2^{J-1}-1}$ at the smallest scale.

Example:

$f(x) = \sqrt{|\cos 2\pi x|} + \text{noise}$ (discretized on $1024 = 2^{10}$ values)

Example: $f(x) = \sqrt{|\cos 2\pi x|} + \text{noise}$
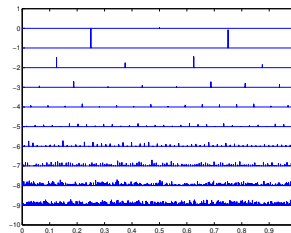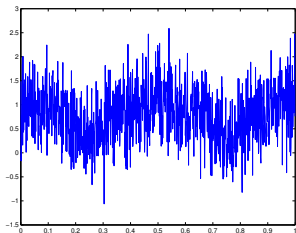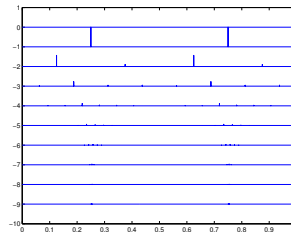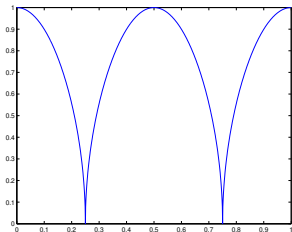
Example: $f(x) = \sqrt{|\cos 2\pi x|} + \text{noise}$
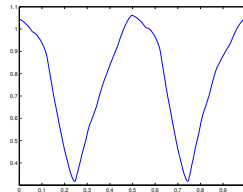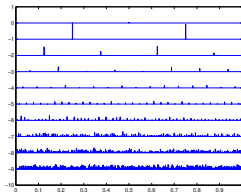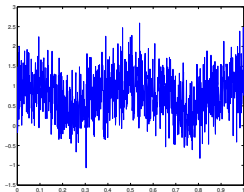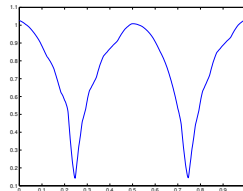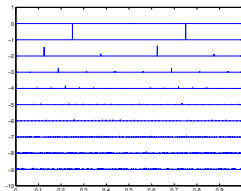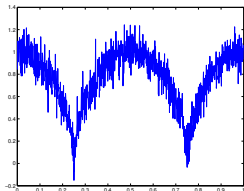
Example: Piece-Regular
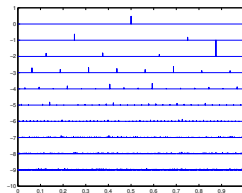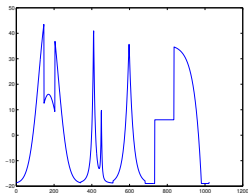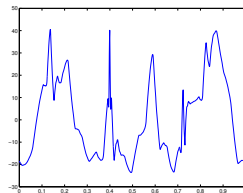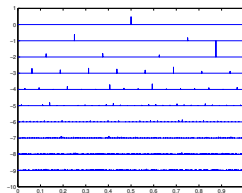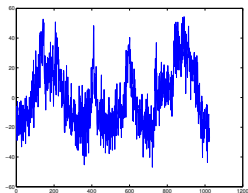
Example: WaveLab denoising function

```matlab
%Generation of a signal y
n=1024; dx=1/n; x=(0:n-1)/n;
alpha=0.1 % noise coefficient
y=sqrt(abs(cos(2*pi*x)));
or
y=MakeSignal('Piece-Regular',n);
%
y=y+alpha*randn(size(y)); % add Gaussian noise
plot(x,y) % plot of the noisy signal
% Denoising by hardtresholding on orthonormal wavelet coeff 'Symmlet 4'
out=ThreshWave(y);
plot(x,out) % plot the denoised signal
```

Sparse representation and approximation

Analysis vs. synthesis

- **Analysis**: $\Phi(f) = \{\langle f, \phi_p \rangle\}_{p \in \Gamma}$
- **Synthesis**: $f = \sum_p \langle f, \phi_p \rangle \phi_p$

Suppose that a sparse family of vectors $\{\phi_p\}_{p \in \Lambda}$ has been selected to approximate a signal f. An approximation can be recovered as an orthogonal projection in the space $\mathbf{V}_\Lambda$ generated by these vectors.

