

Chapter 5

Spectral and Time–Frequency Analyses and Signal Processing

5.1 Fourier Theory and Wavelets

5.1.1 Contribution of the Fourier Analysis to Regular and Stationary Series: An Approach of Linearities

Even if the Fourier analysis dates back before the Fourier's work and even if the different Fourier analysis developments have been done after him, Fourier is an icon whose influence is fundamental still today. In 1822, Fourier¹ in his work entitled “Analytical Theory of Heat”, explained the way in which the *linear equations of partial derivatives could describe the propagation of Heat in a simple form*. In brief, it stated that any periodic function can be expressed as a sum of sinusoids, i.e. sines and cosine of different frequencies: This is the Fourier series. Then, by extension it is said that any periodic curve, even if it is discontinuous, can be *decomposed into a sum of smooth curves*. Consequently, an irregular or jagged curve and the sum of sinusoids are representations of the same thing, but one of them has an empirical nature and the other is the result of an algebraic decomposition. The decomposition method uses the amplitude of sinusoids by assigning to them coefficients and uses the phases. It is important to underline that we can reconstruct the function from the Fourier series (or Fourier transform) without loss of information.

The Fourier transform is an operation that consists in the decomposition of a function according to its frequencies, i.e. we transform a function that depends on time into a function that depends on the frequency. This new function, which depends on frequencies, shows how much sine and cosine of each frequency are contained in the original function. The new function obtained is called the Fourier series.² For stock market fluctuations for example, or signals which vary in the

¹ Although the first writings date back to 1807.

² *Fourier series for a periodic function:* The Fourier series of a periodic function contain only the sinusoids of frequencies equal to multiple integers of the fundamental frequency. When function is not periodic, but when it decreases rather quickly ad infinitum so that the area located under its graph is finite, it is possible to describe it by a superimposition of sine and cosine; but it is however

course of time, the frequency is generally expressed in cycles per second or in hertz. A function and its Fourier transform are thus the two faces of the same information. The first one favors the information about time and masks the information about frequencies, while the second highlights the information about frequencies and hides the information about time. *One of the fundamental stakes of this method is to be able to consider a curve of empirical origin as a function, even if it is irregular.*

The Fourier method was applied in many technical and scientific fields such as geophysics, astronomy, acoustics, medical research and more generally signal analysis.³ The analyzed signals are often of irregular nature, and the Fourier analysis of these signals translates these curves into an analytical form. This analytical form transforms a chronicle which varies in the course of the time into a function which is the Fourier transform, and more exactly the Fourier series giving the quantity of sine and cosine of each frequency contained in the signal.

At this stage, it is important to underline that the different sinusoids composing a signal, highlighted by the Fourier analysis, can represent the *veritable physical waves* which compose the signal. Indeed, they can be for example an acoustic signal, radio or electromagnetic waves, which are individually identifiable by their respective frequencies. The Fourier method has demonstrated its capacity to analyze, decompose and decode the natural phenomena. In Physics the *space of Fourier transforms* (i.e. “Fourier space”) makes it possible to *study the characteristics of an elementary particle* and more exactly *its quantity of motion* (i.e. momentum⁴), which makes it possible to *assimilate a particle to a wave*. Whereas “spontaneously”, we tend to characterize a particle by its position in spatial and temporal coordinates. However, one of the main difficulties is that it is not possible to have simultaneously for the particle its precise position and its precise quantity of motion (momentum). This principle, called the Heisenberg uncertainty principle, is a consequence of the Fourier analysis. Besides, it is possible to transpose this principle to any signal analysis, in Finance for example. Indeed, in the Fourier space the results of the Fourier transform on a time series give information about the frequencies which compose it (and not about the quantity of motion). But obviously, the frequencies are measured only on a time segment, i.e. a variable duration or period, which can correspond to all the length of the signal until one of its segments which cannot be reduced to a point. Thus it is not possible to have at the same time a precise frequency and the position of this frequency at a point of the studied time series.

This same observation often leads to use the Fourier transform on segments of the time series.⁵ Segments that slide along the signal to identify the various frequencies contained in this segment. Then, usually, this segment (i.e. duration) finds its transposition in the Fourier space in the form of another segment which expresses

necessary to calculate the coefficients at all the possible frequencies; this decomposition is called then the Fourier series of the original function.

³ Used also in the following scientific fields: Seismology, ground analysis, radio-telescope imaging, medical imaging.

⁴ *Momentum:* In Physics it is a measure of the motion of a body equal to the product of its mass and velocity. Also called linear momentum. Impetus of a physical object in motion.

⁵ This method is called the “Windowed Fourier analysis” proposed by Gabor (1945).

a range of frequencies from the lowest to the highest contained in the considered duration. These segments, respectively in time and frequency, depict rectangles that replace the points, because we cannot have a frequency in a point. The technique is simply called the *short term Fourier transform*. Obviously the risk incurred in this case is that (due to the limitation of the duration through the time segment) the long frequencies are not captured (i.e. depicted) by this type of Fourier decomposition. The graphic illustration of this type of representation in the time–frequency planes is often named: *Heisenberg boxes* (or *Heisenberg rectangles*).

At an intermediate stage during the calculation of the Fourier transform, the sinusoids are provided with Fourier coefficients. The transformation of an arbitrary signal into a sum of sinusoids can create difficulties because we do not know the number of terms contained in the sum. We could even imagine an infinite number of terms. The number of coefficients obtained would be then infinite also and the calculations would become unrealizable. However, more often, a limited number of coefficients is sufficient, and the more the frequencies increase, the more the coefficients tend towards zero. This observation leads to establish the relation with the *sampling theorem*⁶ in the information theory, which is a consequence of the Fourier analysis. In substance, if we position in the Fourier space, the theorem is equivalent to saying that if a signal contains only a limited number of frequencies, it is possible to reproduce it with a finite number of points. Generally, this allows the calculation and digitalization.

The Fourier method does not adapt to all the signals or time series, nor to all the difficulties. Indeed, it is advisable to say that the Fourier analysis is adapted to the resolution of linear problems, i.e. to phenomena where “the effect is proportional to the cause”. The writing of a Fourier series is linear, and in front of a problem of nonlinear nature, recently yet, we could treat the difficulty as if it were of a linear nature, and thus omitting the true problems related to the nonlinearities, i.e. the existence of transitory or turbulent behavior and also the instability of solutions. Furthermore, it is a problem with which the economists had to face in the past, in particular due to the limited capacities of calculation, which led to reduce a phenomenon which is not linear to a linear model. *In the nonlinear system, the most negligible parametric variations can modify more than proportionally the results.* Thus, forecasts and anticipations become more difficult. As explained before, the *Fourier analysis uses the sines and the cosine which oscillate infinitely*, each one with a fixed frequency. However, the signals which have “changing frequencies are not well adapted to this context of infinite time”. Thus, the Fourier space, privileging the frequencies and hiding information about time, can use the same sinusoids to represent very different moments of a time series. That means that any moment of the signal in the Fourier space is confused with any other moment. Even if information about time is not lost, because we can rebuild it by an inverse transformation, there are however

⁶ **Theorem (Sampling Theorem).** *In order for a band-limited (i.e. one with a zero power spectrum for frequencies $\eta > A$) baseband ($\eta > 0$) signal to be reconstructed fully, it must be sampled at a rate $\eta \geq 2A$. A signal sampled at $\eta = 2A$ is said to be Nyquist sampled, and $\eta = 2A$ is called the Nyquist frequency. No information is lost if a signal is sampled at the Nyquist frequency, and no additional information is gained by sampling faster than this rate.*

indeed a confusion or a dissembling of information about time in the Fourier space. For example, a completely flat signal except for its extremities where there are very strong oscillations, will be represented only by the sum of sinusoids oscillating individually in an infinite manner, which will have completely to be compensated and canceled in phase for the flat part of the signal, but however which should be able to represent the strong variations of the signal localized at the extremities. *This type of difficulty is symptomatic of the maladaptation of the Fourier analysis to the signals which have abrupt (or sharp) variations.* However, these abrupt variations contain crucial information, which if they are deformed or not well restored, damage the comprehension of subjacent phenomena.

We can express the problem in a different way, it is possible to say that *the constraint of the infinite oscillation of each sinusoid* at each frequency tends to *diffuse a singularity* of the signal *among all the frequencies of the transform of the signal* so that this singularity is represented (with more or less fidelity). That means that each sinusoid will be able to contain partial information on the aforementioned “singularity”, which will be restored “the least inaccurately” possible only by the cumulative set of all the sinusoids. A discontinuity for example will be represented only by the cumulative set of all possible frequencies. However, the short term Fourier transform, mentioned previously, is a technique which makes it possible to avoid partially the pitfall of the diffusion of abrupt variations. But the principle of the signal segmentation in “sliding windows” creates a problem, indeed, this method cannot depict the long frequencies which exceed the size of the segment or the window chosen, and thus also penalizes the restitution of the signal.

5.1.2 Contribution of the Wavelet Analysis to Irregular and Non-Stationary Time Series: An Approach of Nonlinearities

At this stage, the wavelets are intervened.⁷ They have the characteristic to decompose a signal, at the same time, into time and frequency, which in the Fourier analysis was not realizable in a simultaneous way. The origin and the history of wavelets are difficult to reconstitute, so much the sources are different, however we positions it in 1930. But it is at the same time in mathematics and physics that the subject found an elaborated structure.⁸ The creation of wavelets is attributed to a geophysicist Jean Morlet, within the framework of petroleum prospecting. The former techniques (of the 1960s) used the Fourier analysis of echoes to analyze the soil layers. However, the Fourier analysis tends to diffuse the components of the different layers, the ones in the others, “mixing up” the decomposition of echoes as by interference of layers. These phenomena of *diffusion* and *interference* have been

⁷ We will note the fundamental contributions of J. Morlet, A. Grossmann and Y. Meyer in the analysis and the comprehension of the wavelets.

⁸ The first description in 1971 by K. Wilson, Nobel Prize of physics. Other very different work made it possible to build wavelets called the “self-similar functions of Gabor”.

resolved in 1975 by the (previously mentioned) use of a segmented analysis of the signal by temporal sliding-windows which are also sometimes overlapped with the aim of having a finer local definition. Thus, in this technique the window slid along the signal but always preserved the same length. And inside the window, it is the frequency or the number of oscillations which changed.

A new step was initiated by Morlet, which consisted in preserving constant the number of oscillations while varying the size of the (centered) window. In order to increase the size of the window, as by dilation, by keeping the constant number of oscillations, has the consequence to decrease the frequency inside the window. Indeed, the stretching of the window stretches the oscillations. We can say that this is one of the principles which allows to understand the wavelet technique. *The objective is to locate the high frequencies with small windows and the low frequencies with large windows.* One calls a “mother wavelet” the basic shape which is the subject of the dilation or compression. One of the characteristics of mother wavelets is that they are not necessarily composed of a continuous sinusoids at a unique frequency but can be (now) composed of a combination of various frequencies.

At this stage, it is advisable to distinguish the Gabor functions (sometimes called “gaborets” or “gaborettes”) also called “Gabor wavelets”, which are an intermediate stage between the *short term Fourier analysis* and the wavelets created by Morlet and presented above which have the name of *constant form wavelets*.

It is through the operation of convolution the between the signal and the wavelets, at successive stages of the dilation, that the analysis and decomposition work is carried out. Each *dilation level* (or stretching) is often called *resolution level*. A large window, i.e. a mother wavelet largely stretched (large wavelet), highlights the components of long duration, of low frequency; the term used is the *low resolution*. And a compressed “mother wavelet” (narrow wavelet) highlights the high frequencies, i.e. the components of short duration and transients; the term used is the *high resolution*. This technique is called the *multiresolution*. It was also proved that during the wavelet transform the energy of the signal remains unchanged. The Energy is measured by the average of the square of the amplitude. This means that convolution and the deconvolution, i.e. the return to the initial signal, preserves the energy of the studied time series.

Unlike the *traditional Fourier analysis*, which transforms a signal into a function with one variable, which is the frequency, the wavelet transformation produces a transformation through two variables; the time and the frequency. Thus, the convolution must use “double integrals”. However, the approximation methods allowed to avoid the double integration facilitating thus largely the calculations. Furthermore, it was proved that these approximation methods do not generate miscalculation.⁹ Like what occurs in the Fourier analysis where the objective of the transformation is to calculate the coefficients assigned to each sinusoid, the wavelet analysis produces also coefficients. These coefficients can be used besides through different manners, such as filters for example. A coefficient corresponds to the calculation of the *integral* of the product of the signal with a wavelet at a given resolution level. As

⁹ Proof: see J. Morlet et A. Grossmann.

explained above, in order to facilitate the calculations, approximation methods are preferred in order to avoid the double integration. The set of coefficients obtained at each resolution level synthesizes the set of *information* of the studied signal. By convention, each resolution level is twice finer than the previous (and usually five or six different resolution levels are used). When we say that the resolution level is twice finer, that means that the doubling of the resolution reveals the elements of the signal which have frequencies twice higher. On the other hand, it is possible to decompose the time series in intermediate levels of resolution. The term of “resolution” can also be understood as the number of wavelets used to decompose a signal, indeed the more the resolution increases, the more the wavelet number increases.

In the literature about the wavelets, the term “resolution” corresponds to the terms of “scale” and “octave”. The term “scale” corresponds to the “correlation” between the size of the wavelet and the size of the elements of the signal that we can identify and highlight.

At this stage, it is essential to redefine that *a wavelet does not have necessarily a precise frequency* unlike a sine or a cosine. Indeed, the phenomenon of dilatation or compression of the mother wavelet, related to the size of the window which is used, modifies its frequency. A second element essential for the comprehension of the structure of a wavelet is that the integral of a wavelet is equal to zero (zero integral). This means that during the convolution between the signal and the wavelet, for example, if the signal is perfectly stationary, the coefficient emanating from the calculation will be also equal to zero. The direct consequence of *this observation is that the wavelets highlight the changes in a signal*. However, it is logical to think that the changes in a signal contain important information interesting to analyze. Obviously, the stationary elements are also interesting, but they are generally apprehended by former techniques belonging to approaches of linear nature: i.e. linear systems or Fourier analysis.

Thus, in the decomposition of a signal, we will be able to understand the need for using the “hybrid” methods (or “mixed methods”) which utilize at the same time the properties of the wavelet analysis and those of the Fourier analysis: The first one to highlight the non-stationarities which are potentially related to nonlinearities, and the second one to highlight the stationarities which are potentially related to the linearities.

5.1.3 A Statistical Theory of the Time–Frequency Analysis Remains to Be Developed

The methods of Fourier transforms and the wavelets transforms are merged within the scope of the time–frequency analysis (or time-scale analysis). This analysis framework is largely widespread in many technical fields. Today, we have thus new tools which are particularly interesting for signal analysis specialists. These methods have a common purpose: Capture the best part of a signal, which by direct

reading offers a weak readability. The extraction of its characteristics and intrinsic structures is not necessarily solved and treated in an exhaustive way by the statistical work that we can apply to a signal. The transformation of a sound, or a sonorous echo (or others), medical images for example, exceed the framework of the current statistical analysis. Indeed the statistical analysis does not bring all the answers to the full knowledge of a signal. Thus, there is still a qualitative work to do to assemble the time–frequency analysis and the statistical analysis. This concerns moreover deterministic signals or random signals. Indeed, the signal analysis specialists also work on random or “noized” times series containing sometimes possible non-stationary components. And these subjects can also touch the community of statisticians, which generally neglects the spectral analysis of time series. The use of wavelet bases is increasingly widely used, and many applications, in engineering in particular, tried to utilize the continuous Gabor transforms or the wavelet transforms for statistical purposes. For example to detect, to denoise or to reconstruct a signal, and even more important, to carry out the spectral analysis on non-stationary signals.

The statisticians did not yet really share the benefit of this type of work. The traditional spectral analysis of random or deterministic stationary processes constituted the core of the time–frequency analysis. Today, a very particular attention is focused on the sampling of stationary continuous signals, but also non-stationary, within the framework of time–frequency representations. The main question is to apprehend the importance of time–frequency representations resulting from the Gabor transformations and from the wavelet transforms. Indeed, their potential is large, but still requires an academic validation of the statisticians’ community. A remarkable attempt to bring closer both fields has been presented by R. Carmona, W.L. Hwang and B. Torrésani (1998). This work tries to establish the relation between the contribution of this type of analysis and their statistical interest. Moreover, the authors believe in *the capacity of these methods to provide a toolbox with various advantages and talents for the spectral analysis of non-stationary signals*. Unfortunately, *the corresponding statistical theory is not completely developed yet*. The authors revisit the traditional elements of spectral theory of the stationary random processes in the light of the new tools of the time–frequency analysis.

To revisit the traditional elements of the spectral theory implies the evolution of the *Wiener’s deterministic spectral theory* towards a deterministic spectral theory of the time series in general, by highlighting for example *the duality and the relation between correlogram and periodogram (power spectrum)*.¹⁰ This also implies the necessity to develop methods of non-parametric spectral estimation that aim for example to carry out the analysis of non-stationary processes which can be *locally stationary*, or the analysis of *stationary random processes*.

¹⁰ See section entitled “Wiener theory and time–frequency analysis” which presents some elements of R. Carmona, W. Hwang and B. Torrésani works.

5.2 A Brief Typology of Information Transformations in Signal Analysis

5.2.1 Fourier, Wavelet and Hybrid Analyses

There exist different types of transformation which are either the expression of technological developments or an adaptation to the needs of different application fields. We can list the various methods:¹¹

1. The *Fourier transform*. The analyzing functions are sines and cosines, which oscillate indefinitely. This method is suitable for stationary signals corresponding to constant (or predictable) laws.
2. The *windowed Fourier transform*, i.e. Fourier transform with sliding windows (but of fixed size), also called short term Fourier transform or Gabor transform. The analyzing function is a wave limited in time, multiplied by trigonometric oscillations. That is suitable for quasi-stationary signals, i.e. stationary at the scale of the window.
3. The *wavelet transform* (with variable window). The analyzing functions is a wave limited in time, with a fixed number of oscillations.
4. The *adaptative windowed Fourier transform*, i.e. Fourier transform with adaptive windows, with an *analyzing function of particular forms*, also called *Malvar wavelets*.
5. The *wavelet packets transform*. This method corresponds to a wavelet multiplied by trigonometrical functions. The frequency, position and scale are independent parameters. That is suitable for signals that combine non-stationary and stationary elements (e.g. fingerprint analysis).
6. The Transformation by means of the *Matching Pursuit* algorithm. The analyzing function is a gaussian of variable window size, multiplied by trigonometric functions. The frequency, position of the window and size of the window can change independently. That is suitable for highly non-stationary signals composed of very different elements.

These types of transformation come in a great number of variants.

5.3 The Fourier Transform

5.3.1 Fourier Series and Fourier Transform

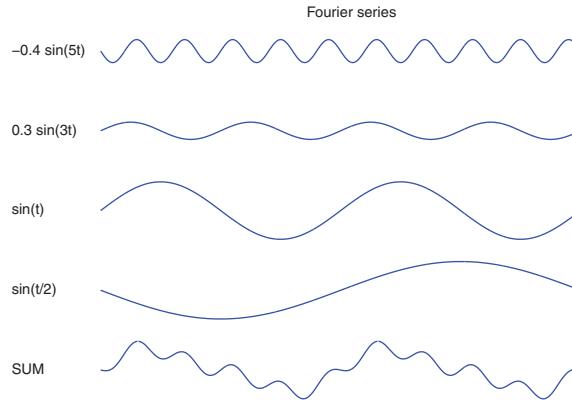
An arbitrary periodic function, even if it is discontinuous, can be represented by a sum of sinusoids of different frequencies, each one is provided with a coefficient. The set of these coefficients makes it possible to reconstitute the function or the

¹¹ Refer to Hubbard (1998) for an interesting report of the information transformations methods.

studied initial series. This is the principle of the Fourier series. The Fourier series allow to transform a time series or a function into a series of independent (differential) equations. Each one of these equations shows the chronological evolution of the coefficients of one of the sinusoids which will compose the initial series. Thus, any periodic curve can be decomposed into a sum of *smooth curves*, because they result from sinusoidal functions. Fourier series can be written for example (without cosine):

$$f(x) = c_1 \cdot \sin \varphi_1(x) + c_2 \cdot \sin \varphi_2(x) + c_3 \cdot \sin \varphi_3(x) + \dots \quad (5.1)$$

or also $f(x) = \sin x + \frac{1}{a} \sin a \cdot x + \frac{1}{b} \sin b \cdot x + \dots$. The Fourier coefficients of this series are $(1, 1/a, 1/b, \dots)$. We can illustrate this remark by means of a numerical and graphic example, i.e.: $f(x) = (-0.4 \sin(5x)) + (0.3 \sin(3x)) + (\sin(x)) + (\sin(x/2))$. We observe that this series is written as the sum of sinusoidal functions. The series and the functions are represented in the graph below:



The measurement of Fourier coefficients at a given frequency, is done using the calculation of the integral of the function or series:

The continuous Fourier transform is written:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \quad (5.2)$$

or, $e^{i\omega t}$ are sinusoids.¹² And *the discrete Fourier transform* $\hat{f}[k]$ is written:

$$\hat{f}[k] = \sum_{n=0}^{N-1} f[n] \exp\left(\frac{-i2\pi kn}{N}\right) \quad \text{for } 0 \leq k \leq N. \quad (5.3)$$

The traditional Fourier transform compares all the signal with infinite sinusoids of various frequencies. On the other hand, the Fourier transform with “sliding window” compares a *segment* of the signal with segments of sinusoids of different frequencies. We can also speak of *local frequencies*. when a segment of the signal was

¹² When we are in the *discrete mode*, the notation $[.]$ is used to symbolize the *sequences*.

analyzed, we can slide the window along the signal, and analyze the following part. Gabor introduced the *Fourier atoms* with window, which had vocation to measure the variations of frequencies:

$$g_{u,\xi}(t) = e^{i\xi t} g(t-u). \quad (5.4)$$

The corresponding *continuous transform with window*¹³ of f is written:

$$\hat{Sf}(\omega) = \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{-i\xi t} dt. \quad (5.5)$$

The *discrete transform* with window¹⁴ is written:

$$Sf[m, l] = \langle f, g_{m,l} \rangle = \sum_{n=0}^{N-1} f[n]g[n-m] \exp\left(\frac{-i2\pi kn}{N}\right). \quad (5.6)$$

5.3.2 Interpretation of Fourier Coefficients

5.3.2.1 Frequencies

For the functions, or signals which vary with time (as it is the case with *the stock exchange fluctuations*), the frequency is usually expressed in Hertz, or in cycles per second. Below two sinusoids of different frequencies:

1. The frequency $\sin(2\pi t)$ corresponds to 1 cycle per second (1 Hz) (Fig. 5.1a).
2. The frequency $\sin(2\pi 2t)$ corresponds to 2 cycles per second (2 Hz) (Fig. 5.1b).