1. In a *dual-synthesis* problem, the orthogonal projection f_Λ of f in $\mathbf{V}_\Lambda$ is computed as above from the inner products $\{\langle f, \phi_p \rangle\}$ provided by an analysis operator, whose only a subset of such inner products is selected and possibly thresholded.

2. In a *dual-analysis* problem, the decomposition coefficients of f_Λ must be computed on a family of selected vectors $\{\phi_p\}_{p \in \Lambda}$, by pursuit algorithms which compute approximation supports in highly redundant dictionaries.

Sparse representation and approximation

Approximation in bases vs. redundant dictionaries

1. Choose an **orthogonal basis** $\mathcal{B} = \{\phi_p\}_{p \in \Gamma}$ for which the representation is **not redundant** at all, so we get a representation which is sparse and stable
   - By selecting the first M coefficients *(linear approximation)*
   - By selecting the M largest coefficients *(non-linear approximation)*

   The size support $M = |\Lambda|$ of $f_M \equiv f_\Lambda$ needed to have a good approximation error $\|f - f_M\|$ depends on the regularity of f.

2. Choose a **dictionary** $\mathcal{D} = \{\phi_p\}_{p \in \Gamma}$ which is **highly redundant** in order to obtain a more sparse representation (e.g. natural languages use redundant dictionaries). Identifying *patterns* or *features* consist on finding which vectors (*atoms*, words, ...) to choose to approximate

$$f \approx f_\Lambda = \sum_{p \in \Lambda} \alpha_p \phi_p$$

Famous algorithms: Matching pursuit, Orthogonal Matching Pursuit (OMP), Basis pursuit, ...

Sparse representation and approximation
Moving from transforms to dictionaries

- "Xlets" (curvelets, bandlets, contourlet, ...) take advantage of the image geometric regularity
- Redundant dictionaries can improve approximation, compression and denoising
- Optimal approximation finding is NP-hard, only approximated with matching or basis pursuits
- Great impact to inverse problems
  - Compressed sensing
  - Super-resolution
  - Source separation
- Can be used for patter recognition but problems of instabilities
- Deep learning made a breakthrough in classification and pattern recognition (dictionaries are learned i.e linear operators/filters, but need a lot of examples). Increase the level of adaptability.

$\Rightarrow$ have wavelets become *has-been*?

The Scattering Transform

Understanding deep convolutional networks
Supervised learning against high dimension

- Data in high dimension $x \in \mathbb{R}^d$ with $d \approx 10^6$
- $f(x)$ represents a label of a class (whose can be also big, e.g $2 \cdot 10^3$ for ImageNet) for **classification** tasks, or a real for **regression**.
- Training set of n samples $\{x_i, y_i = f(x_i)\}_{i \leq n}$ (few samples per class)
- Supervised learning aims at generalizing from the samples to predict $f(x)$ for new datas.

Intuitively, to do an **interpolation** in x we need somehow to average among known samples $\{x_i, y_i\}$ in the neighborhood of x, saying:

$$\forall x \in [0,1]^d, \exists x_i \in [0,1]^d, \quad \|x - x_i\| \leqslant \epsilon$$

then if the x_i's are uniformly distributed, it would require ϵ^{-d} points to cover $[0,1]^d$ entirely!

Points are far away in high dimension $\Rightarrow$ Curse of dimensionality

Understanding deep convolutional networks

Kernel learning

1. **Representation**. Change of variable $\Phi(x) = \{\phi_k(x)\}_{k \leqslant d'}$ (*features*) in order to nearly linearize class bounderies:

$$x = (v_1, \ldots, v_d) \xrightarrow{\Phi} \Phi(x) = (v_1', \ldots, v_d')$$

2. **Classifier**. Find an hyperplan (that is an vector w orthogonal to the hyperplan) which seperates the transformed data:

$$\tilde{f}(x) = \text{sign}(\langle \Phi(x), w \rangle + b) = \text{sign}\left(\sum_k w_k v_k' + b\right)$$

Questions:

- How to construct such a representation Φ?
- What regularity is needed?
- Can wavelets be useful to understand and draw CNN architectures?

Understanding deep convolutional networks

CNN architecture

J. Hinton, Y. LeCun

Hierarchical invariance

Le, Ranzato, Ng et. al.:
unsupervised training
1 billion variables

$\Phi(x)$

x
Input

C1 feature maps

S1 feature maps

C2 feature maps

S2 feature maps

Output

SVM/RBF

y

Face detector

Body detector

Full Connection

filtering

Subsampling

filtering

Subsampling

filtering

Rectification

Contrast Normalization

Pooling / Subsampling

Wavelet filters
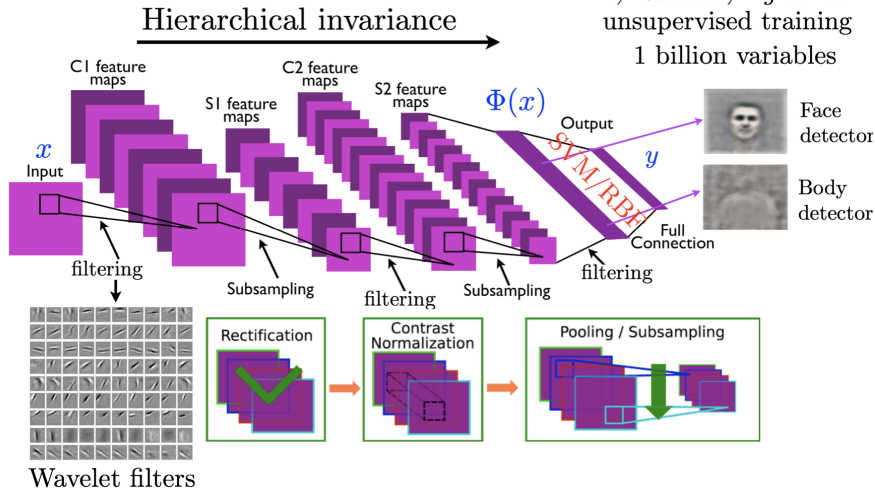
Credits: S. Mallat

Understanding deep convolutional networks

CNN architecture: why are they so efficient for images classification?

- Why convolutions? Which filters?
- Why pooling? Why multi-stage and how deep?
- Why and which non-linearities?
- Why normalization?
- What is the role of sparsity?

⇒ what are the mathematical operators behind such architectures?
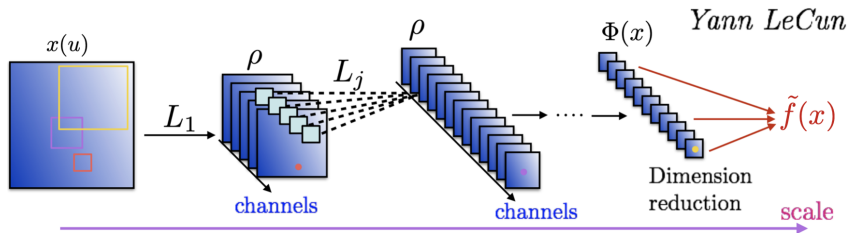
Figure: L_j: sum of spatial convolutions across channels, subsampling. ρ: scalar non-linearity ($\max(u,0)$, $|u|$, ...)

Credits: S. Mallat

Understanding deep convolutional networks

The "3S" ingredients for reducing the dimensionality problem

1. **Separability**: variables separation can reduce the dimensionality from d to K problems of dimension $q \ll d$ (e.g decomposing an image $10^3 \times 10^3$ in small independant patches 8×8, whose interactions between pixels are essentially local $\Rightarrow$ SIFT). It is important to make **scales separation** but also to capture their interaction: deeper neurons can "see" greater portion of the image.

2. **Symmetry**: spatial symmetries produce **translation/rotation/flip invariance** (e.g convolution filters induce translation invariance) and reduce the dimensionality by eliminating some variables.

3. **Sparsity**: pattern recognition consists on decomposing the problem on sparse **elementary structures** in dictionaries (cat's hears, human's eyes, ...) in particular through the activation functions.

$\Rightarrow$ take advantage both of *a priori* information hard-coded in the network architecture and learning to design Φ.

Symmetry group

To know the regularity of f one can study it through local but also global transformation such that symmetry group of f:

$$G = \{g : \forall x \in \Omega, \quad f(g.x) = f(x)\}$$

- The functions g preserve the level sets $\Omega_t = \{x : f(x) = t\}$, that is if $x \in \Omega_t$ and $g \in G$ then $g.x \in \Omega_t$. So it is easy to verify the solutions of a level set has a structure of group.