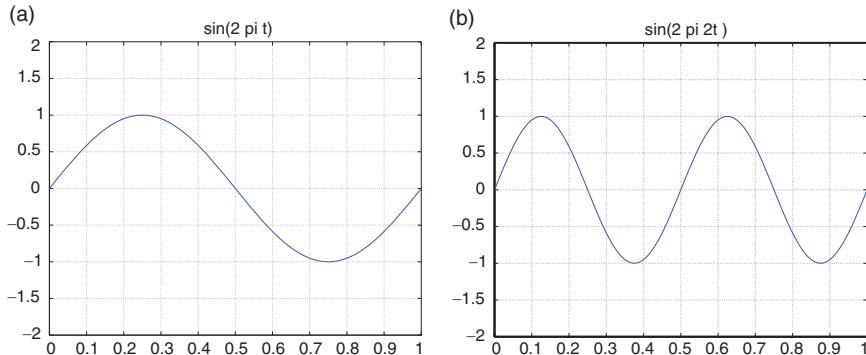


Fig. 5.1 (a) $\sin(2\pi t)$. (b) $\sin(2\pi 2t)$

¹³ Also called “windowed continuous Fourier transform”.

¹⁴ Also called “windowed discrete Fourier transform”.

We transform also functions which vary with space. The transform of a finger-print would present maxima in the neighborhood of the “spatial frequency” equal to 15 *striae per centimeter*. The temporal frequency is the inverse of time. The spatial frequency also called *number of waves*, is the inverse of the length. A function and its Fourier transform are two aspects of the same information. *The Fourier transform however reveals information on frequencies and hides information on the temporal evolution.*

5.3.2.2 Series and Coefficients of Fourier

Series

Any periodic function can be written in Fourier series. If the period is 1 (i.e. if $f(t) = f(t + 1)$), this series takes the form:

$$f(t) = \frac{1}{2}a_0 + (a_1 \cos 2\pi t + b_1 \sin 2\pi t) + (a_2 \cos 2\pi 2t + b_2 \sin 2\pi 2t) + \dots \quad (5.7)$$

Coefficients

The coefficient a_k measures the “quantity” of cosine function $\cos 2\pi kt$ of frequency k contained in the signal; the coefficient b_k measures the “quantity” of sine function $\sin 2\pi kt$ of frequency k contained in the signal. The Fourier series includes only sinusoids of frequencies equal to multiple integers of the fundamental frequency (this fundamental frequency is the inverse of the period). We can write the formula above in a more usual way:

$$f(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} (a_k \cos 2\pi kt + b_k \sin 2\pi kt), \quad (5.8)$$

where k represents the frequency. The formula above makes it possible to reconstruct a function using its Fourier coefficients: we multiply each sinusoid by its coefficient (we modify its amplitude) and we add, point by point, the functions thus obtained; the first term is divided by 2. By convention, usually ξ corresponds to x for the transform of a signal which varies with space. All 2π make heavier the formulas but are inevitable to express the periodicity: The function ($\sin 2\pi t$) is of period 1, while $\sin(t)$ is of period 2π (Fig. 5.2).

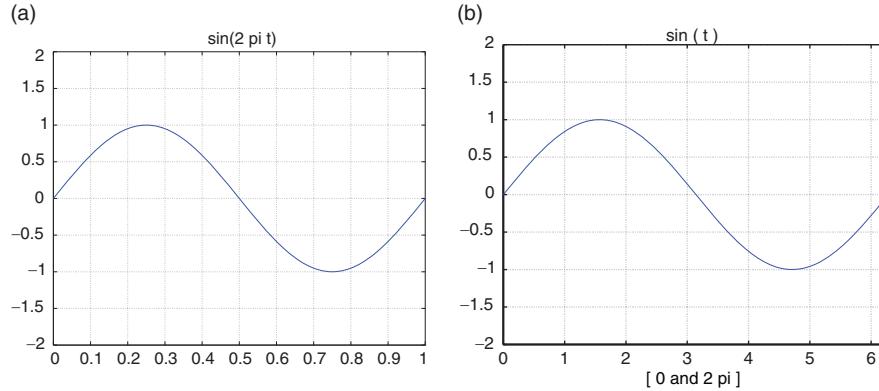


Fig. 5.2 (a) $\sin(2\pi t)$. (b) $\sin(t)$, between 0 and 2π

We obtain Fourier coefficients of a function $f(t)$ of period 1 using the integrals:

$$a_k = 2 \int_0^1 f(t) \cos 2\pi k t dt \quad \text{and} \quad b_k = 2 \int_0^1 f(t) \sin 2\pi k t dt \quad (5.9)$$

that means that we calculate the product of the function f by the sine or cosine of frequency k , we integrate it, and we multiply the result by 2. Integrate a function comes down to measure the area located under a portion of its curve.

5.4 The Gabor Transform: A Stage Between the Short Term Fourier Transform and the Wavelet Transform

5.4.1 The Gabor Function

In the *windowed Fourier analysis* the size of the window remains fixed. It is the number of oscillations inside the window which varies, at the same time making vary the frequency. In the short term Fourier analysis the window is not fixed, but it moves along the signal to decompose it by successive segments. This technique has inspired the conception of a new technique. It is the *Gabor transform*, which is the first step towards the wavelet analysis. It is an intermediate stage between the short term Fourier analysis and the wavelet analysis.

The method uses a Gabor function (also known as *gaboret* or also *Gabor atom*).¹⁵ The length of the window of the Gabor function remains constant, unlike what occurs with a wavelet. For a wavelet the stretching of the window modifies its length and in certain cases its amplitude also. In fact the length and amplitude of a Gabor function do not change. The only element which changes is the frequency inside the interval of the Gabor function. This principle induces a type of signal analysis by

¹⁵ The Gabor function is also called the “gaboret” or “gaborette”.

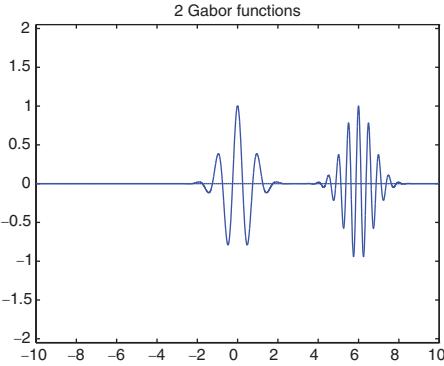


Fig. 5.3 Two Gabor functions: $g_{(0,2\pi)}(x)$ and $g_{(6,4\pi)}(x)$

segment. The Gabor function moves along the signal, it combines with the signal and produces its coefficients which are as many sampled decompositions of the signal by frequency. The displacement of the Gabor function along the signal is done by means of the window and *centering parameter b*. Thus, *b* indicates the *position* in time of the Gabor function and of the window

$$g_{(b,\omega)}(x) = g(x-b)e^{i\omega(x-b)}. \quad (5.10)$$

Using an example of window such as:

$$g(x-b) = \exp^{-(t-b)^2}. \quad (5.11)$$

We can illustrate the subject with a quantified example and Fig. 5.3.

$$g_{(0,2\pi)}(x) = g(x)e^{i2\pi x} = e^{-t^2}e^{i2\pi x} \quad (5.12)$$

$$g_{(6,4\pi)}(x) = g(x-6)e^{i4\pi(x-6)} = e^{-(t-6)^2}e^{i4\pi(x-6)}. \quad (5.13)$$

5.4.2 The Gabor Transform with a Sliding Window: The “Gabor Wavelet”

The principle of the *Gabor function* such as it is quickly described above, shows that its use for the analysis of a signal will be done by sliding the Gabor function (Gaboret or gaborette) along the signal and each segment of the aforesaid signal will be analyzed by a “range (or scale)” of Gabor functions of different internal frequencies, but of identical amplitude and length. The Gabor transform illustrates one of the types of windowed transform. The Gabor transform of an unspecified function $f(x)$ is written:

$$\int f(x)g(x-b)e^{-i\omega x}dx \quad (5.14)$$

or can be written in a more practical form:

$$G_f(b, \omega) = \int f(x)g(x-b)^*e^{-i\omega(x-b)}dx = \langle f, g_{(b,\omega)} \rangle \quad (5.15)$$

with a *Gabor wavelet* $g_{(b,\omega)}$:

$$g_{(b,\omega)} = g(x-b)e^{i\omega(x-b)}. \quad (5.16)$$

The *reconstruction* of the function or initial series is carried out by the inversion of the Gabor transform.

5.5 The Wavelet Transform

5.5.1 A Wavelet ψ Is a Function of Zero Average, i.e. Zero-Integral: $\int_{-\infty}^{+\infty} \psi(t)dt = 0$

The above condition is a characteristic of wavelets and of all the functions which are potentially used as such. Although we can find some exceptions (e.g. Gauss pseudo-wavelet that is not considered as a true wavelet), the fact that the integral of a wavelet is equal to zero allows during the operation of convolution between the signal and the wavelet to identify the signal components which are of a *non-stationary nature*. In this case the wavelet coefficient resulting from the convolution will be different from zero (Fig. 5.4).

5.5.1.1 Conditions for Obtaining a Good Wavelet

It is said commonly that any function with zero-integral can be regarded as a wavelet. However additional conditions make it possible to obtain good wavelets. Firstly, it is said that they must be “regular”, i.e. to have a support limited in

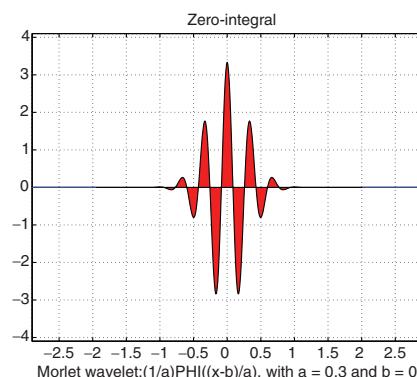


Fig. 5.4 Zero integral, zero area

frequency. It is necessary also that they have a support limited in time. Secondly, beyond the basic condition stated above: $\int_{-\infty}^{+\infty} \psi(t)dt = 0$, we add the following constraints relative to the *moments*:

$$\int_{-\infty}^{+\infty} t\psi(t)dt = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} t^n\psi(t)dt = 0. \quad (5.17)$$

Note that all the wavelets do not satisfy necessarily these conditions.

5.5.2 Wavelets and Variable-Window

The principle inherent to the wavelets is the *variable window* which is applied to the *mother wavelet*. The size of a variable-window is dilated or compressed. Thus, only the frequency and sometimes the amplitude of oscillations contained in the window vary. This technique is truly at the root of the construction of a wavelet transform.

5.5.3 The Wavelet Transform

5.5.3.1 Scale of Frequencies and Level of Resolution: Illustrations

The principle of the wavelet transform is different from the transformation by the *Gabor function* (also called “Gabor atom”). Indeed, contrary to the principle of the construction of the Gabor function, which varies its frequency inside the same segment which moves on the signal, the wavelet, in this case, has a variable amplitude, or more exactly it can be “dilated” (i.e. extended) or “contracted” (i.e. compressed). The purpose is to analyze the signal by a range of wavelets more adapted to the shape of the signal. Thus, by using a type of basic wavelet (of which there is an infinity) adapted to specificities of the signal, we can obtain a certain number of sub-wavelets which are only the variation through the amplitude, height and length of the initial wavelet. Thus, we contract or we dilate the wavelet to change the size of the “window” and therefore the size of the “scale” to which we observe the signal; the “frequency” of the wavelet changes at the same time as the size of the window. *The scale is the level of resolution of the observation*. Indeed, with large wavelets the examination of the signal is carried out at a coarse resolution (with a small number of coefficients). That is why the wavelets have sometimes been called the “mathematical microscope”. The compression of the wavelet increases the enlargement (i.e. magnification) and the resolution which highlights the increasingly precise details of the signal. The level of resolution is also called the “number of octaves” (usually six or five). As we will see again in next sections, it is possible to analyze a signal with intermediate resolutions by the octaves which offer partially “redundant” information. *Scale, resolution or octave* express the same reality. The transformation gives a “number of samples” of the signal. The compute of each transform at a given

scale provides a “sample” (or observation) of the signal. The method can be applied to the entire signal (or possibly to segments of it by using sliding-windows). Here are two examples of wavelet transform.

Example 5.1. (Discrete wavelet transform). For an arbitrary signal, see Fig. 5.5.

Figure 5.5 shows a discrete wavelet transform on five scales (with cubic spline wavelet). Note that the transforms, somehow, stationarize the signal (see supra).

Example 5.2. (Continuous wavelet transform). For French stock index, see Fig. 5.6.

Figure 5.6 shows a continuous wavelet transform using a Gauss pseudo-wavelet which is a very particular case where the signal is not stationarized. The finest resolution of the wavelet transform at the octave 5 is in the higher part of the figure, it provides the most details. We observe the transform at given scale.

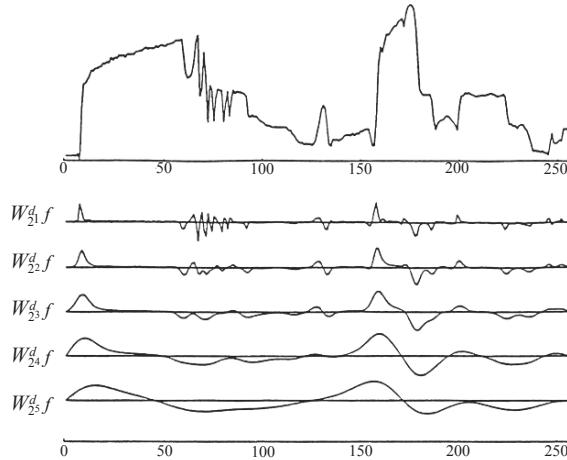


Fig. 5.5 Discrete (dyadic) wavelet transform of an arbitrary signal

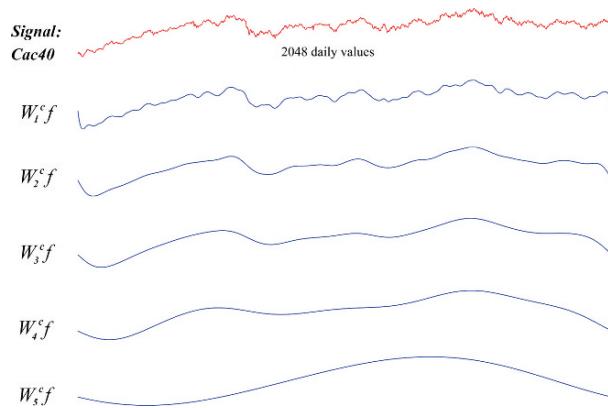


Fig. 5.6 Continuous wavelet transforms of a stock index by Gauss pseudo-wavelet

The term *scale* indicates the *correlation between the size of the wavelet and the size of the components that one can see*. Note that the scales are usually sampled in geometric progression. The term *octave* indicates that the doubling of the resolution causes to raise the frequency of wavelets in order to observe the components of double frequencies. The effective work of the window of the *wave* consists in extracting the “transitory” components of the signal at the high frequencies with a narrow window and the long duration components at the low frequencies with a large window. Thus, the size of the window is *variable*. Thus, the variable elements (after the window) are the scale and the frequency. We’ll give later two examples of this variation from a Morlet basic wavelet (i.e. Morlet mother wavelet). If we write:

$$\text{Mother wavelet: } \psi(x), \quad (5.18)$$

$$\text{Wavelet family: } \psi_{(a,b)}(x) = \frac{1}{a} \psi\left(\frac{x-b}{a}\right), \quad (5.19)$$

obviously, we do not calculate a wavelet transform of a finite length signal for each possible value of the two parameters a and b , we restrict to finite sets and regular scales. The length of the signal divided by the length of the *wavelet support* noted I_ψ provides the maximum of the *scale parameter* “ a ”, that we note here A_{\max} . The floor of the scale parameter is provided by dividing the *central frequency* of the wavelet by the sampling rate of the signal (Torresani 1995). The extraction of the transform at each value of the scale provides a “sample” of the signal. The size of the wavelet is variable. It is possible to create by using for example a Morlet “mother wavelet” written as follows:

$$\psi(x) = e^{-x^2/2} e^{i\omega x}, \quad \text{with } \omega = 5.5 \quad (5.20)$$

a *wavelet family* by means of the variation of the parameter couple (b, a) . Due to the fact that the lateral displacements of the wavelet according to the variation of b are simple to imagine, we keep the parameter of “centering” b equal to zero. However, it is interesting to observe the form of the wavelet, if we vary the scale parameter a and for example between values $1/5$ and 1 .

5.5.3.2 Variation of the Dilatation Parameter

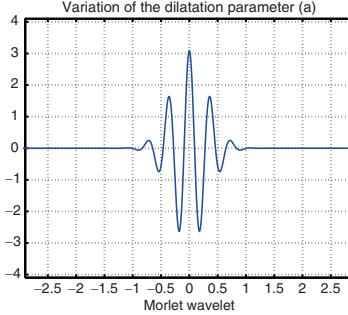
Thus, the *wavelet family* for $a = (1/5) \dots 1$, is written:

$$\psi_{(a,0)}(x) = \frac{1}{a} \psi\left(\frac{x}{a}\right) \quad (5.21)$$

and we can write for better represent them, by way of illustration, two sub-wavelets extracted from the family expressed above.

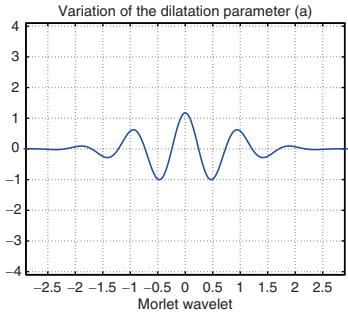
- For $a = 1/3$, the *sub-wavelet*, also called “daughter wavelet”, is written:

$$\psi_{((1/3),0)}(x) = \frac{1}{(1/3)} \psi\left(\frac{x}{1/3}\right) = 3 \times e^{-(3x)^2/2} e^{i(5.5)3x}. \quad (5.22)$$



- For $a = 0.83$ the wavelet is written:

$$\psi_{(0.83),0}(x) = \frac{1}{0.83} \psi\left(\frac{x}{0.83}\right) = \frac{1}{0.83} \times e^{(-x/0.83)^2/2} e^{i(5.5)(x/0.83)}. \quad (5.23)$$



Generally, in this type of construction, when the scale parameter increases, the length of the “central frequency” of the wavelet tends to increase and, on the other hand, the height tends to decrease.

5.5.3.3 Concomitant Variation of the Parameters (b,a)

This provides a complete range, or a family of wavelets, which is written (as described previously): $\psi_{(b,a)}(x) = e^{-((x-b)/a)^2/2} e^{i\omega x/a}$. We can represent two wavelets of this family, for example $\psi_{(0,1)}$ and $\psi_{(6,2)}$.

In Fig. 5.7, we observe the lateral displacement, the change of amplitude, but also the shrinking of the segment inside of which the central frequency is observable. Another important aspect of the use of sliding windows is related to the good adequacy of the size of the window to the size of the studied object. (Recall: a is the dilation parameter or *scale* and b is the *position*.)

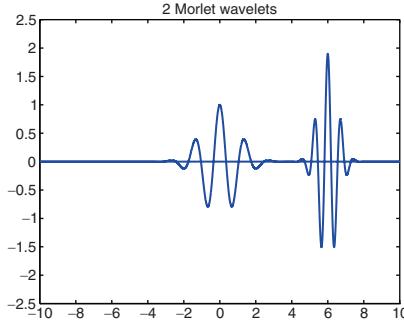


Fig. 5.7 $[\psi_{(0,1)}(x); \psi_{(6,2)}(x)]$: two Morlet wavelets $\psi_{(b,a)}(x)$, with $\omega = 5.5$

5.5.3.4 Transform and Inverse Transform

The wavelet transform can be written:

$$T_f(b, a) = \langle f, \psi_{(b,a)} \rangle = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{a}} \psi^* \left(\frac{x-b}{a} \right) dx. \quad (5.24)$$

Like for the inverse Fourier transform, we can write the inverse wavelet transform.¹⁶
The inverse transform is written:

$$f(x) = c_\psi \int_{-\infty}^{+\infty} \int_0^{+\infty} T_f(b, a) \frac{\psi_{(b,a)}}{a} da db, \quad (5.25)$$

where the constant c_ψ is written: $c_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|}{\omega} d\omega < +\infty$.

5.5.4 Wavelet Transform and Reconstruction

5.5.4.1 Continuous Wavelet Transform and Reconstruction

The wavelet model has first been formalized by Grossmann and Morlet. For the sake of simplification, let us use the following notation: Let $\psi_s(x)$ be the dilation of the wavelet $\psi(x)$ by a factor s :

$$\psi_s(x) = \frac{1}{s} \psi \left(\frac{x}{s} \right). \quad (5.26)$$

The wavelet transform of a function $f(x)$ at the scale s and position x is given by the convolution product:

$$W_s f(x) = f * \psi_s(x). \quad (5.27)$$

¹⁶ Theorem of Calderon, Grossmann, Morlet.

Given $\widehat{\psi}(\omega)$ the *Fourier transform* of $\psi(x)$. Morlet and Grossmann proved that if the wavelet $\psi(x)$ has a Fourier transform equal to zero at $\omega = 0$, then the wavelet transform satisfies an *energy conservation equation* and $f(x)$ can be reconstructed from its wavelet transform. The wavelet $\psi(x)$ can be understood as the *impulse response of a band-pass filter* and the *wavelet transform as a convolution with a band-pass filter which is dilated*. When the scale s is large, $W_s f(x)$ detects the lower frequency components of the signal $f(x)$. When the scale s decreases, the support of $\psi_s(x)$ decreases so the wavelet transform $W_s f(x)$ is sensitive to finer details. The scale s determines the size and the regularity of signal features extracted by the wavelet transform.

5.5.4.2 Dyadic Wavelet Transform and Reconstruction Formula

The wavelet transform depends on two parameters s and x which vary continuously over the set of real numbers. For practical applications, s and x must be discretized. For a specific class of wavelets, the scale parameter can be sampled along the dyadic sequence $[2^j]_{j \in \mathbb{Z}}$, without modifying the properties of the transform. The wavelet transform at the scale 2^j is given by:

$$W_{2^j} f(x) = f * \psi_{2^j}(x). \quad (5.28)$$

At each scale 2^j , the function $W_{2^j} f(x)$ is continuous since it is equal to the convolution of two functions in $\mathbf{L}^2(\mathbb{R})$. The *Fourier transform* of $W_{2^j} f(x)$ is

$$\widehat{W}_{2^j} f(\omega) = \widehat{f}(\omega) \widehat{\psi}(2^j \omega). \quad (5.29)$$

When the following *constraint* is imposed:

$$\sum_{j=-\infty}^{+\infty} |\widehat{\psi}(2^j \omega)|^2 = 1 \quad (5.30)$$

then the whole frequency axis is covered by a dilation of $\widehat{\psi}(\omega)$ by the scales factors $[2^j]_{j \in \mathbb{Z}}$. Any wavelet satisfying the constraint is called a *dyadic wavelet*. The dyadic wavelet transform corresponds also to the sequence of functions:

$$[W_{2^j} f(x)]_{j \in \mathbb{Z}}. \quad (5.31)$$

Let “ \mathcal{W} ” be the *dyadic wavelet operator* defined by $\mathcal{W}f = [W_{2^j} f(x)]_{j \in \mathbb{Z}}$. From the equation of the Fourier transform of $W_{2^j} f(x)$ with its constraint (supra), and by using the Parseval theorem, we can write an *energy conservation equation*:¹⁷

$$\|f\|^2 = \sum_{j=-\infty}^{+\infty} \|W_{2^j} f(x)\|^2. \quad (5.32)$$

¹⁷ Remark: The norm (energy) of $f(x) \in \mathbf{L}^2(\mathbb{R})$ is given by $\|f\|^2 = \int_{-\infty}^{+\infty} |f(x)|^2 dx$.