- **Information *a priori***, a symmetry subgroup $H \subset G$. If $g \in H$ then x and $g.x$ have the same label $f(g.x) = f(x)$, so belong to the same class of equivalence. The quotient of Ω by H is denoted by $\Omega \backslash H$, for $x_0 \in \Omega \backslash H$ then it defines a class of equivalence:

$$H_{x_0} = \{x \in \Omega : g \in H \text{ s.t } g.x = x_0\}$$

Example: if x_0 is an image and $f(x_0)$ its label (cat/dog), then by translating $x = g.x_0 \in H_{x_0}$ the label remains the same $f(x) = f(x_0)$.

- One can then reduce the number of variables (variability) within the class of equivalence (reduction of dimensionality).

Symmetry group

Lie group: infinitely small generators

Reduction of dimensionality in the continuous case:

$$\dim(\Omega \backslash H) = \dim(\Omega) - \dim(H)$$

Diffeomorphisms group

Let $g : [0,1]^2 \to [0,1]^2$ be a $\mathcal{C}^1$ function acting on the underlying variable of x, namely u which is a low-dimensionnal quantity:

$$g.(x(u)) = x(g(u))$$

Examples

- Translation: $g.x(u) = x(u - g)$ with $g \in \mathbb{R}^2$
- Rotation: $g.x(u) = x(\mathbf{R}_g u)$ with $g \in [0, 2\pi]$
- Globally invariant to the translation group $\Rightarrow$ small
- Locally invariant to small diffeomorphisms $\Rightarrow$ **HUGE**

Continuous transports by successive action of generators $f(x_i) = f(x_0)$

$$O_x = \{g.x\}_{g \in G} \quad (\text{orbit} = \text{differentiable surface of iso-label})$$

Understanding deep convolutional networks

Using the information *a priori* on the symmetry group of f to define the representation Φ for the final classification/regression (last layer):

$$\tilde{f}(x) = \langle \Phi(x), w \rangle = \sum_k w_k \phi_k$$

In order that $\tilde{f}$ is a good approximation of f, we impose that it has the **same invariants** $g \in G$ that is G is a symmetry group of Φ.

Two possibilities:

1. G **known and low dimension** (translation, rotation, ...)
   $\Rightarrow$ constructing directly Φ
2. G **unknown and high dimension** (diffeomorphisms)
   $\Rightarrow$ linearization + learning invariant through the classifier.

$$\tilde{f}(x) = \tilde{f}(g.x) \Rightarrow \langle \Phi(x), w \rangle = \langle \Phi(g.x), w \rangle \Rightarrow \langle \Phi(x) - \Phi(g.x), w \rangle = 0$$

$$\Phi(x) - \Phi(g.x) \in V \perp w$$

$\rightsquigarrow$ If V is a hyperplan it implies to linearize transformations, by considering small deformations g.

Linearization of small deformations

- **Linearize group actions**: $g.x = x + \tau.x$ so locally the tangent hyperplan to the orbit O_x is given by τ (Lie algebra).

- **For small deformations** $g.x(u) = x(u - \tau(u))$ we can write the action τ as a "global" action (the translation) and a small "local" action (the deformation), since $\tau(u) \approx \tau(u_0) + \nabla\tau(u_0)(u - u_0)$ then

$$x(u - \tau(u)) = x(\underbrace{(\mathbb{I} - \nabla\tau(u_0))(u - u_0)}_{\text{local deformation}} + \underbrace{u_0 - \tau(u_0)}_{\text{global translation}})$$

- **Distance** for small deformations: $|g|_G = \|\tau\|_\infty + \|\nabla\tau\|_\infty$

- We do not know in advance what is the local range of diffeomorphism symmetries.
  Example: to classify images x of handwritten digits, certain deformations of x will preserve a digit class but modify the class of another digit.

Linearization of small deformations

- We shall linearize small diffeomorphims g via the change of variable $\Phi(x)$, which is say Lipschitz-continuous if

$$\exists C > 0, \forall(x, g) \in \Omega \times G, \quad \|\Phi(g.x) - \Phi(x)\| \leqslant C |g|_G \|x\|$$

- The Radon–Nikodim property proves that the map that transforms g into $\Phi(g.x)$ is almost everywhere differentiable in the sense of Gâteaux. If $|g|_G$ is small, then $\Phi(g.x) - \Phi(x)$ is closely approximated by a bounded linear operator of g, which is the Gâteaux derivative. **Locally, it thus nearly remains in a linear space**.