We pose $\bar{\psi}_{2j}(x) = \psi_{2j}(-x)$. The function $f(x)$ is *reconstructed* from its dyadic wavelet transform by the *reconstruction formula*:

$$f(x) = \sum_{j=-\infty}^{+\infty} W_{2j} f * \bar{\psi}_{2j}(x). \quad (5.33)$$

Given \mathbf{V} , the space of the dyadic wavelet transforms $[W_{2j} f(x)]_{j \in \mathbb{Z}}$, for all the functions $f(x) \in \mathbf{L}^2(\mathbb{R})$.¹⁸ Then, $\mathbf{L}^2(\mathbf{L}^2)$ the Hilbert space of all sequences of functions $[h_j(x)]_{j \in \mathbb{Z}}$, such that $h_j(x) \in \mathbf{L}^2(\mathbb{R})$ and $\sum_{j=-\infty}^{+\infty} \|h_j(x)\|^2 < +\infty$. The energy conservation equation (see supra) proves that \mathbf{V} is a subspace of $\mathbf{L}^2(\mathbf{L}^2)$. Then, \mathcal{W}^{-1} denotes the operator from $\mathbf{L}^2(\mathbf{L}^2)$ to $\mathbf{L}^2(\mathbb{R})$ defined by: $\mathcal{W}^{-1}[h_j(x)]_{j \in \mathbb{Z}} = \sum_{j=-\infty}^{+\infty} h_j * \bar{\psi}_{2j}(x)$. The reconstruction formula (see supra) shows that the restriction of \mathcal{W}^{-1} to the wavelet space \mathbf{V} is the inverse of the dyadic wavelet transform operator \mathcal{W} . Every sequence of functions $[h_j(x)]_{j \in \mathbb{Z}} \in \mathbf{L}^2(\mathbf{L}^2)$ is not – a priori – the dyadic wavelet transform of some function $f(x) \in \mathbf{L}^2(\mathbb{R})$. In fact, if there is a function $f(x) \in \mathbf{L}^2(\mathbb{R})$ such that $[h_j(x)]_{j \in \mathbb{Z}} = \mathcal{W}f$, then we should have $\mathcal{W}[\mathcal{W}^{-1}[h_j(x)]_{j \in \mathbb{Z}}] = [h_j(x)]_{j \in \mathbb{Z}}$. When we replace the operators \mathcal{W} and \mathcal{W}^{-1} by their expression given in the equations $W_{2j} f(x) = f * \psi_{2j}(x)$ (supra: wavelet transform) and $\mathcal{W}^{-1}[h_j(x)]_{j \in \mathbb{Z}} = \sum_{j=-\infty}^{+\infty} h_j * \bar{\psi}_{2j}(x)$ (see supra), then it comes:

$$\sum_{l=-\infty}^{+\infty} h_l * K_{l,j}(x) = h_j(x) \text{ with } j \in \mathbb{Z} \text{ and } K_{l,j}(x) = \bar{\psi}_{2j} * \psi_{2l}(x). \quad (5.34)$$

The sequence $[h_j(x)]_{j \in \mathbb{Z}}$ is a dyadic wavelet transform if and only if the aforementioned equations $\sum_{l=-\infty}^{+\infty} h_l * K_{l,j}(x) = h_j(x)$ and $K_{l,j}(x) = \bar{\psi}_{2j} * \psi_{2l}(x)$ hold. These equations are known as *reproducing kernel equations*, and show the correlation between the functions $W_{2j} f(x)$ of a dyadic wavelet transform. The correlation between the functions $W_{2j} f(x)$ and $W_{2l} f(x)$ at two different scales 2^j and 2^l can be understood by observing their Fourier transform: $\widehat{W}_{2j} f(\omega) = \widehat{f}(\omega) \widehat{\psi}(2^j \omega)$ and $\widehat{W}_{2l} f(\omega) = \widehat{f}(\omega) \widehat{\psi}(2^l \omega)$. The redundancy of $\widehat{W}_{2j} f(\omega)$ and $\widehat{W}_{2l} f(\omega)$ depend on the overlap of the functions $\widehat{\psi}(2^j \omega)$ and $\widehat{\psi}(2^l \omega)$. The energy of this overlap is equal to the energy of the kernel $K_{l,j}(x)$. (It is maximum for $l = j - 1$, $l = j + 1$.) Let \mathbf{P}_V be an operator defined by

$$\mathbf{P}_V = \mathcal{W} \circ \mathcal{W}^{-1} \quad (5.35)$$

this operator is a projector from $\mathbf{L}^2(\mathbf{L}^2)$ on the \mathbf{V} space. It is possible to prove that any sequence of functions $[h_j(x)]_{j \in \mathbb{Z}} \in \mathbf{L}^2(\mathbf{L}^2)$ satisfies $\mathbf{P}_V[h_j(x)]_{j \in \mathbb{Z}} \in \mathbf{V}$, and any element of \mathbf{V} is invariant under the action of this operator.

¹⁸ Recall: The following spaces correspond to the respective functions or associated signals: (1) $\mathbf{L}^2(\mathbb{R})$: Finite energy functions: $\int |f(t)|^2 dt < +\infty$; and consequently is the space of integrable square functions. (2) $\mathbf{L}^p(\mathbb{R})$: Functions such that $\int |f(t)|^p dt < +\infty$. (3) $\mathbf{L}^2(\mathbb{Z})$: Discrete finite energy signals: $\sum_{n=-\infty}^{+\infty} |f(n)|^2 < +\infty$. (4) $\mathbf{L}^p(\mathbb{Z})$: Discrete signals such that: $\sum_{n=-\infty}^{+\infty} |f(n)|^p < +\infty$.

5.5.4.3 Dyadic Wavelet Transform and Maxima

A signal is usually measured with a finite resolution which imposes a finer scale when computing the wavelet transform. For practical purposes, the scale parameter must also vary on a finite range. We are going to explain how to interpret mathematically a dyadic wavelet transform on a finite range. In both previous sections (for the sake of simplification) the model was based on functions of a continuous parameter x , but we have to discretize the abscissa x and explain efficient algorithms for computing a discrete wavelet transform and its inverse.

Interpretation of a Dyadic Wavelet Transform

For practical purposes, it is not possible to compute the wavelet transform at all scales 2^j for j varying between $-\infty$ and $+\infty$. In fact, we are limited by a finite larger scale and a non-zero finer scale. In order to normalize, we suppose that the finer scale is equal to 1 and 2^J is the largest scale. (With $f(x) \in L^2$). Between the scales 1 and 2^J , the wavelet transform $[W_{2^j}f(x)]_{1 \leq j \leq J}$ can be interpreted as the details available when smoothing $f(x)$ at the scale 1 but which have disappeared when $f(x)$ at the larger scale 2^J . At this stage, we introduce a function $\phi(x)$ whose Fourier transform is:

$$\left| \widehat{\phi}(\omega) \right|^2 = \sum_{j=1}^{+\infty} |\widehat{\psi}(2^j\omega)|^2. \quad (5.36)$$

Because we know that the wavelet $\psi(x)$ verifies $\sum_{j=1}^{+\infty} |\widehat{\psi}(2^j\omega)|^2 = 1$, so we have $\lim_{\omega \rightarrow 0} |\widehat{\phi}(\omega)|^2 = 1$. In addition, the energy distribution of the Fourier transform $\widehat{\phi}(\omega)$ is localized in the low frequencies, thus $\phi(x)$ is a smoothing function. Given S_{2^j} the smoothing operator defined as follows:

$$S_{2^j}f(x) = f * \phi_{2^j}(x) \quad \text{where} \quad \phi_{2^j}(x) = \frac{1}{2^j} \phi_{2^j}\left(\frac{x}{2^j}\right). \quad (5.37)$$

The larger the scale 2^j , the more details of $f(x)$ are removed by the smoothing operator S_{2^j} . The dyadic wavelet transform $[W_{2^j}f(x)]_{1 \leq j \leq J}$ between the scale 1 and 2^J give the details available in $S_1f(x)$ but not for $S_{2^J}f(x)$. The Fourier transforms $\widehat{S}_1f(\omega)$, $\widehat{S}_{2^J}f(\omega)$, $\widehat{W}_{2^j}f(\omega)$ of $S_1f(x)$, $S_{2^J}f(x)$, $W_{2^j}f(x)$ are respectively given by:

$$\widehat{S}_1f(\omega) = \widehat{\phi}(\omega)\widehat{f}(\omega), \quad \widehat{S}_{2^J}f(\omega) = \widehat{\phi}(2^J\omega)\widehat{f}(\omega), \quad \widehat{W}_{2^j}f(\omega) = \widehat{\psi}(2^j\omega)\widehat{f}(\omega). \quad (5.38)$$

The first equation $|\widehat{\phi}(\omega)|^2 = \sum_{j=1}^{+\infty} |\widehat{\psi}(2^j\omega)|^2$ (i.e. Fourier transform of $\phi(x)$) gives:

$$\left| \widehat{\phi}(\omega) \right|^2 = \sum_{j=1}^J |\widehat{\psi}(2^j\omega)|^2 + \left| \widehat{\phi}(2^J\omega) \right|^2. \quad (5.39)$$

From the equation above and from the three equations of Fourier transforms (see supra) $\widehat{S}_1 f(\omega)$, $\widehat{S}_{2^J} f(\omega)$, $\widehat{W}_{2^J} f(\omega)$ and by using Parseval's theorem, we obtain the following energy conservation equation:

$$\|S_1 f(x)\|^2 = \sum_{j=1}^J \|W_{2^j} f(x)\|^2 + \|S_{2^J} f(x)\|^2. \quad (5.40)$$

Such an equation shows that the higher frequencies of $S_1 f(x)$ which disappeared in $S_{2^J} f(x)$ can be recovered from the dyadic wavelet transform $[W_{2^j} f(x)]_{1 \leq j \leq J}$ between the scales 1 and 2^J . The functions $\{S_{2^J} f(x), [W_{2^j} f(x)]_{1 \leq j \leq J}\}$ are known as the *finite scale wavelet transform* of $S_1 f(x)$. For practical purposes, the signals are given by discrete sequences of values, and we know that any *discrete signal* D of finite energy can be interpreted as the uniform sampling of some function smoothed at the scale 1. (*Remark: Given* $D = [d_n]_{n \in \mathbb{Z}}$ *a discrete signal of finite energy:* $\sum_{n=-\infty}^{+\infty} |d_n|^2 < +\infty$. *Suppose that the Fourier transform* $\widehat{\phi}(\omega)$ *verifies:* $\exists C_1 > 0, \exists C_2 > 0, C_1 \leq \sum_{n=-\infty}^{+\infty} |\widehat{\phi}(\omega + 2n\pi)|^2 \leq C_2$ *where* $\omega \in \mathbb{R}$. *There exists a function* $f(x) \in L^2(\mathbb{R})$ *which is not unique, such that for any* $n \in \mathbb{Z}$, $S_1 f(n) = d_n$.) Thus the discrete signal D can be rewritten as follows $D = [S_1 f(n)]_{n \in \mathbb{Z}}$. For a specific class of wavelets $\psi(x)$, the samples $[S_1 f(n)]_{n \in \mathbb{Z}}$ allow to compute a uniform sampling of the finite scale wavelet transform of $S_1 f(x)$: $\{[S_{2^J} f(n)]_{n \in \mathbb{Z}}, [[W_{2^j} f(n + \rho)]_{n \in \mathbb{Z}}]_{1 \leq j \leq J}\}$ where ρ is the sampling shift which depends on the wavelet $\psi(x)$ (Meyer 1989, Appendix 2, pp. 267–268). Then, we pose: $W_{2^J}^d f = [W_{2^j} f(n + \rho)]_{n \in \mathbb{Z}}$ and $S_{2^J}^d f = [S_{2^J} f(n)]_{n \in \mathbb{Z}}$. The sequence of discrete signal $\{S_{2^J}^d f, [W_{2^j} f]_{1 \leq j \leq J}\}$ is known as a *discrete dyadic wavelet transform* of the signals $D = [S_1 f(n)]_{n \in \mathbb{Z}}$. Let \mathcal{W}^d be the discrete wavelet transform operator which associates to a signal D the discrete wavelet transform previously defined. (*Remark: This operator uses a fast algorithm; note that if the signal possesses N non-zero samples, the level of complexity of such an algorithm is O(N log(N)); the algorithm uses a cascade of convolutions with two discrete filters. It is also possible to compute the discrete inverse wavelet transform* W^{-1d} *which allows the reconstruction of the signal D from its discrete dyadic wavelet transform*). A model of a cubic spline wavelet $\psi(x)$ is shown in Fig. 5.8 (left). The shape of the function $\phi(x)$ corresponding to the wavelet $\psi(x)$ is shown in Fig. 5.8 (right) ($\phi(x)$ is also a cubic spline but with a compact support of size 4).

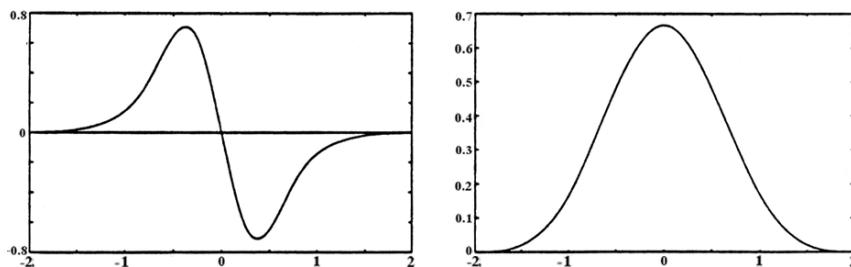


Fig. 5.8 Left: $\psi(x)$. Right: $\phi(x)$

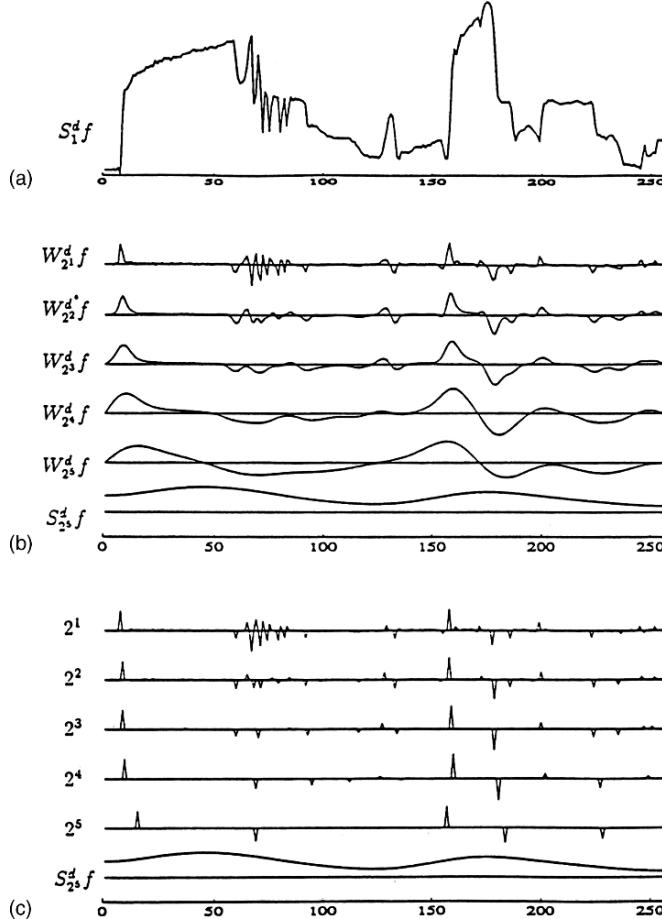


Fig. 5.9 (a) An arbitrary signal, (b) wavelet transforms (c) maxima (ref. to Mallat, Meyer 1989)

Figure 5.9a shows a *signal* (256 samples), Fig. 5.9b is the *discrete wavelet transform* of the signal decomposed on five scales with a cubic spline wavelet. Figure 5.9c corresponds to the *maxima representation* of the signal. At each scale, the *Diracs* show the position and amplitude of a local maximum of the wavelet transform given in Fig. 5.9b.

The curve denoted $S_{2^5}^d f$ is the coarse signal. Since the wavelet used is the first derivative of a smoothing function, the maxima of the wavelet transform show the points of sharper variation at each scale. The wavelet maxima representation of a signal D is defined by the discrete maxima of its discrete wavelet transform $W_{2^j}^d f$ for each scale $1 \leq 2^j \leq 2^J$, plus the coarse signal $S_{2^J}^d f$. The maxima show at different scales the position of the signal sharper variations. (Note that for the functions $\phi(x)$ whose support is larger than 2, the discrete maxima detection produces errors; note also that the function $\phi(x)$ used previously has a compact support of size 4.)

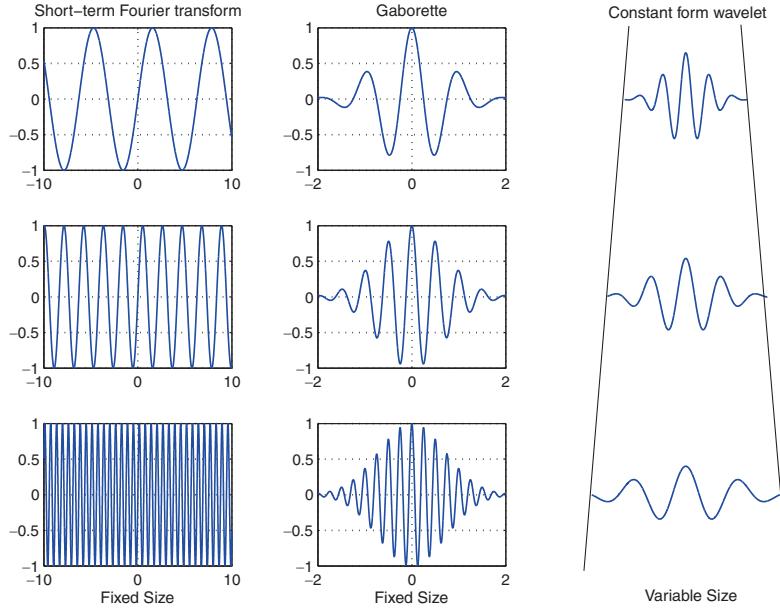


Fig. 5.10 Distinction between the different window mechanisms

5.6 Distinction of Different Window Mechanisms by Type of Transformation

In previous sections, three types of basic transformation have been distinguished, i.e. the Fourier transform (of short term), the Gabor transform and the wavelet transform. In order to distinguish the different window mechanisms, it seems useful to represent them in a graph. We observe clearly in Fig. 5.10, that for *the short term Fourier transform and the Gabor transform*, the window has a fixed size and consequently, it slides along the signal. On the other hand, *the size of the window is variable for the wavelet transform*. Indeed, we clearly distinguish the dilation of the wavelet by increase in the size of the window; the mechanism is easily visible.

5.7 Wavelet Transform of Function or Time Series

5.7.1 The Wavelets Identify the Variations of a Signal

The wavelets are a natural extension of the Fourier analysis. The purpose is always to transform a signal into coefficients, from which we can reconstruct the initial signal. The transformation of a signal is carried out with the sum of different wavelets.

If a variable size window of a wavelet is chosen narrow, we capture the *sharp* and short *variations* of the signal, i.e. events of the peak type, or the discontinuities. If the window is large, we capture *long oscillations* and long variations, i.e. low frequencies. The choice of the window is conditioned by what we want to extract. The wavelets highlight with precision the variations of the signal. An absence of variation or a very slow variation will provide a very small coefficient or equal to 0. A *wavelet* has *shapes more technical*, more complex and more adapted than a *sinusoid*. A wavelet coefficient measures the correlation between the wavelet and the signal (or portion of signal to which we apply it). Recall: a wavelet has a zero integral, and the positive areas neutralize negative areas. The wavelets are also a way of highlighting *changes of state* of a signal (i.e. transitions) with a reduced number of non-zero coefficients. The size of a window allows to capture the different components of a signal.

Figure 5.11 is shown to understand the mechanism of signal analysis and decomposition by a wavelet. A signal is artificially built, then a wavelet. Then we displace along the signal, a wavelet whose window size remains constant, and in so doing we carry out a convolution between the wavelet (whose size remains constant) and the signal. The convolution (*product*) of each segment of the signal with the wavelet provides a *new curve*. The *area* (i.e. the *integral* of this *curve*) gives the *wavelet coefficient*. The *segment* of the signal, which has “*similarities*” with the *wavelet*, will have a product (segment * wavelet) providing a curve with a *high coefficient*.