$\Rightarrow$ The Lipschitz property of Φ is difficult to be obtained. Indeed, a local deformation is a dilation, so **the representation will have to be based on dilations**, that is we will need to separate scales with the wavelet transform.

Stable invariants

If $\Phi(x) = \{|\hat{x}(\omega)|\}_\omega$ then:

- **Invariance to translations** $x_c(t) = x(t - c)$

$$\forall c \in \mathbb{R}, \quad \Phi(x_c) = \Phi(x)$$

- Not Lipschitz stable to small deformation $x_\tau(t) = x(t - \tau(t))$
  where $\tau(t) = \epsilon t$ for example. The Fourier transform of
  $x(t - \tau(t)) = x((1 - \epsilon)t)$ is $\hat{x}(\omega(1 + \epsilon))$, so two "bumps" centered
  in $\omega = \pm\omega_0$ will be "shifted" toward low frequencies by a quantity
  $\epsilon\omega_0$, such that they are not superposed anymore and then

$$\|\Phi(x_\tau) - \Phi(x)\| \neq \epsilon$$

.

$\Rightarrow$ Wavelets are localized waveforms and are thus stable to
deformations, as opposed to Fourier sinusoidal waves

Stable invariants

- Wavelets are uniformly stable to deformations:
  If $\psi_{\lambda,\tau}(t) = \psi_\lambda(t - \tau(t))$ then

$$\|\psi_\lambda - \psi_{\lambda,\tau}\| \leq C \sup_t |\nabla \tau(t)|$$

- Wavelet separate multiscale information
- Wavelets provide sparse representation

Multiscale Wavelet Transform

- Complex wavelet $\psi(u) = \psi^a(u) + i\psi^b(u)$
- Dilated 1D wavelet: $\psi_\lambda(u) = 2^{-j/Q}\psi(2^{-j/Q}u)$ with $\lambda = 2^{-j/Q}$
- For images with two variables $u = (u_1, u_2)$ add a rotation $r \in G$ of angles $2k\pi/K$ for $0 \le k < K$:

$$\psi_\lambda(u) = 2^{-2j}\psi(2^{-j}r^{-1}u), \quad \lambda = (2^{-j}, r)$$

- Wavelet transform:

$$Wx = \left(\begin{array}{c} x \star \phi(u) \\ x \star \psi_\lambda(u) \end{array} \right)_{u,\lambda}$$

- If $|\widehat{\phi}(\omega)|^2 + \sum_\lambda |\widehat{\psi}_\lambda(\omega)|^2 = 1$ then W is unitary: $\|Wx\|^2 = \|x\|^2$

Stable translation invariance

- $x \star \psi_\lambda$ is translation covariant, not invariant and

$$\int x \star \psi_\lambda(u) \, \mathrm{d}u = 0$$

- Translation invariant representation: $\int M(x \star \psi_\lambda)(u) \, \mathrm{d}u$
- Diffeomorphism stability: M commutes with diffeomorphims
- L^2 stability: $\|Mh\| = \|h\|$ and $\|Mg - Mh\| \leq \|g - h\|$

$$\Rightarrow M(h)(u) = |h(u)| = \sqrt{|h^a(u)|^2 + |h^b(u)|^2}$$

Wavelet translation invariance

- The modulus $|x \star \psi_{\lambda_1}| = \sqrt{|x \star \psi_{\lambda_1}^a|^2 + |x \star \psi_{\lambda_1}^b|^2}$ (*pooling*) is a regular envelop
- The average $|x \star \psi_{\lambda_1}| \star \phi(t)$ is invariant to small translations relatively to the support of ϕ
- Full translation invariance at the limit:

$$\lim_{\phi \to 1} |x \star \psi_{\lambda_1}| = \int |x \star \psi_{\lambda_1}(u)| \mathrm{d}u = \|x \star \psi_{\lambda_1}\|_1$$

- First Wavelet transform modulus:

$$\rho W_1 = |W_1|x = \left(\begin{array}{c} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \end{array} \right)_{\lambda_1}$$
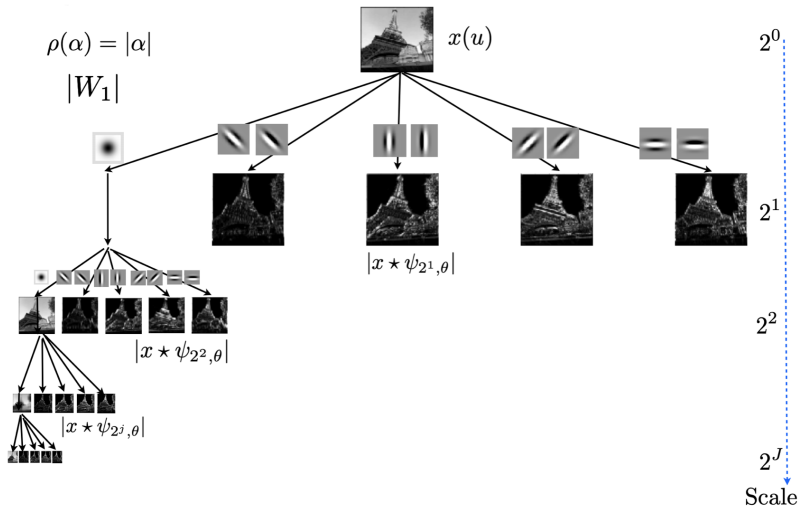
- Second Wavelet transform modulus (for recovering high freq. lost):

$$|W_2||x \star \psi_{\lambda_1}| = \left(\begin{array}{c} |x \star \psi_{\lambda_1}| \star \phi_{2^J} \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \end{array} \right)_{\lambda_2}$$
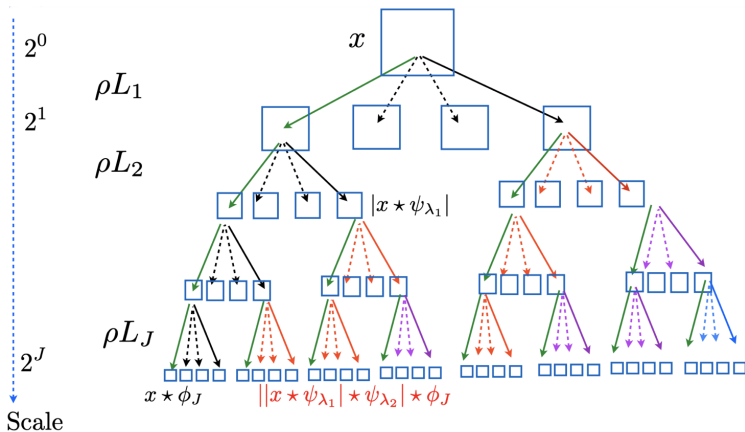
- Translation invariance by averaging $||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J}, \quad \forall \lambda_1, \lambda_2$

Scattering Network

$\rho(\alpha) = |\alpha|$

$|W_1|$

$x(u)$

2^0

2^1

$|x \star \psi_{2^1,\theta}|$

2^2

$|x \star \psi_{2^2,\theta}|$

$|x \star \psi_{2^j,\theta}|$

2^J

Scale

Credits: S. Mallat

Scattering Network

$$S_J = \rho\, W_1 \;\; \rho\, W_2 \; \cdots \; \rho\, W_J$$

$$\rho(\alpha) = |\alpha| \qquad S_J x = \left\{ |||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2} \star \ldots | \star \psi_{\lambda_m} | \star \phi_J \right\}_{\lambda_k}$$

Interactions across scales

Credits: S. Mallat

Scattering Properties

$$S_J x = \begin{pmatrix} x \star \phi_{2^J} \\ |x \star \psi_{\lambda_1}| \star \phi_{2^J} \\ ||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \phi_{2^J} \\ |||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \psi_{\lambda_3}| \star \phi_{2^J} \\ \vdots \end{pmatrix}_{\lambda_1, \lambda_2, \lambda_1, \dots} = \cdots |W_3||W_2||W_1|x$$

Lemma: $\|W_k D_\tau - D_\tau W - k\| \leq C\|\nabla \tau\|_\infty$ where $D_\tau x(u) = x(u - \tau(u))$

Theorem (Mallat et al.)