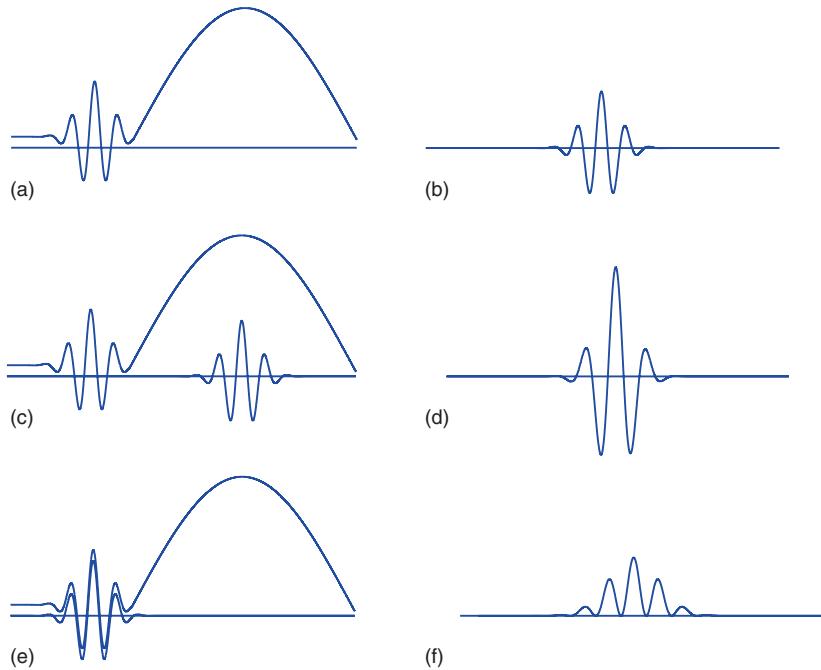


Fig. 5.11 (a) An arbitrary signal, (b) a wavelet, (c, e) signal analyzed by a wavelet, (d, f) product = signal * wavelet

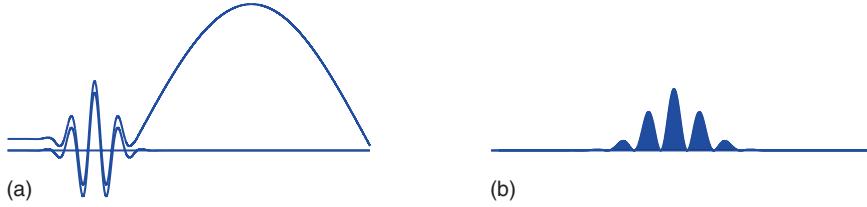


Fig. 5.12 (a) Signal analyzed by a wavelet, (b) *Area = coefficient*

Conversely, the integral of the product will be small or zero, if there is *no similarity*, because the negative values of the integral will be compensated by the positive values of areas. The coefficients of the wavelet applied to the signal, correspond to the surface located between the curve and the axis centered at zero. Note that when the wavelet analyzes the signal which has the same form as the wavelet itself, their product is entirely positive (curves of real parts) and thus the area is also positive and has a high enough value (Fig. 5.12). On the other hand, if the wavelet analyzes a *flat signal*, then the *integral* of the product (signal by wavelet) will be *zero*, because by definition the integral of the wavelet is zero. If the signal offers small curves without strong variation, the integral of the product will provide low coefficients. The displacement of the wavelet along the signal is carried out by means of the parameter (u):

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right). \quad (5.41)$$

Usually, five or six groups of wavelet transforms are used to represent the initial signal, we can also speak of octaves. Each group is decomposed into intermediate transforms, usually six also. Thus, a resolution 6 provides 36 wavelet transforms of the initial signal.

5.7.2 Continuous Wavelet Transform

The continuous wavelet transform of a function or a signal f at the scale s and at the position u is calculated by “correlating” f with a wavelet also called “wavelet atom”:

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt, \quad (a)$$

where ψ^* is the *complex conjugate* of ψ in \mathbb{C} . It is also possible to write this equation as the following convolution product:

$$Wf(u,s) = f \star \bar{\psi}_s(u), \quad (b)$$

where $\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^*\left(-\frac{t}{s}\right)$. The Fourier transform of $\bar{\psi}_s(t)$ is given by:

$$\hat{\bar{\psi}}_s(\omega) = \sqrt{s} \hat{\psi}^*(s\omega) \quad (5.42)$$

with

$$\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (5.43)$$

It appears that $\hat{\psi}$ is the *transfer function* of a *band-pass filter* of frequencies. This convolution computes the wavelet transform with *band-pass filters*. The graphic expression of a continuous wavelet transform is in a plane whose abscissa (u) is the unit of time and ordinate ($\text{Log}_2(s)$) is the frequency scale.

5.7.3 Discrete Wavelet Transform

$$Wf[n, a^j] = \sum_{n=0}^{N-1} f[m] \frac{1}{\sqrt{a^j}} \psi_j^*[m-n] \quad (c)$$

with a wavelet:

$$\psi_j[n] = \frac{1}{\sqrt{a^j}} \psi\left(\frac{n}{a^j}\right). \quad (5.44)$$

The discrete transformation can also be written as a *circular convolution* like in the continuous case:

$$Wf[n, a^j] = f \circledast \bar{\psi}_j[n] \quad (d)$$

with

$$\bar{\psi}_j[n] = \psi_j^*[-n] \quad (5.45)$$

this circular convolution is calculated with the Fast Fourier Transform which requires $O(N \log 2N)$ operations.¹⁹ If

$$a = 2^{\frac{1}{v}} \quad (5.46)$$

there are $v \log 2(N/2K)$ scales $a^j \in [2N^{-1}, K^{-1}]$. The number of operations necessary to compute the transformation is $O(vN(\log 2N)^2)$.

5.7.4 Wavelet Models: “Gauss Pseudo-Wavelet”, Gauss-Derivative, Morlet and Sombrero

5.7.4.1 The Case of “Gauss Pseudo-Wavelet” or “Gauss Window”

Construction of a “Gauss pseudo-wavelet” (constructed by Gabor):

$$g(t) = e^{-\omega_0^2 t^2 / 2}. \quad (5.47)$$

¹⁹ $f[n] = O(g[n])$, there exists K such that $(f[n] \leq Kg[n])$. $(f[n]) = o(g[n])$, $\lim_{n \rightarrow +\infty} \frac{f[n]}{g[n]} = 0$.

Fig. 5.13 Gauss pseudo-wavelet, also called Gauss window

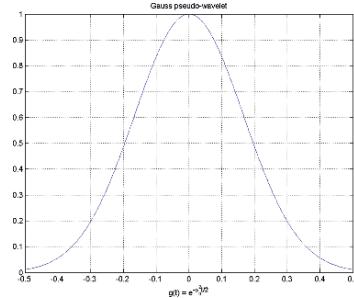
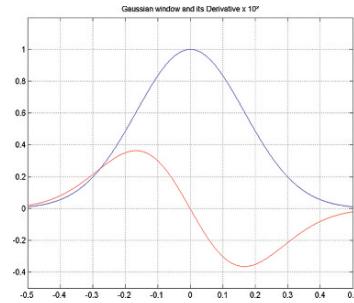


Fig. 5.14 Gauss window and its derivative $\times 10^2$



A *Gauss pseudo-wavelet* is an *isolated structure* in the set of wavelet forms which were constructed subsequently (Fig. 5.13). Indeed, its bell shape does not offer many oscillations around the x -axis (abscissa). It is an almost rudimentary form which offers a weak adaptation to signals with complex structures. Certain works still evoke this structure as a wavelet, however we will prefer to consider it as a *Gauss window* or as a *filter*. Indeed, it does not satisfy all the conditions necessary to the wavelet construction.

A Particularity of the Gauss Window and Gauss-Derivative Wavelet

It will be noted, first, that the derivative of a Gauss window can be regarded as a wavelet (the integral of this derivative is zero). The Gauss window and its derivative are represented in Fig. 5.14.

Let us consider the equivalent of the differential of an arbitrary signal denoted $f'(x)$ and a Gauss window $g(x)$ (i.e. “Gauss pseudo-wavelet”), then we can write by carrying out an *integration by parts*:²⁰

$$\int f'(x)g(x)dx = - \int f(x)g'(x)dx + [\cdot], \quad (5.48)$$

where $[\cdot] = 0$. Thus:

$$\int f'(x)g(x)dx = - \int f(x)g'(x)dx. \quad (5.49)$$

²⁰ Integration by parts: $\int f'g = [fg] - \int fg'$.

Applied to a financial signal (French stock index: Cac40), thus:

$$\int cac'(x)g(x)dx = - \int cac(x)g'(x)dx. \quad (5.50)$$

This particularity is interesting because this shows that the Gauss pseudo-wavelet transform of the first-differences of the Cac40 is equivalent to the Gauss-derivative wavelet transform of the Cac40 itself (up to a sign). And their representations in the time–frequency plane are similar.

5.7.4.2 Construction of a Morlet Wavelet

A Morlet wavelet (Fig. 5.15) is written:

$$\psi(t) = e^{-i\omega_0 t} e^{-t^2/2}. \quad (5.51)$$

5.7.4.3 Construction of a Sombrero Wavelet

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3}\sigma} \left(\frac{t^2}{\sigma^2} - 1 \right) e^{-t^2/2\sigma^2} \quad (5.52)$$

(Fig. 5.16).

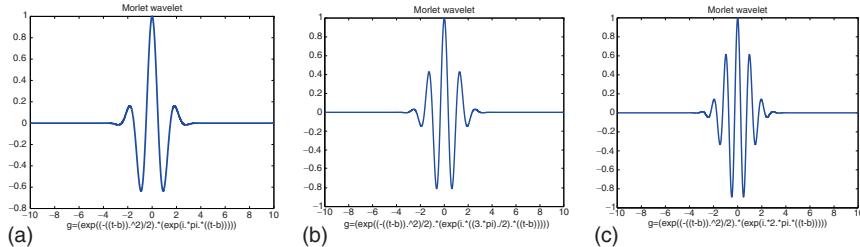


Fig. 5.15 (a) $\omega = \pi$, (b) $\omega = \frac{3\pi}{2}$, (c) $\omega = 2\pi$

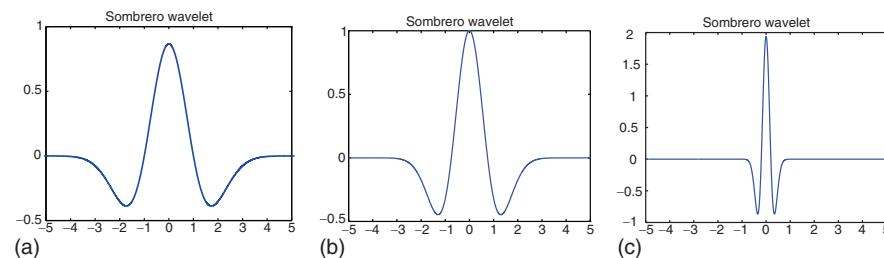
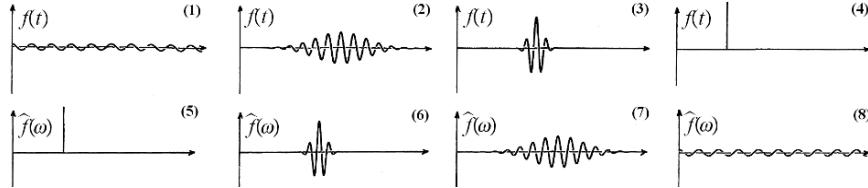


Fig. 5.16 (a) $\sigma = 1$, (b) $\sigma = 3/4$, (c) $\sigma = 1/10$

5.8 Aliasing and Sampling

In the figure below,²¹ we show at the top a few “wave functions” $f(t)$ in the space of positions and at the bottom the corresponding Fourier transforms in the space of impulse responses. The graphs (2),(3),(6),(7) shows wave packets.



A signal f is discretized by recording its sample values $\{f(nT)\}_{n \in \mathbb{Z}}$ at interval T . A discrete signal can be represented as a sum of Diracs. Any sample $f(nT)$ is associated with a Dirac $f(nT)\delta(t - nT)$ located at $t = nT$. So, a uniform sampling of f corresponds to the weighted Dirac sum $f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\delta(t - nT)$. The Fourier transform of $\delta(t - nT)$ is $e^{-inT\omega}$. Thus the Fourier transform of f_d is a Fourier series $\hat{f}_d(\omega) = \sum_{n=-\infty}^{+\infty} f(nT)e^{-inT\omega}$. \hat{f} and \hat{f}_d are related by $\hat{f}_d(\omega) = (1/T) \sum_{k=-\infty}^{+\infty} \hat{f}(\omega - (2k\pi/T))$. The sampling theorem is given by:

Theorem 5.1 (Shannon, Whittaker). *If the support of the Fourier transform \hat{f} is included in $[-\pi/T, \pi/T]$ then $f(t) = \sum_{n=-\infty}^{+\infty} f(nT)\vartheta_T(t - nT)$ with $\vartheta_T(t) = \sin(\pi t/T)/(\pi t/T)$.*

This theorem imposes that the support of \hat{f} is included in $[-\pi/T, \pi/T]$ which prevents abrupt variations of f between consecutive samples. However the sampling interval is generally imposed and the support of \hat{f} is not included in $[-\pi/T, \pi/T]$. So the formula of the theorem does not recover f . A filtering technique to reduce error is then used, i.e. *aliasing*.²² Consider a support of \hat{f} going beyond $[-\pi/T, \pi/T]$, usually the support of $\hat{f}(\omega - (2k\pi/T))$ intersects $[-\pi/T, \pi/T]$ for several k , see infra Fig (ii). When there is aliasing, the interpolated signal $\vartheta_T * f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\vartheta_T(t - nT)$ possesses a Fourier transform $\hat{f}_d(\omega)\hat{\vartheta}_T(\omega) = T\hat{f}_d(\omega)\mathbf{1}_{[-\pi/T, \pi/T]}(\omega) = \mathbf{1}_{[-\pi/T, \pi/T]}(\omega)\sum_{k=-\infty}^{+\infty} \hat{f}(\omega - (2k\pi/T))$ which can be very different from $\hat{f}(\omega)$ over $[-\pi/T, \pi/T]$. The signal $\vartheta_T * f_d(t)$ can be an inappropriate approximation of f , see infra (†) in Fig. (ii). In Fig. (i): (a) shows a signal with its Fourier transform; (b) a uniform sampling of the signal generates a *periodic* Fourier transform; (c) low-pass filter; (d) the filtering of (b) with (c) recovers f . In Fig. (ii): (a)' shows a signal with its Fourier transform; (b)' aliasing generated by an overlapping of $\hat{f}(-2k\pi/T)$ for different k , see dashed lines; (c)' low-pass filter; (d)' the filtering of (b)' with (c)' generates a low-frequency signal different from f , see (†).

²¹ Remark: The Heisenberg uncertainty relations are illustrated by the fact that a large uncertainty in position is associated with a low uncertainty in impulse response and vice versa.

²² **Definition (Aliasing).** Introduction of error into the computed amplitudes of the lower frequencies in a Fourier analysis of a function carried out using discrete time samplings whose interval does not allow the proper analysis of the higher frequencies present in the analyzed function.

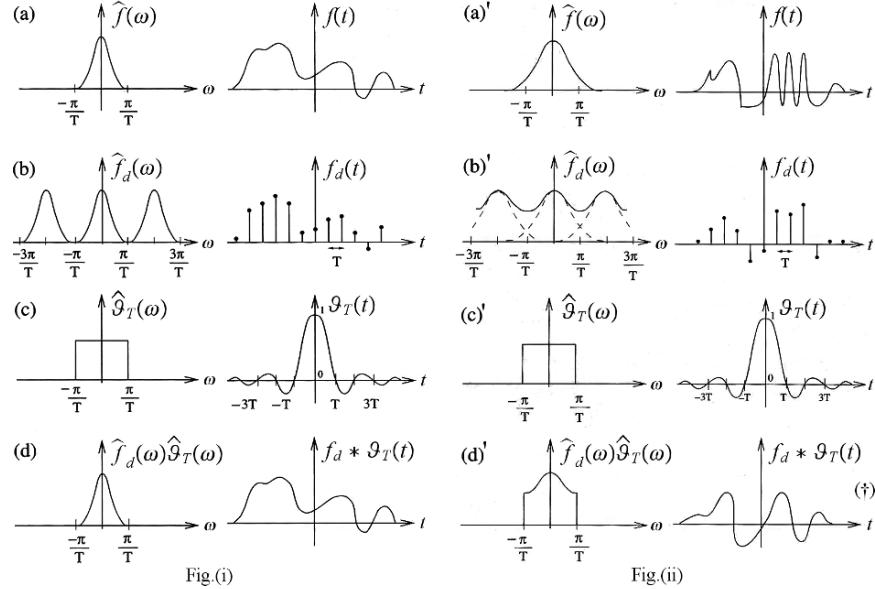


Fig.(i)

Fig.(ii)

5.9 Time-Scale Plane (b,a), Cone of Influence

5.9.1 Cone of Influence and Time-Scale Plane

Let $f(x)$ be a function and $\psi(x)$ a wavelet which occupies the space: $I_\psi = [-X, X]$. The behavior of the function $f(x)$ in a precise point, x , is mainly observed on the wavelet transform in the cone $b \in a I_\psi + x = [-aX + x, aX + x]$ (Fig. 5.17).

$\psi(x)$ is a mother wavelet and $\psi_{(a,b)}(x)$ is a wavelet family created by the changes of the parameters a, b . As described in previous sections, “ b ” represents the position in time of the wavelet, i.e. its position along the analyzed signal. “ a ” represents the

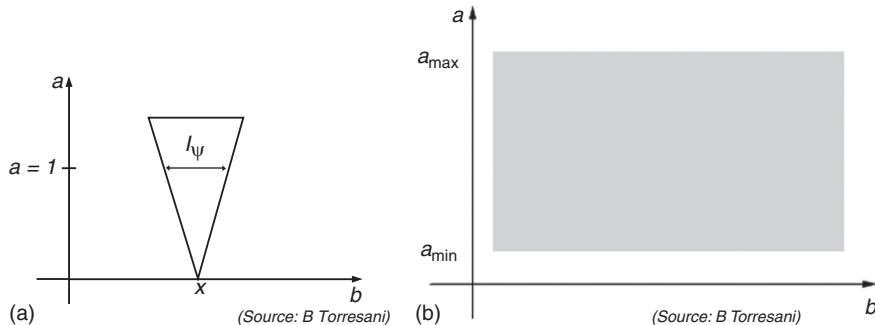


Fig. 5.17 (a) Cone of influence of a point x of b . (b) (b,a) -plane

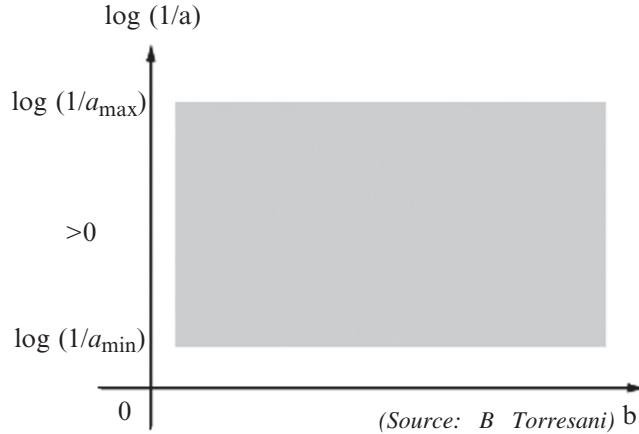


Fig. 5.18 Phase-space (hyperbolic plane)

dilation and contraction of the wavelet. The wavelet transform $T_f(b,a)$ is represented in the *hyperbolic space* of transforms (Fig. 5.18) (*Time-frequency plane = Phase space*). We mentioned it in the preceding section, in certain types of construction, when the scale parameter “ a ” increases, the length of the “central frequency” of the wave tends to increase. This is an analysis which is valid for example within the framework of the continuous wavelet transform. Thus, in the plane which precedes, the coarsest or less fine “long waves” calculated by the transformation, will be in the upper part of the plane. On the contrary, the finest wavelets, which analyze with more precision the signal, will be represented in the lower part of the plane. Hyperbolic planes such as these allow a form of symmetry around the abscissa (x -axis). We often use the scale “log 2”, which offers the property to represent the transforms of a signal in the positive plane above the abscissa, if the scale parameter is taken as a divider, i.e. $\log 2(1/a)$. In the opposite case, the image of transforms will be represented in the plane below the abscissa. The plane itself will obviously have a minimum and a maximum that determine thus the limits of the image (Fig. 5.19).

Note that if we select a scale parameter as divider of transforms, the ordinates of the plane will be positive, otherwise, the ordinates will be negative. The representation in negative ordinates is rather frequent (see Torrésani or Mallat). If the signal which is the subject of a transformation has a dyadic length, we will use *log 2* rather than *log*. We cannot numerically calculate a wavelet transform for all the values of b and a . We sample the wavelet transform. The extreme values of the scale are computed starting from the following extrema, as follows:

$$a_{\max} = \frac{\text{Size of the signal}}{\text{Size of the wavelet}},$$

$$a_{\min} = \frac{\text{Central frequency of the wavelet}}{\text{Sampling frequency of the signal}}.$$

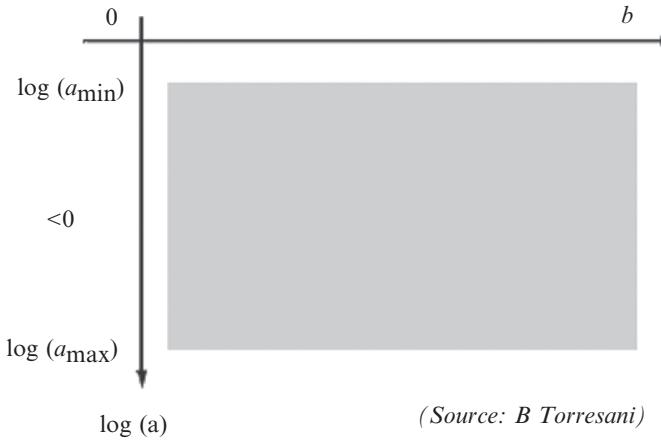


Fig. 5.19 Hyperbolic plane, logarithmic scale

5.9.2 Time–Frequency Plane

The time–frequency plane is used in particular in the transformations with sliding window of the Gabor type, or in the wavelet multiresolution programs. The Mallat and Zhang “Pursuit” algorithm uses also the time–frequency plane. The general principle: Since the wavelet and the window are localized along the signal, it is possible – by using segments which are sampled and localized in time – to make correspond to these segments a frequency which will occupy by projection a segment on the ordinate axis. Thus, we construct “time–frequency” rectangles or “time–frequency” squares localized such as “tiles” in the plane. This process is close to the principle of Heisenberg boxes described in the section which follows. The construction of such time–frequency planes requires to constitute a grid intended to receive these time–frequency rectangles: $(\Delta t, \Delta \omega)$ or $(\sigma_t, \sigma_\omega)$. The choice of the type of grid can be large enough, we can speak of the Fourier grid or Gabor grid. Then, we establish a scale corresponding to frequency levels, for example: gray nuances. Finally, we assign to each “time–frequency tile” a nuance of gray corresponding to its eigen frequency. There are other types of construction of time–frequency plane, in particular the plane of the Wigner–Ville distribution, the plane of the energy distribution, or the spectrograms, for which representations are more diffuse and segmented, just as for the representation of the energy distribution resulting from the “Pursuit” algorithm.

5.10 Heisenberg Boxes and Time–Frequency Plane

Briefly, we introduce the Heisenberg theorem which leads to the construction of Heisenberg boxes, which then allows to underline the two different approaches between the wavelets and “Fourier”, and this by using *the bivalent concept of*

time–frequency atom which makes it possible to unify, but also to have in hand equivalences from a field to another of the analysis.

Theorem 5.2 (Heisenberg uncertainty). *The temporal variance σ_t^2 and the frequency variance σ_ω^2 of the function f (which belongs to the L^2 space), verify:*

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4} \quad (5.53)$$

or, in a similar way by means of the standard deviations $\sigma_t \sigma_\omega \geq \frac{1}{2}$. The inequality becomes an equality if and only if there exist $(a, b, u, \xi) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that $f(t) = a \exp(i\xi t - b(t-u)^2)$.

This observation leads to a kind of compromise between the *temporal* resolution and the *freqential* resolution. *The localization in time–frequency can be reached only in standard deviation.* This localization is representable by means of a Heisenberg box. The cover of the time–frequency plane is carried out through different manners in accordance with the Fourier atoms or wavelet atoms.