For appropriate wavelets, a scattering is contractive

$$\|S_J x - S_J y\| \leq \|x - y\|,$$

translations invariance and deformation stability:

$$\lim_{J \to +\infty} \|S_J D_\tau x - S_J x\| \leq C\|\nabla \tau\|_\infty \|x\|$$
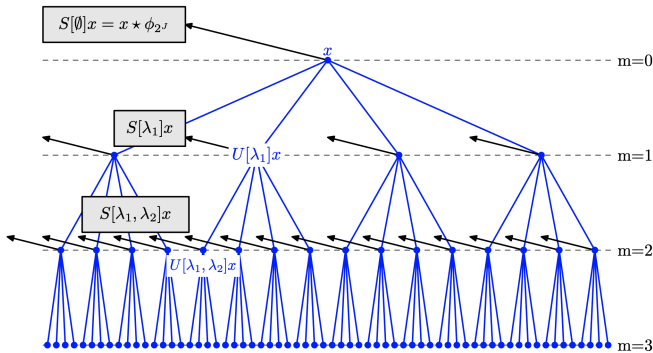
Scattering Network

Fig. 2. A scattering propagator $\widetilde{W}$ applied to x computes the first layer of wavelet coefficients modulus $U[\lambda_1]x = |x \star \psi_{\lambda_1}|$ and outputs its local average $S[\emptyset]x = x \star \phi_{2^J}$ (black arrow). Applying $\widetilde{W}$ to the first layer signals $U[\lambda_1]x$ outputs first order scattering coefficients $S[\lambda_1] = U[\lambda_1] \star \phi_{2^J}$ (black arrows) and computes the propagated signal $U[\lambda_1, \lambda_2]x$ of the second layer. Applying $\widetilde{W}$ to each propagated signal $U[p]x$ outputs $S[p]x = U[p]x \star \phi_{2^J}$ (black arrows) and computes a next layer of propagated signals.
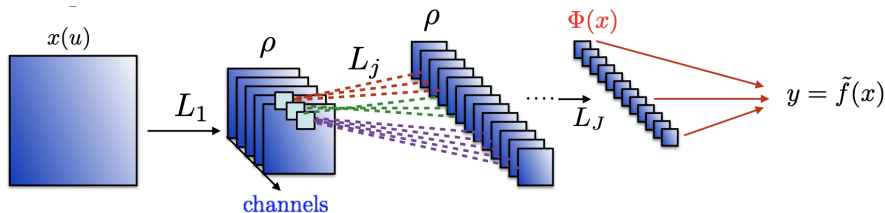
Credits: S. Mallat

Understanding deep convolutional networks

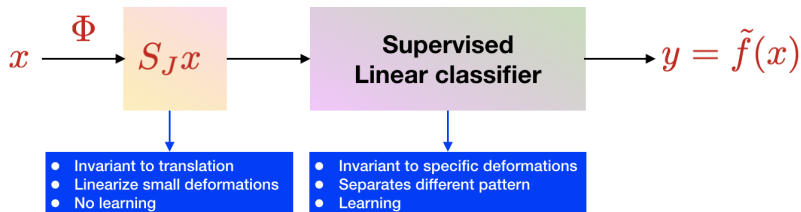Simplified architecture: Deep Convolutional Trees

Architecture

- Convolutional filters L_j: band-limited wavelets
- Pooling: L^1 norm as averaging
- Nonlinear activation ρ: modulus

$$\Phi(x) = S_J x \text{ (scattering vector)}$$


Credits: S. Mallat

Experiments and results

- MNIST dataset for **digit classification**: for a training of 50,000 digits the classification error of the Scattering Network was similar to the Convolutional Network's (0.4 %)

Experiments and results

- CUReT dataset for **textures classification**: for a small training set of textures 200×200 in 61 classes (46 per class), the classification error with the Scattering Network achieves 0.2 %, far better than Fourier transform's one (1 %)
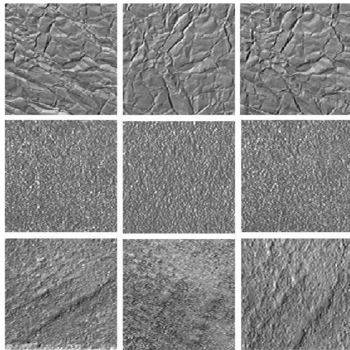
Experiments and results
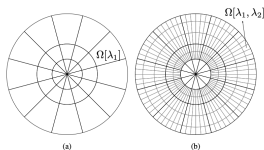
Scattering coefficients