5.10.1 Concept of Time–Frequency Atom: Concept of Waveform Family

We know that the transformation of an unspecified signal in the time–frequency plane can be carried out either by sines and cosines or by wavelets, thus, we can indicate in a general way these forms which enter into the mechanics of transformation as *waveforms*; a *family of waveforms* is also called *time–frequency atoms* and is written: $\{\phi_\gamma\}$ where γ is a vector of parameters. Generally, it is agreed that ϕ_γ such that $\|\phi_\gamma\| = 1$.²³ The transformation of a signal $f(t)$ by this family of atoms ϕ_γ , can be written:²⁴

$$Tf = \int_{-\infty}^{+\infty} f(t) \phi_\gamma^*(t) dt = \langle f, \phi_\gamma \rangle. \quad (5.54)$$

For a *windowed Fourier transform*, a *Fourier atom* by translation and modulation is written:

$$\phi_\gamma(t) = g_{u,\xi}(t) = g(t-u) e^{i\xi(t-u)}. \quad (5.55)$$

For a *wavelet transform*, a *wavelet atom* by translation and modulation are written:

$$\phi_\gamma(t) = \psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right). \quad (5.56)$$

²³ ϕ_γ : belongs to the integrable functions $\mathbf{L}^2(\mathbb{R})$.

²⁴ By the well-known Parseval formula we know that:

$$Tf = \int_{-\infty}^{+\infty} f(t) \phi_\gamma^*(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \widehat{f}(\omega) \widehat{\phi}_\gamma^*(\omega) d\omega.$$

5.10.2 Energy Density, Probability Distribution and Heisenberg Boxes

After having defined the concept of time–frequency atom, we present the concept of density of energy (Mallat 1998). These notions are increasingly diffused out of the Physics where they come from.

5.10.2.1 Energy Density and Probability Distribution of a Time–Frequency Atom

For any parameter of translation and modulation (u, ξ) , we suppose that there is a single atom $\phi_{\gamma(u, \xi)}$ which is centered at (u, ξ) in the time–frequency plane. The time–frequency boxes of this atom determines a neighborhood of (u, ξ) where the energy of f is measured by

$$P_T f(u, \xi) = |\langle f, \phi_{\gamma(u, \xi)} \rangle|^2 = \left| \int_{-\infty}^{+\infty} f(t) \phi_{\gamma(f(u, \xi))}^*(t) dt \right|^2. \quad (5.57)$$

The generic transformation $Tf = \langle f, \phi_\gamma \rangle$ in the time–frequency plane (t, ω) of a signal depends on the nature of the atom ϕ_γ and depends on its *time–frequency “spread”*, i.e. as well in time as in frequency. Thus, one attempts to approach this notion of *spread* of a time–frequency atom. We know that:

$$\|\phi_\gamma\|^2 = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 dt = 1, \quad (5.58)$$

the element which is the subject of the integration, i.e. the square of the absolute value of the atom, is understood as a *probability distribution* centered at u_γ , whose *center* (or average) is:

$$u_\gamma = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 \cdot t dt. \quad (5.59)$$

And its *variance* measures its *spread around u*:

$$\sigma_t^2(\gamma) = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 (t - u_\gamma)^2 dt. \quad (5.60)$$

So that we can symbolically write that the atom follows a *probability law*:

$$\phi_\gamma \sim Law(u_\gamma, \sigma_t^2(\gamma)). \quad (5.61)$$

If we are located in the space of frequencies (of the trigonometrical functions or wavelets), we can wonder what becomes the center, the variance and the spread of this atom which (it) is expressed simultaneously in time and frequency. It is by using the Plancherel transfer formula, which *provides an equivalence in both fields frequency and time, that we can determine the center frequency of the atom and*

its spread in the frequency (frequency support). The Plancherel formula, which is written $\int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 dt = 1/2\pi \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 d\omega$, provides this equivalence between both fields, and applied to ϕ_γ , this gives:

$$\int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 d\omega. \quad (5.62)$$

In addition, we know that for an unspecified²⁵ function f we have $\|f\|^2 = \langle f, f \rangle = \int_{-\infty}^{+\infty} |f(t)|^2 dt$, consequently for the atom ϕ_γ , there is a similar result: $\|\phi_\gamma\|^2 = \langle \phi_\gamma, \phi_\gamma \rangle = \int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 dt$, thus the Plancherel formula for the atom ϕ_γ is written:

$$\int_{-\infty}^{+\infty} |\phi_\gamma(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 d\omega, \quad (5.63)$$

this can also be written:

$$2\pi \|\phi_\gamma\|^2 = \int_{-\infty}^{+\infty} |\hat{\phi}_\gamma(\omega)|^2 d\omega. \quad (5.64)$$

Thus, endowed with *these equivalences*, we can determine the *center frequency* of the atom and *its “frequency spread” (measured by the variance)*. The *center frequency* ξ_γ of $\hat{\phi}_\gamma$ is written:

$$\xi_\gamma = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega |\hat{\phi}_\gamma(\omega)|^2 d\omega, \quad (5.65)$$

then the *spread* around the center frequency ξ_γ is written:

$$\sigma_\omega^2(\gamma) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega - \xi_\gamma)^2 |\hat{\phi}_\gamma(\omega)|^2 d\omega. \quad (5.66)$$

We are thus endowed with the necessary tools for a representation in the time–frequency plane of a basic atom ϕ_γ by means of the two following pairs (u_γ, σ_t^2) and $(\xi_\gamma, \sigma_\omega^2)$, i.e. respectively the temporal center and spread, then the center frequency and spread in the frequency. The time–frequency resolution is depicted in the time–frequency plane (t, ω) by means of the Heisenberg boxes centered at (u_γ, ξ_γ) of which the width along time is $\sigma_t(\gamma)$ and the width along frequency is $\sigma_\omega(\gamma)$. Moreover, we note that $\sigma_t \sigma_\omega$ is the area of a Heisenberg box. *By referring to the Heisenberg uncertainty theorem* which explains that the temporal variance σ_t^2 and the frequency variance σ_ω^2 of the function f , verify:

$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4}, \quad (5.67)$$

²⁵ Even if it is a *discontinuous* function f .

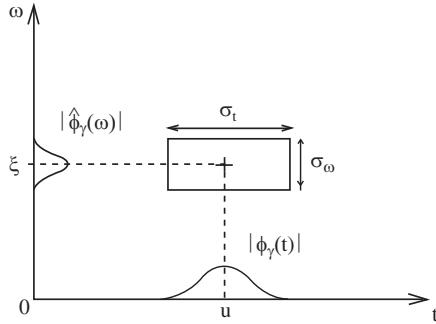


Fig. 5.20 Heisenberg Boxes representing an atom ϕ_γ

or, in an equivalent way: $\sigma_t \sigma_\omega \geq 1/2$, this *constraint* on the variances says that the *surface of the rectangle representing the atom is higher or equal to 1/2*. Thus, we do not represent the atom by a point but by rectangles, we will be able to refer to the figure which follows in the next section.

5.10.2.2 Heisenberg Boxes of Time–Frequency Atoms: Fourier Atoms and Wavelet Atoms

Generic presentation of the Heisenberg box: One represents the time–frequency localization of a basic atom (time–frequency atom) by means of a Heisenberg box, located in the time frequency plane, which is a rectangle of dimensions σ_t and σ_ω , centered at the point of coordinates: (temporal center, center frequency), as described in the previous section (Fig. 5.20).

Presentation of the Heisenberg box for a windowed Fourier atom: Remember that the principle of the windowed Fourier transform can be defined as follows:

$$Sf(u, \xi) = \langle f, g_{u, \xi} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{i\xi(t-u)}dt. \quad (5.68)$$

The *atom* used is a *sinusoid multiplied by a window g*. The analysis vector family is obtained by translation and modulation of the window as follows: $g_{u, \xi}(t) = g(t-u)e^{i\xi(t-u)}$. This function is centered for the frequencies at ξ and is symmetric in relation to u . The standard-deviation in frequency is constant. The family is thus obtained by translation in time and frequency of a single window. Here is an example of Heisenberg boxes of windowed Fourier atoms (Fig. 5.21).

Presentation of the Heisenberg box for the “wavelet atoms”: Let us recall the principle of the wavelet transform which can be defined as follows:

$$Wf(u, s) = \langle f, \psi_{u, s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt, \quad (5.69)$$

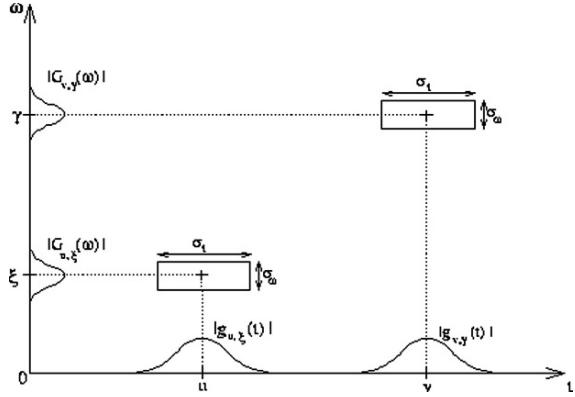


Fig. 5.21 Heisenberg rectangles, or time–frequency boxes, symbolizing the *energy spread* of two Gabor atoms

where the basic atom, or the basic wavelet, ψ is a function of zero average, centered at the neighborhood of 0 and of a finite energy. The family of vectors is obtained by translation and dilation of the atom $\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$. This function is centered at the neighborhood u , as the windowed Fourier atom. If the center frequency of ψ is indicated by η , the center frequency of the dilated function is at η/s . The temporal standard-deviation is proportional to s . The standard-deviation in frequency is inversely proportional to s . Thus, one can present an example of Heisenberg boxes of wavelet atoms (Fig. 5.22).

5.10.3 Spectrogram, Scalogram and Energy Conservation

The square of the modulus of the windowed Fourier transform described in the preceding section corresponds to the spectrogram:

$$P_S f(u, \xi) = |Sf(u, \xi)|^2 = \left| \int_{-\infty}^{+\infty} f(t) g(t-u) e^{i\xi(t-u)} dt \right|^2. \quad (5.70)$$

In order to calculate the *scalogram*, which is the equivalent for the *wavelets* of the *spectrogram* for the *Fourier transform*, one takes for example η as the center frequency of the wavelet, then the center frequency of a dilated wavelet is $\xi = \eta/s$.

The scalogram is written then:

$$P_W f(u, \xi) = |Wf(u, s)|^2 = \left| Wf \left(u, \frac{\eta}{\xi} \right) \right|^2. \quad (5.71)$$

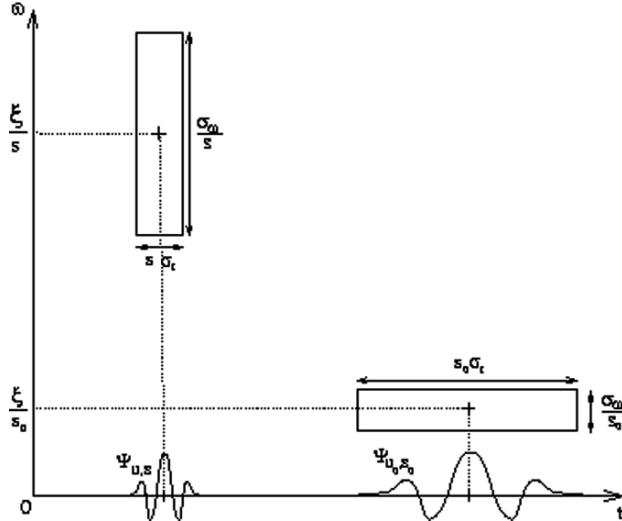


Fig. 5.22 Example for two wavelets $\psi_{u,s}(t)$ and $\psi_{u_0,s_0}(t)$. When the scale s decreases, the time-segment is reduced but the “spread” of the frequency increases

The *scalogram* is also called *energy density*. This interpretation of the scalogram as energy density is shown by the following theorem, which shows that the energy is preserved during the transformation of a real unspecified signal f :

Theorem 5.3 (The scalogram is an energy density).²⁶ For any $f \in L^2(\mathbb{R})$

$$Wf(u,s) = \frac{1}{2} Wf_a(u,s) \quad (5.72)$$

if the admissibility condition of the wavelet written:

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|}{\omega} d\omega < +\infty \quad (5.73)$$

is verified and if f is real, then:

$$f(t) = \frac{2}{C_\psi} \operatorname{real} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s) \psi_s(t-u) du \frac{ds}{s^2} \right] \quad (5.74)$$

and

$$\|f(t)\|^2 = \frac{2}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(u,s)|^2 du \frac{ds}{s^2}. \quad (5.75)$$

²⁶ Proof in Mallat (1998).

5.10.4 Reconstruction Formulas of Signal: Stable and Complete Representations

The objective of this section is to point out the reconstruction formulas of a signal, which result from the Fourier transform with window and from the wavelet transform.

5.10.4.1 Reconstruction by the Windowed Fourier Transform

For the Fourier atoms $g_{u,\xi}(t)$ when the couple of indexes (u, ξ) varies on \mathbb{R}^2 , the Heisenberg boxes of the atoms cover the whole of the time–frequency plane. This assertion has a corollary, indeed it is possible to rebuild the signal by means of its windowed Fourier transform $Sf(u, \xi)$. The *reconstruction formula* of the signal is written:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u, \xi) g(t-u) e^{it\xi} d\xi du, \quad (5.76)$$

with $\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u, \xi)|^2 g(t-u) e^{it\xi} d\xi du$. Moreover, the *reconstruction formula preserves identical the energy of the signal*. The reconstruction formula can also be rewritten as follows:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \langle f, g_{u,\xi} \rangle g_{u,\xi}(t) d\xi du. \quad (5.77)$$

5.10.4.2 Reconstruction by the Wavelet Transform

Similarly, it is possible to rebuild the signal from the wavelet transform. The (real) wavelet transform is complete and preserves the quantity of energy (as long as the wavelet verifies the weak admissibility condition written: $C_\psi = \int_0^{+\infty} (|\widehat{\psi}(\omega)| / \omega) d\omega < +\infty$). The *reconstruction formula is given by*:

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u, s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) du \frac{ds}{s^2}, \quad (5.78)$$

with

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(u, s)|^2 du \frac{ds}{s^2}. \quad (5.79)$$

5.11 Wiener Theory and Time–Frequency Analysis

5.11.1 Introduction to the Correlogram–Periodogram Duality: Similarities and Resemblances Researches

The correlogram and periodogram are the base of spectral analysis. Spectral analysis can be interpreted as the decomposition of the variance of a time series into the field of frequencies. It is commonly said that the function of the temporal correlation corresponds to the Fourier transform of the power spectrum of a time series. The purpose of spectral analysis is to identify similarities and resemblances of a signal (on itself). One will choose here in this preamble a simplified presentation of these periodogram and correlogram concepts, which are also described within the framework of the Wiener–Khintchine theorem. Consider a process $\{x_t\}$ with a temporal index $t = 1, \dots, T$. If one chooses a stochastic process weakly stationary, it will satisfy the conditions:

$$E|x_t|^2 < \infty, \quad \forall t, \quad (5.80)$$

$$Ex_t = \mu, \quad \forall t, \quad (5.81)$$

$$E(x_t - \mu)(x_s - \mu) = E(x_{t+\tau} - \mu)(x_{s+\tau} - \mu), \quad \forall s, t, \tau. \quad (5.82)$$

5.11.1.1 Autocovariance and Autocorrelation Functions

Definition 5.1 (Autocovariance). The *autocovariance function* at the given delay τ of a process x_t is defined for $\tau \in \{0, \dots, T-1\}$ such that:

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=\tau+1}^T (x_t - \bar{x})(x_{t-\tau} - \bar{x}), \quad (5.83)$$

where $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ is an arithmetic mean of x_t . The *autocorrelations* $\hat{\rho}(\tau)$ are defined by standardizing the autocovariance function $\hat{\gamma}(\tau)$ by $\hat{\gamma}(0)$ (i.e. the variance of the process):

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}. \quad (5.84)$$

(The confidence interval on the correlograms is written $\pm 2/\sqrt{T}$.) Another measure often used is *the partial autocorrelation*. A partial autocorrelation with order $\tau \geq 2$ is calculated as a correlation of two residuals obtained after regression of $x_{\tau+1}$ and x_1 on the intermediate observations x_2, \dots, x_τ . The partial autocorrelation at the given delay equal to 1 is defined as the correlation between x_1 and x_2 .

5.11.1.2 Estimator of the Spectral Density

The periodogram is the equivalent, in the field of frequencies, to the autocorrelation function in the field of time. The analysis of the frequency field is carried out by the decomposition of the observed series into periodic components. The principal tool for the spectral analysis of a series is the spectrum which is defined by:

$$f(\lambda) = \frac{1}{2\pi} \left[\gamma(0) + 2 \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(\lambda \tau) \right], \quad (5.85)$$

with λ the angular frequencies $[-\pi, \pi]$ and $\gamma(\tau)$ the theoretical autocovariances ($\tau = 1, \dots, \infty$). The spectrum is symmetric around zero, the analysis is restricted at the frequencies $[0, \pi]$. For a sample of T observations, one considers the harmonic frequencies, or Fourier frequencies $\lambda_j = 2\pi j/T$, $j = 1, \dots, [T/2]$. The periodogram, by using the spectrum, is defined by:

$$I(\lambda) = \frac{1}{2\pi} \left[\hat{\gamma}(0) + 2 \sum_{\tau=1}^{T-1} \hat{\gamma}(\tau) \cos(\lambda \tau) \right]. \quad (5.86)$$

An estimator is said to be “consistent” if its variance tends towards zero when the number of observations tends towards the infinite. The consistency implies that the estimator becomes gradually more precise when the quantity of acquired information increases. If one chooses a stationary random process of order two, it can be shown that the estimator obtained is centered but does not converge. It is appropriate only for the time series which have strict periodicities and its graph (periodogram is equivalent to the autocorrelation function in the field of time) has an “abrupt” shape which makes it difficult to interpret it. In order to solve this problem of readability of the graph, spectral windows are used. Indeed, in order to solve this readability problem of graphs, we do not estimate for each frequency the value $I(\lambda)$, but an average value on equal frequency bands, whose juxtaposition covers the interval given by λ . In fact, we carry out a smoothing of the spectrum, i.e. a kind of filtering. We take, around a number of equidistant points of the frequency axis, a weighted average on the “neighboring frequencies”, i.e. one builds a window in the space of the frequency. The choice of the window is important. According to its form, i.e. according to the chosen weighting factors, spectral “leaks” at the adjacent frequencies for a given frequency are possible. For example, if at a frequency λ_j of the spectrum of a time series there exists a “peak” with a high spectral power, in relation to the adjacent frequencies, the secondary lobes inherent to the spectral window, can generate, at their frequencies, significant spectral powers. It is thus difficult to choose a good spectral window, because it is necessary that the power of secondary lobes is low and that the power at the adjacent frequencies are not correlated. There are two types of window often used in spectral analysis, they are the Tukey–Hanning and Parzen windows. For the first, it is said that the secondary peaks are lower or

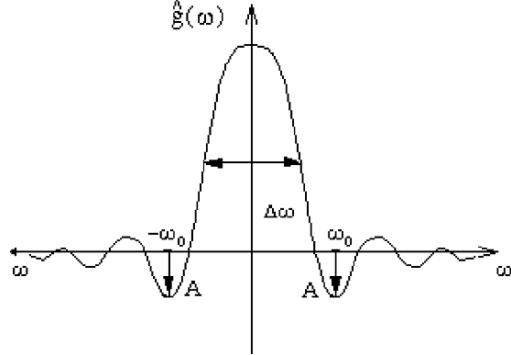


Fig. 5.23 Plane (x -axis: frequency; y -axis: window function \hat{g}): The energy spread of \hat{g} measured by its bandwidth $\Delta\omega$ and the maximum amplitude A of the first side-lobes (localized at $\omega = \pm\omega_0$)

equal to 2% of the main peak and the spectral estimators at the frequencies λ_j and λ_{j+2} are not correlated.²⁷

Construction of a Fourier Windows

Let g be the window in the (windowed) Fourier transform, whose energy is concentrated around 0 (Fig. 5.23). There are three parameters which determine the energy spread:

- (1) The root mean square bandwidth $\Delta\omega$ written as follows:

$$\frac{|\hat{g}(\Delta\omega/2)|^2}{|\hat{g}(0)|^2} = \frac{1}{2}. \quad (5.87)$$

If $\Delta\omega$ is small, the energy of the window is well concentrated around 0.

- (2) The maximum amplitude A of its first side-lobes, measured in decibels:

$$A = 10 \log_{10} \frac{|\hat{g}(\omega_0)|^2}{|\hat{g}(0)|^2}. \quad (5.88)$$

It is possible that these side-lobes create *shadows* on each side of the center frequency.

- (3) The polynomial exponent which describes the decay of the window of the Fourier transform for broad frequencies:

$$|\hat{g}(\omega_0)| = O(\omega^{-p-1}) \quad (5.89)$$

It represents the behavior of the Fourier transform beyond the first side-lobes.

²⁷ For a discuss about the subject one will be able to refer to Bourbonnais and Terraza (1998).

Traditional Spectral Windows: Rectangular, Gaussian, Hamming, Hanning, Blackman

As presented previously the Heisenberg uncertainty theorem imposes that the standard-deviation in time σ_t and in frequency σ_ω of a function verify: $\sigma_t \sigma_\omega \geq 1/2$. This observation leads to a compromise between the temporal resolution and the frequential resolution. The localization in time–frequency can be reached only in standard deviation (or variance). This localization is representable by the Heisenberg box. However, it will be noted that *the limit $\sigma_t \sigma_\omega = 1/2$ is reached only if the window is Gaussian*. The Gaussian, Hamming, Hanning, Blackman windows obviously have very close structures, but nevertheless different. The following table gives the values of parameters which were presented in the previous section for traditional windows, normalized so that $g(0) = 1$.

Window	$g(t)$	$\Delta\omega$	A_{\max} (dB)
Rectangular	$g_{rec}(t) = 1,$	0.89	-13
Gaussian	$g_g(x) = \exp(-18t^2),$	1.55	-55
Hamming	$g_{hm}(t) = 0.54 + 0.46 \cdot \cos(2\pi t),$	1.36	-43
Hanning	$g_{hn}(x) = \cos^2(\pi t),$	1.44	-32
Blackman	$g_b(x) = 0.42 + 0.5 \cdot \cos(2\pi t) + 0.08 \cdot \cos(4\pi t).$	1.68	-58

A_{\max} is the maximum amplitude measured in decibels (dB) and $\Delta\omega$ the bandwidth. Hereafter their graphs (the rectangular window is not represented) (Fig. 5.24).