Fig. 3. To display scattering coefficients, the disk covering the image frequency support is partitioned into sectors $\Omega[p]$, which depend upon the path p. (a): For $m = 1$, each $\Omega[\lambda_1]$ is a sector rotated by r_1 which approximates the frequency support of $\hat{\psi}_{\lambda_1}$. (b): For $m = 2$, all $\Omega[\lambda_1, \lambda_2]$ are obtained by subdividing each $\Omega[\lambda_1]$.
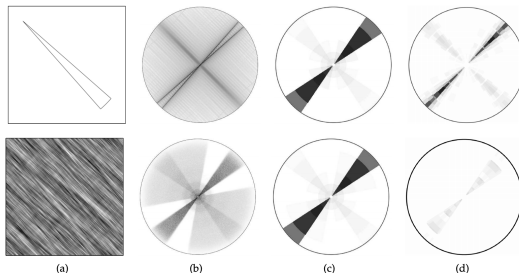

Fig. 4. (a) Two images $x(u)$. (b) Fourier modulus $|\hat{x}(\omega)|$. (c) First order scattering coefficients $Sx[\lambda_1]$ displayed over the frequency sectors of Figure 3(a). They are the same for both images. (d) Second order scattering coefficients $Sx[\lambda_1, \lambda_2]$ over the frequency sectors of Figure 3(b). They are different for each image.

Experiments and results

Scattering coefficients
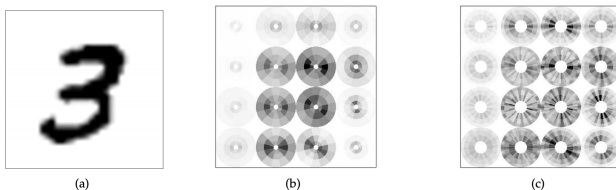

Fig. 7. (a): Image $X(u)$ of a digit '3'. (b): Arrays of windowed scattering coefficients $S[p]X(u)$ of order $m = 1$, with u sampled at intervals of $2^J = 8$ pixels. (c): Windowed scattering coefficients $S[p]X(u)$ of order $m = 2$.
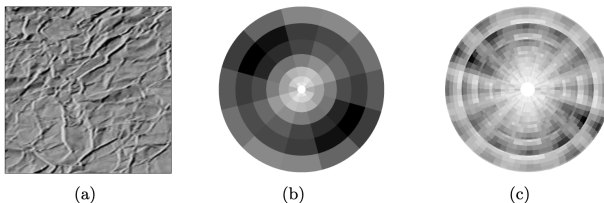

Figure 4.3: (a): Example of CureT texture $X(u)$. (b): Scattering coefficients $S_J[p]X$, for $m = 1$ and 2^J equal to the image width. (c): Scattering coefficients $S_J[p]X(u)$, for $m = 2$.
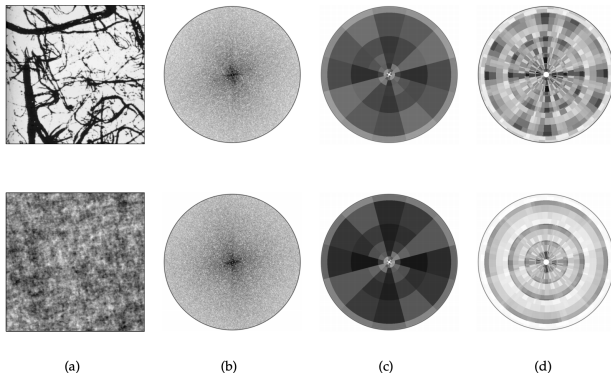
Experiments and results

Scattering coefficients

Fig. 5. (a) Realizations of two stationary processes $X(u)$. Top: Brodatz texture. Bottom: Gaussian process. (b) The power spectrum estimated from each realization is nearly the same. (c) First order scattering coefficients $S[p]X$ are nearly the same, for 2^J equal to the image width. (d) Second order scattering coefficients $S[p]X$ are clearly different.

Take home message

Interpretation of convolutional networks

- Deep convolutional network are really efficients to approximate functions in very high dimension
- Compute multiscale **invariants** of complex symmetries and learn sparse patterns
- Many mathematical questions still open (notion of regularity, complexity, approximation theorems, ...)

References

- J. Bruna & S. Mallat, Invariant scattering convolution networks. IEEE transactions on pattern analysis and machine intelligence (2013)
- S. Mallat, Understanding deep convolutional networks. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences (2016)