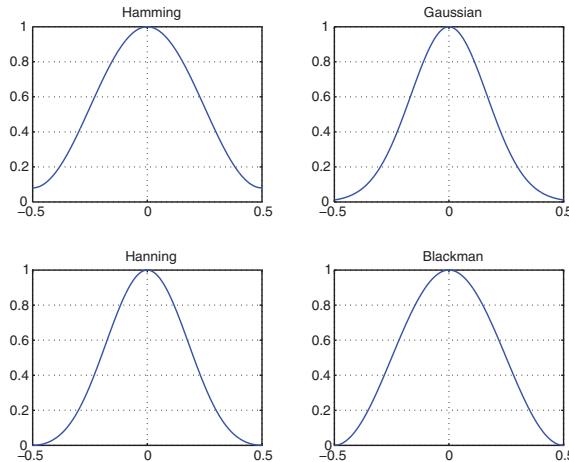


Fig. 5.24 Four windows on the support: $[-0.5, 0.5]$

5.11.2 Elements of Wiener Spectral Theory and Extensions

In this section we consider the time–frequency analysis for signals of dimension one and we present basic tools. These tools are particularly interesting for the specialists who work on random or noized (disturbed) signals with possible non-stationary characteristics. These subjects can also concern the statisticians who usually are rarely involved in the spectral time series analysis. The utilization of *wavelet bases* is increasingly widespread and many applications in particular in engineering have tried to exploit the *continuous Gabor transforms* or the *wavelet transforms* with *statistical goals*. It could be a question of *detecting*, *denoizing* or *reconstruction* of signals and more important *to carry out the spectral analysis of non-stationary signals*. It seems that the statisticians did not share yet the benefit of this type of work.

The *traditional spectral analysis of stationary processes (random or deterministic)* constitutes the hard core of the analysis. But, a very particular interest can be granted to the *sampling of stationary or non-stationary continuous signals* within the framework of time–frequency representations. One of the important questions is to apprehend the time–frequency representations resulting from the *Gabor transformations* and wavelet transform. Their potential is vast but still requires an academic validation.

A synthetic work has been presented in 1998 by R. Carmona, W.L. Hwang and B. Torrésani (Carmona et al. 1998). This work tries to establish the link between the contribution of this type of analysis and their statistical interest. Moreover, the authors believe in the capacity of these methods to provide tools with advantages and varied talents for the spectral analysis of non-stationary signals. Unfortunately, the corresponding statistical theory is not completely developed yet. The authors *revisit* in their works the traditional elements of the spectral theory of stationary random processes in the light of these new tools.

5.11.2.1 Wiener’s Deterministic (Classical) Spectral Theory

This theory of deterministic signals is a set of mathematical concepts which are presented in the sub-sections which follow. First, one considers continuous signals. Let us consider a function f , one can write the *autocovariance function* $C_f(\tau)$ of f for all τ , as follows:

$$C_f(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(x + \tau) f(x) dx. \quad (5.90)$$

Its value at the origin is written:

$$C_f(0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(x)^2 dx. \quad (5.91)$$

This value at the origin is called the *power* of the signal f , which is finite. The function of autocovariance of a signal with *real values* makes it possible to write

also $C_f(\tau) = C_f(-\tau)$. If one confines oneself with a class of function f with finite power and possibly complex valued, it comes:

$$C_f(0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} |f(x)|^2 dx. \quad (5.92)$$

And if this value exists, then one can write the inner product of two functions:

$$\langle f, g \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(x) \overline{g(x)} dx. \quad (5.93)$$

And the *Schwartz inequality* is written:

$$|C_f(\tau)| \leq C_f(0). \quad (5.94)$$

The autocovariance is defined as non-negative:

$$\sum_{j,k=1}^n z_j \overline{z_k} C_f(x_j - x_k) \geq 0 \quad (5.95)$$

whatsoever the choice of the complex numbers z_j and elements x_j . This equation implies the existence of non-negative finite measurements v_f satisfying:²⁸

$$C_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\tau\omega} v_f d\omega. \quad (5.96)$$

The measure v_f is called the *spectral measure* of the signal f . The spectral analysis of a signal f consists in finding the properties of this measure. In general, when we speak of measurements, we are interested in the significance of non-negative finite measurement. The following stage relates to the *Lebesgue decomposition into three elements of this measurement*, which is written:

$$v_f = v_{f,pp} + v_{f,sc} + v_{f,ac}, \quad (5.97)$$

- $v_{f,pp}$ is a *pure point* measure (i.e. a weighted sum of Dirac point masses).
- $v_{f,sc}$ is a *singular*, i.e. concentrated on a set of Lebesgue measure zero, and *continuous* because $v_{f,sc}(\{\omega\}) = 0$ for any singleton ω measure. Thus $v_{f,sc}$ is a *continuous singular*.
- $v_{f,ac}$ is *absolutely continuous*, i.e. is given by a density with respect to the Lebesgue measure, i.e. $dv_{f,ac}(\omega) = v_{f,ac}(\omega)d\omega$, for some non-negative integrable functions v_f .

In a similar way, the *Lebesgue decomposition* of the spectral measure provides a decomposition of the autocovariance function, such that:

$$C_f = C_{f,pp} + C_{f,sc} + C_{f,ac} \quad (5.98)$$

²⁸ **Theorem (Bochner's theorem).** Among the continuous functions on \mathbb{R}^n , the positive definite functions are those functions which are the Fourier transforms of finite measures.

with

$$C_{f,\dots}(\tau) = \int_{-\infty}^{+\infty} e^{i\tau\omega} v_{f,\dots}(d\tau). \quad (5.99)$$

This *decomposition means that the signal can be understood as the sum of three orthogonal signals*, i.e. *uncorrelated*, with *pure spectra* given by the components of original spectral components in the Lebesgue decomposition (Carmona et al. 1998). The following function $\rho_f(\omega) = v_f(\{\omega\})$ is called the *spectral mass function*, and the density $v_{f,ac}(\omega)$ is called the *spectral density function*. The interpretation of the component $C_{f,sc}$ of the covariance is difficult. However, in practice this component relative to a measure $v_{f,sc}$ is rare and usually is equal to zero: $v_{f,sc} = 0$. The measure $v_{f,pp}$ is written:

$$v_{f,pp} = \sum_k \rho_k \delta_{\omega_k}, \quad (5.100)$$

considering the non-negative weights $\rho_k = \rho_f(\omega_k)$ and the possibly *unit point masses* δ_{ω_k} , then the *pure point* part of the autocovariance function has the following form of a potentially complex trigonometrical polynomial when the number of frequencies ω_k is finite,

$$C_{f,pp}(\tau) = \frac{1}{2\pi} \sum_k \rho_k e^{i\omega_k \tau} \quad (5.101)$$

and it has generally the form of an almost periodic function. It will be noted that: $\frac{1}{2\pi} \sum_k \rho_k = C_{f,pp}(0) < +\infty$.

Case of the Spectrum of an Almost Periodic Function

If one considers an *almost periodic function* of the following form:

$$f(x) = \sum_{j=-\infty}^{+\infty} c_j e^{i\lambda_j x} \quad (5.102)$$

λ_j are distinct *real numbers* and the coefficients c_j are *potentially complex numbers*, satisfying:

$$\sum_{j=-\infty}^{+\infty} |c_j|^2 < \infty. \quad (5.103)$$

One can write:

$$C_f(\tau) = \sum_{j=-\infty}^{+\infty} |c_j|^2 e^{i\lambda_j \tau}, \quad (5.104)$$

this formula expresses that the autocovariance function is a pure point, which means that $C_{f,sc}(\tau) = C_{f,ac}(\tau) = 0$, and the spectrum is concentrated on a set of λ_j . Thus, the *spectral mass function* $\rho_f(\omega)$ can be written:

$$\rho_f(\omega) = \begin{cases} |c_j|^2 & \text{if } \omega = \lambda_j, \\ 0 & \text{otherwise.} \end{cases} \quad (5.105)$$

It is highlighted that there is an information loss of phases. That means that the autocovariance function $C_f(\tau)$ is only a function of the modulus of the coefficient c_j (i.e. $|c_j|^2$) and the arguments of these *complex numbers* cannot be obtained from the knowledge of the spectrum. (Because “these phases values measure the displacements of *harmonic components* relative to a fixed time origin, and they cannot be retained by the spectral analysis because the latter is based on a definition of the autocovariance function which wipes out any natural notion of time origin” [Carmona et al. 1998, p. 44].)

The Wiener Theory is Extended to Random Signals

The Wiener theory is extended to random signals if the limit defining the autocovariance function $C_f(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(x + \tau) f(x) dx$ exists and if the limit is non-random. This is indeed the case for *ergodic* and *stationary processes*. But spectral theory of these processes has advantages which exceed the spectral Wiener theory of deterministic signals considered previously.

Indeed, it admits a representation of a signal *as a random superposition of complex exponentials* and this representation is very important in spectral analysis of these random signals (Carmona et al. 1998, pp. 74–79). But there is no decomposition adapted in the Wiener theory of deterministic signals.

5.11.2.2 Extensions: Deterministic Spectral Theory for Time Series

Note that this type of analysis could be illustrated by taking for example time series resulting from *stock exchange* or *financial markets*.²⁹

Correlogram and Memory of a Series

In this section, let us consider the *finite time series*

$$f = \{f_0, f_1, \dots, f_{N-1}\} \quad (5.106)$$

of real numbers and it is discussed, in different ways, of quantifying the statistical correlations between the successive values of the f_j . The most used *dependence measure* is the *sample autocovariance function*, defined as follows:

$$c_f(j) = \frac{1}{N} \sum_{k=0}^{N-1-|j|} (f_k - \bar{f})(f_{k+|j|} - \bar{f}), \quad (a)$$

²⁹ It is in particular about the evolution of the logarithms of the prices of a “*futures*” contracts of Japanese yen that R. Carmona, W.L. Hwang and B. Torrésani develop their argument.

where \bar{f} is the average of the sample which is simply written:

$$\bar{f} = \frac{1}{N} \sum_{j=0}^{N-1} f_j. \quad (5.107)$$

The writing (a) of *the autocovariance function above* is usually preferred to the following writing (b):

$$\tilde{c}_f(j) = \frac{1}{N-|j|} \sum_{k=0}^{N-1-|j|} (f_k - \bar{f})(f_{k+|j|} - \bar{f}) = \frac{N}{N-|j|} c_f(j). \quad (b)$$

j in the expression above is considered as a *lag* and often corresponds to *non-negative values*. From this autocovariance function we are led to the notion of *sample autocorrelation function* of the signal f , which is written:

$$\rho_f(j) = \frac{c_f(j)}{c_f(0)} = \frac{\sum_{j'=0}^{N-1-j} (f_{j'} - \bar{f})(f_{j+j'} - \bar{f})}{\sum_{j'=0}^{N-1-j} (f_{j'} - \bar{f})^2}, \quad j = 0, 1, \dots, N-1. \quad (c)$$

The *graph* of the *autocorrelation function* $\rho_f(j)$ of j is called a *correlogram*. (The *maximum lag corresponds to $N-1$* , even if we use, in practice, values which usually are much smaller.) Commonly, it is said that a series has a “long memory” when the correlogram decreases slowly, and it is said that a series has a short memory when the correlogram decrease quickly. However when the series is “detrended” by differentiation for example, the phenomenon of long memory disappears and different new correlograms take place. The *correlograms of stock exchange indexes* on the long term generally highlight these *phenomena of long memories* where a *slow decrease of correlograms* is observed.

5.11.2.3 Periodogram: Interpretations

The objective of spectral theory is to represent a signal as a sum of trigonometric functions with specific phases and amplitudes. All the phases involved in the representation are called the *spectrum* of the series and the size of amplitudes is summarized in what is usually called the *periodogram*. If one quickly points out the expression of the *discrete Fourier transform* (DFT) one writes

$$\hat{f}_j = \sum_{k=0}^{N-1} f_k e^{-2i\pi j \omega_k}, \quad k = 0, 1, \dots, N-1, \quad (5.108)$$

ω_k is the *natural frequency* $\omega_k = k/N$. Usually in the Fourier transform it is ξ which is equivalent to the relation: $\xi = 2\pi\omega$. It is also pointed out that the original signal

can be reconstituted from an inverse Fourier transform:

$$f_j = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}_j e^{2i\pi j \omega_k}, \quad k = 0, 1, \dots, N-1, \quad (5.109)$$

the inversion formula above expresses that any finite series f can be written as a linear combination of complex exponentials. From this decomposition, in particular when the original series is real, one can write that: $a_k = \frac{1}{N} \Re \hat{f}_k$ and $b_k = \frac{1}{N} \Im \hat{f}_k$. Moreover, we have:³⁰

$$f_j - \bar{f} = \sum_{k=2}^{N-1} e_j(k), \quad (5.110)$$

the trigonometric functions e_j are defined by

$$e_j(k) = r_k \cos(2\pi j \omega_k - \varphi_k), \quad (5.111)$$

and $r_k = \sqrt{a_k^2 + b_k^2}$, $\varphi_k = \tan^{-1}(b_k/a_k)$. Thus, $k = 1, \dots, N/2$, and we have

$$a_k = a_{N-k} \quad \text{and} \quad b_k = -b_{N-k}, \quad (5.112)$$

and then,

$$r_k^2 = r_{N-k}^2 \quad \text{and} \quad e_j(k) = e_j(N-k). \quad (5.113)$$

What precedes implies the following trigonometrical representation:

$$f_j - \bar{f} = \begin{cases} 2 \sum_{k=1}^{N/2} r_k \cos(2\pi j \omega_k) & \text{if } N \text{ odd,} \\ 2 \sum_{k=1}^{(N/2)-1} r_k \cos(2\pi j \omega_k) + r_{N/2} \cos(2\pi j \omega_{N/2}) & \text{if } N \text{ even.} \end{cases}$$

The purpose of the equation above is to “untangle the dependence among the complex exponentials appearing in the inverse Fourier transform”. Such a dependence is not present any more in the equation above because the trigonometric functions $e_j(k) = r_k \cos(2\pi j \omega_k)$ appearing in the function above are mutually orthogonal. As expressed previously, any series f can be written as a sum of cosine functions of frequencies $\omega_1 = 1/N, \omega_2 = 1/N, \dots, \omega_{N/2} = (N/2)/N$. (Remember that $r_k^2 = \frac{1}{N^2} \left| \sum_{j=0}^{N-1} f_j e^{-2i\pi j \omega_k} \right|^2$, with $k = 0, 1, \dots, N/2$ and $\omega_k = k/N$.) The graph of $N_{r_k^2}$ for ω_k is called the *periodogram* of the series f , and the function $v_f(\omega)$ defined for $\omega \in [0, 1]$, by:

$$v_f(\omega) = \frac{1}{N} \left| \sum_{j=0}^{N-1} f_j e^{-2i\pi j \omega} \right|^2 : \text{Sample spectral density} \quad (d)$$

considering $\omega \in [0, 1/2]$ and $v_f(\omega) = v_f(1 - \omega)$, considering $\omega \in [1/2, 1]$ this expression is called the “sample spectral density” of f . Often some values of r_k are

³⁰ Note that $a_1 = \bar{f}$ and $b_1 = 0$.

very large compared with the others, it is one of the reasons *which creates difficulties to analyze the information contained in the periodogram*. However to represent the *logarithm* of the values r_k^2 , instead of the values themselves, helps to see more elements relating to the frequencies. Furthermore, we have:

$$\frac{1}{N} \sum_{j=1}^{N-1} \frac{Nr_k^2}{\sigma^2} = 1. \quad (5.114)$$

If σ^2 is expressed as a variance

$$\sigma^2 = \frac{1}{N} \sum_{j=0}^{N-1} (f_j - \bar{f})^2, \quad (5.115)$$

then in practice we have more graphic information about the contents of the set of frequencies, it is represented graphically as follows:

$$\left[\text{Log} \left(\frac{Nr_k^2}{\sigma^2} \right) \text{ against } \omega_k \right]. \quad (5.116)$$

The remarks which follow are useful for the interpretation of periodograms, which are:

- “Smooth” when the amplitudes of cosine functions of low frequencies are large compared with the other frequencies.
- “Erratic” when the amplitudes of cosine functions of high frequencies are large compared with the other frequencies.
- With a “peak” at the frequency ω when the signal is a cosine or a sine function of period $1/\omega$.
- With a “peak” at the frequency ω and “peaks” at multiple frequencies of ω (harmonic) when the signal is a periodic function of period $1/\omega$ without being a sine or cosine function.

Comments on the Comparison Between Correlogram and Periodogram

Usually a debate is open between the two possible definitions [(a) or (b)] seen previously of the autocovariance function of finite time series. The two definitions differ only when the size N of the data is large, however one can choose. The second definition \tilde{c}_f (corresponding to the equation (b)) is adapted to the point of view of the statistical analysis in particular when samples f_j of random variables form a *stationary stochastic process*. This second definition provides *an unbiased estimation of the true autocovariance function*. But it is generally (a), the first definition c_f which is selected as the autocovariance function. Usually one chooses this definition because (except for the particular cases for which all the f_j are equal to each other): the sequence $c_f = \{c_f(j)\}_j$ defined by $c_f(j) = \frac{1}{N} \sum_{k=0}^{N-1-|j|} (f_k - \bar{f})(f_{k+|j|} - \bar{f})$ when

$|j| < N$ and 0 otherwise the sequence is *defined non-negative*. Which implies due to the *Bochner's theorem*³¹ the existence of a finite non-negative measure v_f on $[0,1]$ whose Fourier transform is the sequence c_f (such a measure is absolutely continuous). One writes this density $v_f(\omega)$

$$v_f(\omega) = \sum_{j=-(N-1)}^{N-1} c_f(j) e^{2i\pi j\omega}, \quad \omega \in [0, 1]. \quad (5.117)$$

Such a $v_f(\omega)$ is called the *spectral density* due to the fact that it is equal to the quantity defined previously having the same name. One can use simple trigonometrical handling to show that the function $v_f(\omega)$ defined above satisfied:

$$v_f(\omega) = \frac{1}{N} \left| \sum_{j=0}^{N-1} f_j e^{-2i\pi j\omega} \right|^2. \quad (5.118)$$

That means that the *autocovariance function* c_f and the *spectral density* $v_f(\omega)$ form a “Fourier pair”, because they are Fourier transforms the one each other. This correspondence is of a quite essential interest.

5.11.2.4 Remark About the Non-Parametric Spectral Estimation: Case of the Non-Stationary Process Locally Stationary

If a *stochastic process is stationary*, then its *autocovariance* defines what is called a *convolution operator* which is *diagonalized by Fourier expansions*. It is one of the starting points of all the non-parametric spectral estimation methods. In a number of practical situations, the function of autocovariance does not allow to see a convolution operator and the problem of the spectral estimation cannot be solved by the standard Fourier analysis. The stationarity may be broken in various manners. One can give two examples of time–frequency or time-scale representations to illustrate the subject:

- The first example is the case of “locally stationary process”, i.e. processes for which there is a *local spectrum varying slowly*. In such case the time–frequency representations provide efficient tools for these local spectral estimation subjects.
- The second example is the case of *self-similar process* such as *fractional Brownian motions* which can have *stationary increases*, and after the *wavelets transform* the non-stationarities disappear.

³¹ **Theorem (Bochner theorem).** Among the continuous functions on \mathbb{R}^n , the positive definite functions are those functions which are the Fourier transform of finite measures.

5.11.2.5 Spectral Theory of Stationary Random Processes

Stationary Processes

Consider a signal made up of real numbers such as: $f = \{f_0, f_1, \dots, f_{N-1}\}$, this signal is *random* if the numbers f_j can be regarded as results of a finite sequence of random variables. An example is provided through observations of *a deterministic signal in the presence of additional noise*: signal + noise. Consequently, the observations can be written in the following way:

$$f_j = f_j^* + \varepsilon_j \quad (5.119)$$

where $f_j^* = \{f_0^*, f_1^*, \dots, f_{N-1}^*\}$ is of “deterministic origin” and $\varepsilon = \{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{N-1}\}$ is the “noise perturbation”. The field of the random series searches to develop tools to carry out forecasts from observed time series. The usual methods use models made by series generators and the majority of these models is based on the notion of *temporal homogeneity* of signals. This notion is regarded as relative to the *stationarity*. The following lines provide some elements of the theory of stationary processes, but it is important to say *that generally the most interesting signals are non-stationary* and for these signals the theory is inoperative. (It will be also noticed that beyond the *non-stationarity*, the majority of *complex signals* are rather *non-Gaussian*.)

The most adapted way to analyze the finite signals is to regard the set of N samples of f_j as a part of a *doubly infinite sequence* $\dots, f_{-1}, f_0, f_1, \dots$ which can be seen as a set of regular samples which comes from a continuous signal but discretized. In fact, one supposes that there is a function f of a continuous variable $x \in \mathbb{R}$, such that:

$$f_j = f(x_j), \quad \text{with } j = 0, 1, \dots, N-1 \quad (5.120)$$

the x_j corresponds to the time at which the measurements are made. One considers the regular samples provided by the observations at times $x_j = x_0 + j\Delta x$ for a fixed sampling interval Δx which allows to sample the signal. One can define the concept of stationarity in the following way:

Stationarity in the Strict Sense

A random signal $f = \{f(x); x \in \mathbb{R}\}$ is *stationary in the strict sense* if for any choice x and $x_1 < \dots < x_k$ of real numbers, the random vectors:

$$[f(x_1), \dots, f(x_k)] \quad \text{and} \quad [f(x_1 + x), \dots, f(x_k + x)] \quad (5.121)$$

have the same distribution in \mathbb{R}^k (or in \mathbb{C}^k if it is a complex signal). Any statistic obtained by translation or shift of such random processes is *invariant*. One can say that *the first moment* $m(x) = E\{f(x)\}$ is constant and independent of the variable x .

Considering the case of a finite signal, the *stationarity* can be defined without the notion of continuous variables functions. Thus, it is written that *a random signal*

$f = \{f_0, f_1, \dots, f_{N-1}\}$ is known as *stationary* if for any choice of j and $j_1 < \dots < j_k$ of integers in $\{0, 1, \dots, N-1\}$, the following random vectors

$$\{f_{j_1}, \dots, f_{j_k}\} \quad \text{and} \quad \{f_{j_1+j}, \dots, f_{j_k+j}\} \quad (5.122)$$

take advantage of the same distribution in \mathbb{R}^k or in \mathbb{C}^k for the complex signals.³² If one considers the case of a *signal* to which one adds a *noise* as previously: $f_j = f_j^* + \varepsilon_j$, the stationarity hypothesis is satisfied for the *noise component*, i.e. for $\varepsilon = \{\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{N-1}\}$, but if $f^* = \{f_0^*, f_1^*, \dots, f_{N-1}^*\}$ is not constant, the stationarity cannot be proved for the *other component* $f = \{f_0, f_1, \dots, f_{N-1}\}$.

Stationarity in the Wide Sense

There exists a weaker definition of the stationarity. Consider a random process $f = \{f(x); x \in \mathbb{R}\}$ of order 2 (for a continuous variable x), that means that all the random variables $f(x)$ are square integrable. This random process is known as stationary in the wide sense if

$$m_f(x) = E\{f(x)\} \quad (5.123)$$

is independent of x , that means $m_f(x) = m$ for constant m , and if its autocovariance

$$C_f(y, x) = E\{(f(y) - m)(\overline{f(x) - m})\} \quad (5.124)$$

is a function of the difference $(y - x)$, i.e. for a one-variable non-negative function noted C_f , if:

$$C_f(y, x) = C_f(y - x). \quad (5.125)$$

By means of the Bochner theorem it is possible to say that there exists a non-negative finite measure v_f on \mathbb{R} which satisfies the definition $C_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\tau\omega} v_f d\omega$. This measure v_f is called the *spectral measure* of the stationary process $f = \{f(x); x \in \mathbb{R}\}$. The expression $C_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\tau\omega} v_f d\omega$ allows to say that the covariance function C_f is the Fourier transform of the spectral measure v_f .

Usually, the covariance function C_f is integrable on \mathbb{R} , then it is possible to introduce the spectral measure without the Bochner theorem. Thus, the following *Fourier transformation* (under the integrability condition) is defined as a function on \mathbb{R} :

$$\vartheta_f(\xi) = \int C_f(x) e^{-ix\xi} d\xi. \quad (5.126)$$

Because of the autocovariance function C_f is defined non-negative, the function v_f is non-negative. This can be called the *spectral function* or *spectral density*.³³ The

³² The addition of integers is understood as mod N .

³³ It is the density of the spectral measure which is possible by means of the Bochner theorem because of $v_f(d\xi) = v_f(\xi)d\xi$, and which allows the utilization of the same notation to express the *spectral measure* and the *spectral function*.

autocorrelation function ρ_f defined as follows

$$\rho_f(y, x) = \frac{C_f}{(C_f(y, y))^{1/2}(C_f(x, x))^{1/2}} \quad (5.127)$$

is also written as a function of the difference $(y - x)$ when the process is stationary, and in such a case there is $\rho_f(y, x) = \rho_f(y - x)$ with:

$$\rho_f(y - x) = \frac{C_f(y - x)}{C_f(0)}. \quad (5.128)$$

In a similar way, it is possible to provide a definition of *the stationarity in the wide sense for the finite random signals*. In all these approaches it is always assumed that the average of a stationary signal in the wide sense is zero.

Comparison with the Wiener Spectral Theory

It is interesting to compare the elements of the spectral theory of this type of process with the Wiener spectral theory, presented previously. *Ergodic theorem*³⁴ explains that the limit:

$$C_f(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} f(x + \tau) f(x) dx \quad (5.129)$$

exists, if f is a *stationary process*. Furthermore, if the process f is *Ergodic*, i.e. if the shift-invariant functions of signal are constant, the limit above results from an almost sure convergence, and is deterministic, and is equivalent to the autocovariance $C_f(\tau)$ seen previously defined when $y = x + \tau$, as follows:

$$C_f(y, x) = E \left\{ (f(y) - m) \overline{(f(x) - m)} \right\} \quad (5.130)$$

and

$$C_f(y, x) = C_f(y - x). \quad (5.131)$$

That means that the “Spectral theory of stationary processes contains the Wiener theory when the analyzed processes are Ergodic”. (“It is known that the estimates calculated from $\bar{f} = \frac{1}{N} \sum_{j=0}^{N-1} f_j$ or from the sample autocovariance function $c_f(j) = \frac{1}{N} \sum_{K=0}^{N-1-|j|} (f_k - \bar{f})(f_{k+|j|} - \bar{f})$ or from the sample autocorrelation function $\rho_f(j) = \frac{c_f(j)}{c_f(0)} = \frac{\sum_{j'=0}^{N-1-j} (f_{j'} - \bar{f})(f_{j+j'} - \bar{f})}{\sum_{j'=0}^{N-1-j} (f_{j'} - \bar{f})^2}$ are estimations of their theoretical counterparts: m , C_f , and ρ_f ”.)

³⁴ Regarded as a generalization of the law of large numbers.

5.12 The Construction of Orthonormal Bases and Riesz Bases

5.12.1 Signal of Finite Energy

The time series or signals are often regarded as vectors, and can also be regarded as measurable functions. Then, one poses the integration Lebesgue theorem, which explains that *a function f is integrable*, if f is of *finite energy*:

$$\int_{-\infty}^{+\infty} |f(t)| dt < +\infty. \quad (5.132)$$

The space of integrable functions is noted $\mathbf{L}^1(\mathbb{R})$ if $\int_{-\infty}^{+\infty} |f_1(t) - f_2(t)| dt = 0$. This means that f_1 and f_2 can differ only on a set of points of measure zero. Roughly, it is said that they are equal almost everywhere. Moreover, the time series or signals which are the subject of transformations in the time–frequency planes are offered to the handling of scientists who practise the signal processing and *define a metric* and exploit the properties of *vector spaces*. The concepts of distance, norm, convergence, integration, orthogonality, projection, basis, are largely used and enriched.

5.12.2 Reminders: Norms and Banach Spaces

Banach space: A complete normed vector space, i.e. a normed space in which any Cauchy sequence converges, is called *Banach space*. In order to define a distance we work inside a vector space \mathbf{H} which admits a norm. The properties of a norm are the following:

$$\forall f \in \mathbf{H}, \|f\| \geq 0 \text{ and } \|f\| = 0 \Leftrightarrow f = 0, \quad (5.133)$$

$$\forall \lambda \in \mathbf{C} \|\lambda f\| = |\lambda| \|f\|, \quad (5.134)$$

$$\forall f, g \in \mathbf{H}, \|f + g\| \leq \|f\| + \|g\|. \quad (5.135)$$

With such a norm, the convergence of a family (or sequence) of positive functions $\{f_n\}_{n \in \mathbb{N}}$ towards f in \mathbf{H} means that:

$$\lim_{n \rightarrow +\infty} f_n = f \Leftrightarrow \lim_{n \rightarrow +\infty} \|f_n - f\| = 0. \quad (5.136)$$

One can impose a complete state of properties using the Cauchy sequence. A family or sequence $\{f_n\}_{n \in \mathbb{N}}$ is a Cauchy sequence if for all $\varepsilon > 0$, if n and p are large enough, then $\|f_n - f_p\| < \varepsilon$. The vector space \mathbf{H} is said complete if any Cauchy sequence in \mathbf{H} converges towards an element of \mathbf{H} . One can evoke the particular case of a space $\mathbf{L}^p(\mathbb{R})$. The space $\mathbf{L}^p(\mathbb{R})$ is composed of measurable functions on \mathbb{R} for which:

$$\|f\|_p = \left(\int_{-\infty}^{+\infty} |f(t)|^p dt \right)^{1/p} < +\infty. \quad (5.137)$$

\mathbf{L}^p is a Banach space and the integral above determines a norm.

5.12.3 Reminders: Inner Products and Hilbert Spaces

A Hilbert space is a space on \mathbb{R} or \mathbb{C} , provided with a *scalar product* whose *associated normed space is complete*. The elements of these spaces historically were functions coming from the *formalization of oscillatory phenomena* and from the *calculus of variations* where the searched solutions (generally integrals) appear as sums of a series of functions, often trigonometric, which one approaches by orthogonal polynomials for a scalar product. A *complete prehilbert space* (for the norm associated with the scalar product) is called *Hilbert space*.

The *Hilbert space* is a *Banach space* provided with an inner product. The inner product of two vectors $\langle f, g \rangle$ is linear and respects:

$$\forall \lambda_1, \lambda_2 \in \mathbf{C}, \langle \lambda_1 f_1 + \lambda_2 f_2, g \rangle = \lambda_1 \langle f_1, g \rangle + \lambda_2 \langle f_2, g \rangle. \quad (5.138)$$

And there is an *Hermitian symmetry*:³⁵

$$\langle f, g \rangle = \langle g, f \rangle^* \quad (5.139)$$

Moreover, $\langle f, f \rangle \geq 0$ and $\langle f, f \rangle = 0 \Leftrightarrow f = 0$. It can be also shown that $\|f\| = \langle f, f \rangle^{1/2}$ is a norm. The inequality of Cauchy–Schwarz can be also shown:

$$|\langle f, g \rangle| \leq \|f\| \|g\|, \quad (5.140)$$

and if f and g are *linearly independent*, there is the following equality:

$$|\langle f, g \rangle| = \|f\| \|g\|. \quad (5.141)$$

5.12.4 Orthonormal Basis

Orthogonal vectors: A family of vector noted $\{e_n\}_{n \in N}$ of a Hilbert space is *orthogonal* if for $n \neq p$:

$$\langle e_n, e_p \rangle = 0. \quad (5.142)$$

(It will be noted that from a statistical point of view, *two orthogonal vectors are uncorrelated*.)

Orthogonal basis: If for $f \in \mathbf{H}$, there exists a sequence $\lambda[n]$ such that:³⁶

$$\lim_{N \rightarrow +\infty} \left\| f - \sum_{n=0}^N \lambda[n] e_n \right\| = 0, \quad (5.143)$$

³⁵ *Property of Hermitian symmetry:* For $f(t) \in R$, on the Fourier transform we obtain: $\hat{f}(-\omega) = \hat{f}^*(\omega)$.

Property of complex conjugates: For $f^*(t)$, on the Fourier transform we obtain: $\hat{f}^*(-\omega)$.

³⁶ $\lambda[n]$: Corresponds to a discrete sequence.

then $\{e_n\}_{n \in N}$ is an *orthogonal basis* of \mathbf{H} . The orthogonality implies that ($\lambda[n]$ is a sequence):

$$\lambda[n] = \frac{\langle f, e_n \rangle}{\|e_n\|^2}, \quad (5.144)$$

and one can write:

$$f = \sum_{n=0}^{+\infty} \frac{\langle f, e_n \rangle}{\|e_n\|^2} e_n. \quad (5.145)$$

The basis is orthonormal if $\|e_n\| = 1$ for any $n \in N$. If one calculates the inner product of $g \in \mathbf{H}$ with the Parseval equation for orthonormal bases:

$$\langle f, g \rangle = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \langle f, e_n \rangle^*. \quad (5.146)$$

When $g = f$, one obtains an *Energy Conservation* derived from the *Plancherel formula*:

$$\|f\|^2 = \sum_{n=0}^{+\infty} |\langle f, e_n \rangle|^2. \quad (5.147)$$

It is possible to construct *orthonormal bases* by means of *local cosine (or sine) functions and wavelets* or *wavelet packets*.

5.12.5 Riesz Basis, Dual Family and Biorthogonality

A *Riesz basis*: A family of vectors denoted $\{e_n\}_{n \in N}$ is a Riesz basis of the Hilbert space \mathbf{H} , if it is *linearly independent* and if there exists $A > 0$ and $B > 0$ such that for any $f \in \mathbf{H}$. It is possible to find a sequence $\lambda[n]$ with $f = \sum_{n=0}^{+\infty} \lambda[n] e_n$, which satisfies

$$\frac{1}{B} \|f\|^2 \leq \sum_n |\lambda[n]|^2 \leq \frac{1}{A} \|f\|^2. \quad (5.148)$$

The *Riesz theorem* proves that there exists \tilde{e}_n such that $\lambda[n] = \langle f, \tilde{e}_n \rangle$, and furthermore the preceding inequations $\frac{1}{B} \|f\|^2 \leq \sum_n |\lambda[n]|^2 \leq \frac{1}{A} \|f\|^2$ implies:

$$\frac{1}{B} \|f\|^2 \leq \sum_n |\langle f, \tilde{e}_n \rangle|^2 \leq \frac{1}{A} \|f\|^2. \quad (5.149)$$

Moreover, it is also possible to write that for any $f \in \mathbf{H}$,

$$A \|f\|^2 \leq \sum_n |\langle f, e_n \rangle|^2 \leq B \|f\|^2, \quad (5.150)$$

And

$$f = \sum_{n=0}^{+\infty} \langle f, \tilde{e}_n \rangle e_n = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \tilde{e}_n, \quad (5.151)$$

$\{\tilde{e}_n\}_{n \in N}$ is a linearly independent *dual family* and is also a *Riesz basis*. If for example $f = e_s$, we have $e_s = \sum_{n=0}^{+\infty} \langle e_s, \tilde{e}_n \rangle e_n$. The linear independence of $\{e_n\}_{n \in N}$ implies a biorthogonality of dual bases, which are called biorthogonal bases:

$$\langle e_n, \tilde{e}_s \rangle = \delta[n - s]. \quad (5.152)$$

5.12.6 Orthogonal Projection

Let \mathbf{V} be a subspace of the space \mathbf{H} . A projector P_V on \mathbf{V} is a linear operator which satisfies:

$$\forall f \in \mathbf{H}, P_V f \in \mathbf{V} \quad \text{and} \quad \forall f \in \mathbf{V}, P_V f = f. \quad (5.153)$$

The projector P_V is orthogonal if:

$$\forall f \in \mathbf{H}, \forall g \in \mathbf{V}, \quad \langle f - P_V f, g \rangle = 0. \quad (5.154)$$

The following properties of projectors can be proposed:

- If $\{e_n\}_{n \in N}$ is an orthogonal basis of \mathbf{V} , then we have:

$$P_V f = \sum_{n=0}^{+\infty} \frac{\langle f, e_n \rangle}{\|e_n\|^2} e_n. \quad (5.155)$$

- If $\{e_n\}_{n \in N}$ is a Riesz basis of \mathbf{V} and $\{\tilde{e}_n\}_{n \in N}$ is the orthogonal basis then we have:

$$P_V f = \sum_{n=0}^{+\infty} \langle f, e_n \rangle \tilde{e}_n = \sum_{n=0}^{+\infty} \langle f, \tilde{e}_n \rangle e_n. \quad (5.156)$$

5.12.7 The Construction of Orthonormal Basis and Calculation of the “Detail” Coefficient on Dyadic Scale

Subsequently, it will be developed the notions of orthonormal basis, Fourier basis and wavelet basis, which are fundamental in the time–frequency analysis, but before to familiarize with this notion, one shows quickly how to build a basis from a dyadic scale. Let $\psi(t)$ be a wavelet whose dilations and translations generate an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$:

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - 2^j n}{2^j}\right) \right\}_{(j,n) \in \mathbb{Z}^2}. \quad (5.157)$$

All the *signals of finite energy* can be *decomposed* on this type of *wavelet basis* $\{\psi_{j,n}(t)\}_{(j,n) \in \mathbb{Z}^2}$, which is written:

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}. \quad (5.158)$$

We know that $\psi(t)$ has a zero-integral or a zero-average and we can write:

$$d_j(t) = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}(t), \quad (5.159)$$

which corresponds to the variations of “detail” at a dyadic scale 2^j . The approximation of f is constructed from the addition of these variations of detail on each scale, whose sum allows to reconstruct the signal f . Note that if the signal has structures rather *smooth*, one can then approach the signal with an approximation from which the “details” at fine scale have been removed, thus the writing of $f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j,n} \rangle \psi_{j,n}$ becomes:

$$f_J(t) = \sum_{j=J}^{+\infty} d_j(t). \quad (5.160)$$

5.13 Concept of Frames

Before quickly presenting the concept of frames, it seemed fundamental to re-introduce two elements of the Fourier theory, without which (even reduced to their rudiments) the time–frequency analysis and the wavelet analysis seem difficult or even impossible to understand. These elements are the Fourier transform in L^2 and the *Parseval and Plancherel formulas* that one has for the latter already mentioned.

5.13.1 The Fourier Transform in $L^2(\mathbb{R})$

Briefly, remember that the following spaces correspond to the respective functions or associated signals:

- $L^2(\mathbb{R})$: Finite energy functions: $\int |f(t)|^2 dt < +\infty$. And consequently is the space of integrable square functions.
- $L^p(\mathbb{R})$: Functions such that $\int |f(t)|^p dt < +\infty$.
- $\ell^2(\mathbb{Z})$: Discrete finite energy signals: $\sum_{n=-\infty}^{+\infty} |f(n)|^2 < +\infty$.
- $\ell^p(\mathbb{Z})$: Discrete signals such that: $\sum_{n=-\infty}^{+\infty} |f(n)|^p < +\infty$.

5.13.1.1 Convergence Limit of Fourier Transform of Function in $L^1(\mathbb{R})$

As seen several times, the Fourier transform of a function $f(t)$ is carried out by the integration of the function f and trigonometric functions $e^{-i\omega t}$. The latter are used to decompose and rewrite the studied function into a sum of sine curves, whose transformation in the Fourier space provides Fourier coefficients, which characterize the studied function in the space of the frequencies (ω). In other words, *the number of sine or cosine functions and their respective frequencies define f*. The Fourier transform is written:

$$\widehat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt. \quad (5.161)$$

If the function f is integrable (i.e. if f is of finite energy $\int_{-\infty}^{+\infty} |f(t)| dt < +\infty$), i.e. if $f \in L^1(\mathbb{R})$, then there exists a bound for the Fourier transform, i.e. there is convergence:

$$|\widehat{f}(\omega)| \leq \int_{-\infty}^{+\infty} |f(t)| dt < +\infty. \quad (5.162)$$

Thus, the Fourier transform through the absolute value $|\widehat{f}(\omega)|$ is bounded by the finite integral of the absolute value of the function f :

$$\int_{-\infty}^{+\infty} |f(t)| dt. \quad (5.163)$$

5.13.1.2 Fourier Transform in the Space $L^2(\mathbb{R})$ of Square Integrable Functions

Certain functions are not integrable, because they are not continuous. This case is not rare and can also be transposed to signals themselves; *discontinuities* or *singularities* can create problems concerning the integrability of a function and by extension of a signal. A *non-integrable function f is not continuous, however its square is integrable*. One can for example evoke the indicator function which is non-continuous.³⁷ This is on the base of this observation that the Fourier transform of integrable functions (and thus continuous) is extended to the Fourier transform of square integrable functions, in order to compensate the difficulties related to problems of discontinuities. The extension of the Fourier transform is done in the space $L^2(\mathbb{R})$ of finite energy functions $\int_{-\infty}^{+\infty} |f(t)| dt < +\infty$ which is also the space of square integrable functions. Consequently, we are located in the *Hilbert* space, where we can be endowed with an *inner product* and a *norm*. It is pointed out that the inner product of f and g belonging to $L^2(\mathbb{R})$ is written

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t) g^*(t) dt, \quad (5.164)$$

³⁷ Example. The Fourier transform of indicator function: $f = \mathbf{1}_{[-1,1]}$ is written as a multiple of cardinal sine: $\widehat{f}(\omega) = \int_{-1}^{+1} e^{-i\omega t} dt = (2 \sin(\omega)) / \omega$.

and that the norm in the Hilbert space $\mathbf{L}^2(\mathbb{R})$ is:

$$\|f\|^2 = \langle f, f \rangle = \int_{-\infty}^{+\infty} |f(t)|^2 dt. \quad (5.165)$$

5.13.1.3 Parseval and Plancherel Formulas: Norm and Inner Product are Conserved by the Fourier Transform up to a Factor of 2π

Theorem 5.4 (Parseval and Plancherel transfer formulas). *If f and g belong to $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$, then*

$$\int_{-\infty}^{+\infty} f(t)g(t)dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{g}^*(\omega)d\omega, \quad (5.166)$$

which is the Parseval formula. When $f = g$, it follows the well-known Plancherel formula:

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2 d\omega. \quad (5.167)$$

5.13.2 Frames

The frames provide a stable and possibly redundant representation of a signal. *The frames are a generalization of the basic notion of a vector space.* A frame is a family of vectors which allows to represent any signal (\mathbf{L}^2) by its scalar products with these vectors. The frames provide a discrete and redundant representation of the signal.

Definition 5.2 (Frame). A sequence³⁸ $\{\phi_n\}_{n \in \Gamma}$ (with the index n belonging to a finite or infinite set noted: Γ) of vectors of a Hilbert space \mathbf{H} is a frame of \mathbf{H} if there exist two constants $A > 0$ and $B > 0$ such that, for any f of \mathbf{H} ,

$$A \|f\|^2 \leq \sum_{n \in \Gamma} |\langle f, \phi_n \rangle|^2 \leq B \|f\|^2, \quad (5.168)$$

If $A = B$, it is said that the frame is tight.³⁹ (A Riesz basis is a frame of independent vectors.)

If the vectors of frame are independent, then $A \leq 1 \leq B$.⁴⁰ The frame is an orthonormal basis if and only if $A = B = 1$. If $A > 1$, the frame is *redundant*.

³⁸ Or family of vectors.

³⁹ Example. We are located in the plane and we consider a family of three unit vectors resulting from the others by rotation of the third of a turn. It forms a tight (adjusted) frame of the plane, with $A = B = 3/2$.

⁴⁰ It is taken vectors of frame normalized to 1.

5.13.2.1 Windowed Fourier Frames and Fourier Atoms

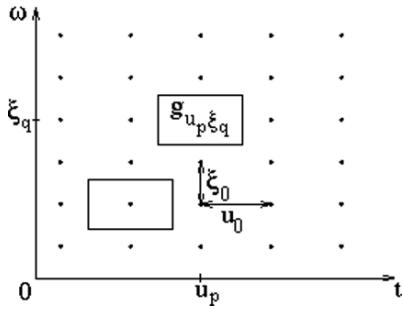
It is pointed out that the windowed Fourier transform is written

$$\langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t)g(t-u)e^{i\xi(t-u)}dt. \quad (5.169)$$

The atom used is a sinusoid multiplied by a window g . The family of vectors of the analysis is obtained by translation and modulation of the window:

$$g_{u,\xi}(t) = g(t-u)e^{i\xi(t-u)}. \quad (5.170)$$

This function is *centered* for the frequencies at ξ and symmetrical in relation to u . *The temporal standard deviation is constant. The standard deviation in frequency is constant.* The family is thus obtained by translation in time and frequency of a single window. The Heisenberg boxes of windowed Fourier atoms have dimensions independent of the center in time and frequency. A frame is obtained by covering the time–frequency plane by discrete boxes. The grid used is uniform rectangular



To obtain a frame, the following condition on the paving (tiling) of the zone proves to be necessary:

$$\frac{2\pi}{u_0\xi_0} \geq 1. \quad (5.171)$$

5.13.2.2 Frames of Wavelets and Atoms of Wavelets

One points out the writing of the wavelet transform

$$Wf(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt, \quad (5.172)$$

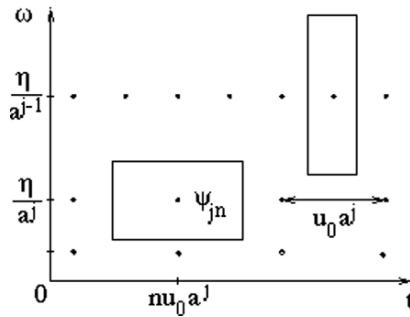
where the atom or the wavelet ψ is a function of zero average, centered at the neighborhood of 0 and of finite energy. The family of vectors is obtained by translation and dilation of the atom:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi \left(\frac{t-u}{s} \right). \quad (5.173)$$

This function is centered at the neighborhood of u , like the windowed Fourier atom. If the center frequency of ψ is indicated by η , the center frequency of the dilated function is in η/s . The standard deviation in time is proportional to s . The standard deviation in frequency is inversely proportional to s . In order to obtain a complete cover, the scale is sampled in an exponential way and denoted a^j and the parameter of translation is uniformly distributed $u = nu_0$. The transform is written:

$$\psi_{j,n}(t) = \frac{1}{\sqrt{a^j}} \psi\left(\frac{t - nu_0 a^j}{a^j}\right). \quad (5.174)$$

In order to pave the time–frequency plane with *wavelet Heisenberg boxes*, one does not use a grid with fixed segmentation, but with time-segments inversely proportional to the frequential segments, which are themselves proportional to the scale.



The wavelet must satisfy the condition which follows:

$$\int_0^{+\infty} \frac{|\widehat{\psi}(\omega)|}{\omega} d\omega < +\infty \quad (5.175)$$

which guarantees the invertibility of the wavelet transform. One will refer to the work of Daubechies or Mallat to know the necessary conditions to the construction of a wavelet frame.

5.13.2.3 Frames and Bases of Time–Frequency Atoms

The windowed Fourier transform and the wavelet transform are written as inner products in $L^2(\mathbb{R})$ with each one their time–frequency atom, that means respectively:

$$Sf(u, \xi) = \langle f, g_{u,\xi} \rangle = \int_{-\infty}^{+\infty} f(t) g_{u,\xi}^*(t) dt, \quad (5.176)$$

$$Wf(u, s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \psi_{u,s}^*(t) dt. \quad (5.177)$$

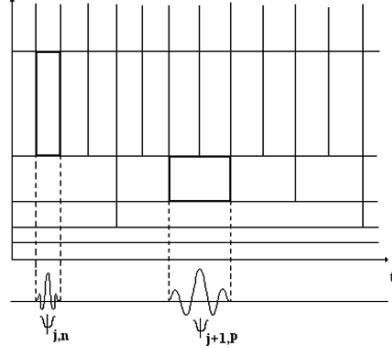


Fig. 5.25 Tiling of the plane by time–frequency boxes of a wavelets basis

A signal can be reconstructed from linear combinations of windowed Fourier atoms or wavelet atoms, which provide a *complete representation* of the signal. The *frame theory* treats conditions of signal reconstructions, stability, completeness and elimination of redundancies in the construction of bases, many notions which are the subject of important developments of which we provided here only preliminary elements.

5.13.3 Tiling of the Time–Frequency Plane by Fourier and Wavelets Bases

A complete orthogonal wavelet basis allows the tiling by time–frequency boxes of wavelets of the whole surface of the plane. Figure 5.25 illustrates this tiling of the plane and highlights two wavelets boxes constructed from $\psi_{j,n}(t)$ and its translation by $2^j n$, i.e. $\psi_{j+1,p}(t)$, which contributes both to the cover of the plane. It is a simple example and a multitude of possibilities is offered to us to cover the plane.

The cover of the time–frequency plane can be carried out from a *Fourier basis of local sinusoids* (i.e. *local cosine bases*) (Fig. 5.26). At this stage, one utilizes *Malvar bases of local sinusoids*, which are constructed with windows $g_p(t)$ on intervals $[a_p, a_{p+1}]$, that one multiplies by cosine functions of the type $\cos(\lambda t + \phi)$ whose frequencies vary. The set is thus written in the following way: $[g_p(t) \cdot \cos(\lambda t + \phi)]$, and allows to carry out translations through the axis of frequencies. The time axis itself is divided by the intervals $[a_p, a_{p+1}]$. This method makes it possible to build a basis of local cosines which divide the time axis by the windows $g_p(t)$.

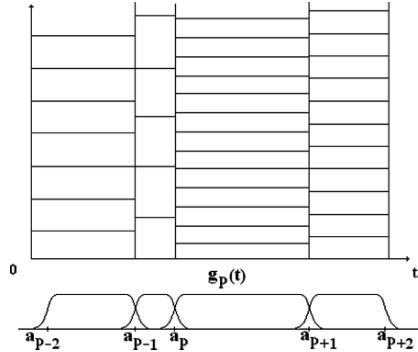


Fig. 5.26 Basis of local sinusoids which cover the time–frequency plane

5.14 Linear and Nonlinear Approximations of a Signal by Projection on an Orthonormal Basis

5.14.1 General Framework of the Linear Approximation and Karhunen–Loève Optimal Basis

Within the framework of the time–frequency analysis, a linear approximation consists in projecting the studied signal f on M vectors (the M first) a priori selected in an orthonormal basis $\{g_m\}_{m \in \mathbb{N}}$. The projection, called *the first M* , is written:

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m, \quad (5.178)$$

and the approximation error is written as *the square of the norm of the difference between the signal and its approximation* ($f - f_M$), or as *the sum of the remaining squared inner products*:

$$\text{Error} = \|f - f_M\|^2 = \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2. \quad (5.179)$$

In a Fourier basis the reduction of the error is related to the *general regularity* of the signal. A *Fourier basis* provides a *good linear approximation* of *uniformly smooth* signals. They are projected on the M first low frequency sinusoids. The linear approximations of these signals have the same properties in the Fourier or wavelet bases.

Choice of an Optimal Basis: This choice consists in minimizing the approximation error from a basis called “Karhunen–Loève basis”, which diagonalizes the covariance matrix of the signal. (This method has to be connected with the projection method of the Singular Spectrum Analysis described in the first part of this book

which consists in operating a projection of the signal on the covariance matrix of this signal.)

5.14.2 Nonlinear Approximation and Adaptive Basis Dependent on the Signal: Regularity and Singularity

For nonlinear structures, the Karhunen–Loève basis is not suitable, it is indeed rather simple to find bases which have errors of approximation smaller than that of Karhunen–Loève. However, there is no in the nonlinear field standard procedure to find an optimal basis as in the linear field. The approximation process is improved by the (no longer “a priori” but) “a posteriori” choice of the M vectors of g_m . These M vectors are identified by their index which belongs to a set $I_M : m \in I_M$. The choice of I_M is crucial, it is done by taking the largest modulus of the inner product: $|\langle f, g_m \rangle|$, i.e. to maximize the amplitude of the inner product. The approximation of f is written:

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m. \quad (5.180)$$

And the approximation error is expressed as the sum of the absolute value of squared inner products, but on the vectors $n \notin I_M$:

$$\text{Error} = \|f - f_M\|^2 = \sum_{n \notin I_M} |\langle f, g_n \rangle|^2. \quad (5.181)$$

During the construction of the basis, the choice of I_M maximizes the amplitude of the inner product and also aims to minimize the error. The notion of the nonlinear approximation comes from the fact that the choice of approximation vectors is modified according to the nature of the signal. (In the previous section in the linear framework, we have seen that when M increases and the approximation error decreases, this can be connected with the global regularity of the signal.) The *global* or *local regularity* and singularity notions of signals are central within the nonlinear approximations. Indeed (see section about the *regularity* concept), the *amplitudes* (i.e. *modulus*) of inner products in a *wavelet* basis are related to the notion of *local regularity* of a signal.

The *irregularities* and *singularities* are symptomatic of nonlinear signals. Because of irregularities of a signal and in particular at the places where this signal is irregular, it is necessary to have an *adaptive approximation grid*, by which the resolution increases where the signal is irregular, which corresponds to a nonlinear approximation having the largest possible inner products of wavelets (maximizing the amplitude of inner products locally). At the places where the signal has singularities, the *nonlinear approximation is much more precise and adapted* than the *linear method* which does not change resolution on the entire signal. Thus, we repeat again, the *amplitude of inner products* in a *wavelet basis* must be linked with the *local regularity* of the signal.

The nonlinear signals can be also approximated by families of vectors which are not built from a unique basis. It will be also noted that the technique of the *best basis* is constructed from *bases of wavelets packets* or from *local sinusoidal functions*. These bases of wavelets packets or sinusoidal local functions are families of orthogonal bases which thus contain various types of time–frequency atoms. The choice of the “best basis” is carried out from a *dictionary of bases*, by the *minimization of a “concave cost function”*. A wavelet packet “best basis” or a best basis of local sinusoidal functions decompose the signal with atoms which are adapted to the time–frequency structures of the signal.

It is also possible to slacken the orthogonality constraint for the choice of bases but that can lead to non-effective or explosive constructions. An approximation of the signal can be produced from M non-orthogonal vectors $\{g_{\gamma_m}\}_{0 \leq m \leq M}$, selected in a redundant dictionary $D = \{g_{\gamma}\}$.

$$f_M = \sum_{m=0}^{M-1} a_m g_{\gamma_m}. \quad (5.182)$$

One can summarize the idea by saying that to optimize the approximation of nonlinear signals, *it is interesting to choose, in an adaptive way, a basis dependent on the signal itself*.

5.14.2.1 Concave Cost Function, Best Basis, Ideal Basis and Entropy

A basis B^* is a *better basis* than the basis B to approximate a signal f if and only if for any concave function $\Phi(u)$ there is:⁴¹

$$\sum_{m=1}^N \Phi\left(\frac{|\langle f, g_m^* \rangle|^2}{\|f\|^2}\right) \leq \sum_{m=1}^N \Phi\left(\frac{|\langle f, g_m \rangle|^2}{\|f\|^2}\right). \quad (5.183)$$

In fact, the comparison of two bases is done by using a single concave function $\Phi(u)$. The *cost of the approximation of a signal f in a basis B^λ* is determined by the sum as follows:⁴²

$$C(f, B^\lambda) = \sum_{m=1}^N \Phi\left(\frac{|\langle f, g_m^\lambda \rangle|^2}{\|f\|^2}\right). \quad (5.184)$$

Coifmann and Wickerhauser construct a *best basis B^** , by *minimizing the cost of a signal f*:

$$C(f, B^*) = \min_{\lambda \in \Lambda} C(f, B^\lambda). \quad (5.185)$$

Even if there is not better basis than B^* , there can exist several bases, which minimize in an equivalent way the cost of f ; then the choice is done according to the

⁴¹ For a demonstration see Mallat (1998, p. 407, Lemma of Hardy, Littlewood, Polya).

⁴² *Schur-concave sum (Ostrowski)*: To refer for example to Tin-Yau Tam (2001). (This concept is also linked to the concept of invariance, S_n -invariance.)

concave function Φ . After having presented how the choice of the “best basis” was carried out, the approximation theory introduces another concept which is the “ideal basis”. An ideal basis is a basis which has one of its vectors proportional to the signal itself, i.e. one can write: $g_m = \alpha f$ with $\alpha \in \mathbb{C}$. Thus the signal f can be approximated and rebuilt from a single vector which is a basis. Any basis B is worse than the ideal basis and better than a “diffusing” basis to approximate a signal f . If $\Phi(0) = 0$, then:

$$\Phi(1) \leq C(f, B) \leq N\Phi\left(\frac{1}{N}\right). \quad (5.186)$$

The concept opposed to the “ideal basis” is the “diffusing basis” for which the approximation quality is lower. All the information contained in the signal is not restored in its integrity. There is an information loss by diffusion, which is linked to the *Entropy concept*. It is written that the entropy $\Phi(x) = -x \log_e x$ is concave for $x \geq 0$. The resulting *cost is regarded as the entropy of the energy distribution*:

$$C(f, B) = - \sum_{m=1}^N \frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \log_e \left(\frac{|\langle f, g_m \rangle|^2}{\|f\|^2} \right). \quad (5.187)$$

Whence it is possible to deduce the bounds of the cost function:

$$0 \leq C(f, B) \leq \log_e N. \quad (5.188)$$

5.14.2.2 Tree Structure, Time–Frequency Plane and Best Basis

A local Fourier “best basis” divides the time–frequency plane in atoms according to the signal itself. The constitution of a dictionary of local sinusoids or wavelet packets requires more than $2^{N/2}$ bases for a signal whose length is N , these are considerable sizes. A best basis minimizes the cost function:

$$C(f, B^*) = \min_{\lambda \in \Lambda} C(f, B^\lambda), \quad (5.189)$$

however to calculate this minimum by a systematic comparison of the cost of each base requires almost $N2^{N/2}$ calculations, which is even more considerable. A “dynamic programming algorithm” developed by *Coifman* and *Wickerhauser* constructs the *best basis* with $O(N \log_2 N)$ calculations. This *algorithm uses the tree structures* constructed from *dictionaries of local Fourier or wavelets bases*. The *trees divide the space into subspaces*, which admit themselves an orthonormal basis of wavelet packets or local sinusoids. In these binary trees of wavelet packets or local sinusoids, each node (at the place of the junction or bifurcation of the fork) corresponds to a space \mathbf{W}_j^p with which an orthonormal basis is associated B_j^p either with wavelet packets, or local sinusoids. The space \mathbf{W}_j^p is divided into two subspaces localized at the nodes:

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \bigoplus \mathbf{W}_{j+1}^{2p+1}. \quad (5.190)$$

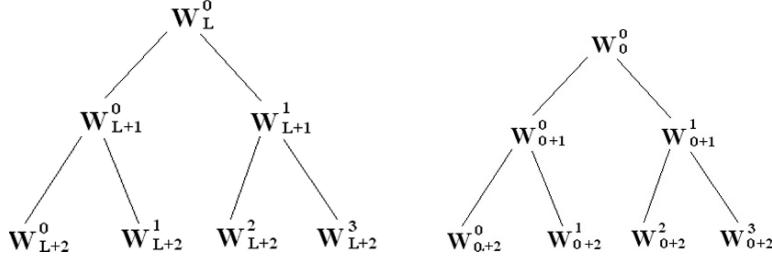


Fig. 5.27 Tree of wavelet packet space (left). Tree of local sinusoids space (right)

The space \mathbf{W}_j^p is thus divided into two subspaces \mathbf{W}_{j+1}^{2p} and \mathbf{W}_{j+1}^{2p+1} , which admit themselves an orthogonal basis each one. The union of the orthogonal bases of \mathbf{W}_{j+1}^{2p} and \mathbf{W}_{j+1}^{2p+1} forms thus the basis associated with \mathbf{W}_j^p . And so on... during the divisions. According to the considered basis (i.e. local sinusoids or the wavelet packets) the initial space is denoted \mathbf{W}_0^0 or \mathbf{W}_L^0 , this space covers⁴³ all the length N of signal, it is the starting point of bifurcations which follow (Fig. 5.27).

The “best basis” of the space \mathbf{W}_j^p is obtained by minimization of the cost function. We have already presented how to write the cost function of a signal f within a basis: $B = \{g_m\}_{0 \leq m < M}$, which is constituted of M vectors. The cost must be calculated according to the division in subspace. Whatever the bases B^0 and B^1 , the idea of Coifman and Wickerhauser is to say that if C is an additive cost function, i.e. if $C(f, B^0 \cup B^1) = C(f, B^0) + C(f, B^1)$, then the *best basis* noted Θ_j^p , corresponds to:

$$\Theta_j^p = \begin{cases} \Theta_{j+1}^{2p} \cup \Theta_{j+1}^{2p+1} & \text{if } C(f, \Theta_{j+1}^{2p}) + C(f, \Theta_{j+1}^{2p+1}) < C(f, B_j^p), \\ B_j^p & \text{if } C(f, \Theta_{j+1}^{2p}) + C(f, \Theta_{j+1}^{2p+1}) \geq C(f, B_j^p). \end{cases} \quad (5.191)$$

This method makes it possible to build the best basis of the space at the root of the tree structure, by calculating in a recursive way the best bases of all spaces \mathbf{W}_j^p inside the tree, from the root downwards. The best bases of the spaces $\{\mathbf{W}_j^p\}_p$ are calculated from the best bases of spaces $\{\mathbf{W}_{j+1}^p\}_p$ by means of the cost calculation method presented previously. The operation is thus repeated until obtaining the best basis on \mathbf{W}_0^0 for the bases of sinusoids and \mathbf{W}_L^0 for the bases of wavelets. The depth of the tree is lower than $\log_2 N$, and the quantity of calculations necessary to obtain the best basis is $O(N \log_2 N)$.

Example of Tree and Wavelet Packet Basis

The multiresolution analysis is a good support to illustrate the wavelet packet trees. Indeed, in the multiresolution analysis (MRA), the space \mathbf{V}_j is reduced to a lower resolution space \mathbf{V}_{j+1} and to a space called the “*detail*” space \mathbf{W}_{j+1} . What means

⁴³ Where $2^L = N^{-1}$.

to divide an orthogonal basis $\{\phi_j(t - 2^j n)\}_{n \in \mathbb{Z}}$ belonging to \mathbf{V}_j into two orthogonal bases (of scaling functions and wavelets):

$$\{\phi_{j+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}} \in \mathbf{V}_{j+1}, \quad \text{and} \quad \{\psi_{j+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}} \in \mathbf{W}_{j+1}. \quad (5.192)$$

In a generic way, a theorem stated by Coifmann, Meyer and Wickerhauser shows that the “conjugate mirror filters” transform an orthogonal basis $\{\theta_j(t - 2^j n)\}_{n \in \mathbb{Z}}$ into two families of orthogonal bases: $\{\theta_{j+1}^0(t - 2^{j+1} n)\}_{n \in \mathbb{Z}}$ and $\{\theta_{j+1}^1(t - 2^{j+1} n)\}_{n \in \mathbb{Z}}$. If we denote \mathbf{U}_{j+1}^0 and \mathbf{U}_{j+1}^1 the spaces associated with these two families, it is said that these spaces are orthogonal and:

$$\mathbf{U}_{j+1}^0 \bigoplus \mathbf{U}_{j+1}^1 = \mathbf{U}_j. \quad (5.193)$$

Consequently, it is possible to establish that $\mathbf{U}_j = \mathbf{W}_j$ and thus to divide these “detail” spaces \mathbf{W}_j into subspaces which have new associated bases. These successive divisions produce (see illustration supra) the binary trees (because division is a division by two), and if the scale for the approximation is dyadic of the type 2^L , it is possible to associate with it the approximation space \mathbf{V}_L and an orthogonal basis of *scaling functions*:

$$\{\phi_L(t - 2^L n)\}_{n \in \mathbb{Z}}, \quad (5.194)$$

(where $\phi_L = 2^{-L/2}\phi(2^{-L}t)$, with $j \geq L$). $j - L \geq 0$ corresponds to the depth of the binary tree. Each junction point (or node) of the binary tree is referred by a couple (j, p) where j is the resolution level and p the “number” of the node starting from the left at the same depth $j - L$. To each one of the node of the tree (indexed by the couples (j, p)) corresponds one space \mathbf{W}_j^p and one basis $\{\psi_j^p(t - 2^j n)\}_{n \in \mathbb{Z}}$. To the starting point of the tree, the space \mathbf{V}_L corresponds to the space \mathbf{W}_j^0 , ($\mathbf{V}_L = \mathbf{W}_j^0$) and the scaling function ϕ_L corresponds to ψ_L^0 . And if one chooses an unspecified division of the tree, one says that the bases $B_{j+1}^{2p} = \{\psi_{j+1}^{2p}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}}$ and $B_{j+1}^{2p+1} = \{\psi_{j+1}^{2p+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}}$ are orthonormal bases of two orthogonal spaces \mathbf{W}_{j+1}^{2p} and \mathbf{W}_{j+1}^{2p+1} , which respect the equality mentioned previously

$$\mathbf{W}_j^p = \mathbf{W}_{j+1}^{2p} \bigoplus \mathbf{W}_{j+1}^{2p+1}. \quad (5.195)$$

Illustration of Tree Structure and Best Basis for a Transient Signal

Tree and Best basis of sinusoids packets space of the signal. Observe in Fig. 5.28 the signal and image of the Heisenberg boxes of Fourier atoms (Fig. 5.28).

Hereafter, in the upper part of the figure on the left: the *sinusoids packets decomposition* within the framework of the *best basis* (level of division: $D = 9 = \log_2(512)$). In the upper part of the figure on the right: the tree of the *best basis of sinusoids packets*. In the lower part on the left: Heisenberg boxes of the Fourier atoms in the time–frequency plane. In the lower right part: *Image of the Heisenberg boxes of Fourier atoms*:

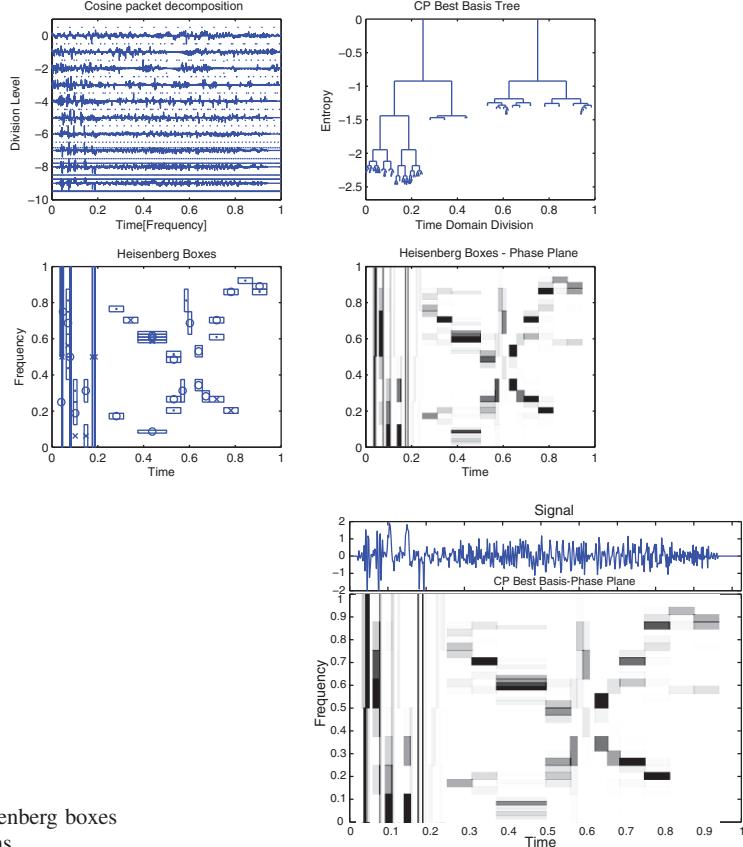


Fig. 5.28 Heisenberg boxes of Fourier atoms

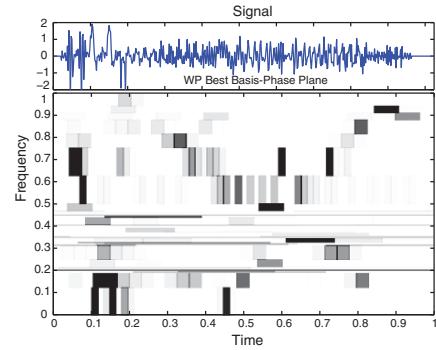


Fig. 5.29 Heisenberg boxes of wavelet atoms

Tree and Best Basis of wavelet packets space of the signal. Observe in Fig. 5.29 the signal and image of the Heisenberg boxes of wavelet atoms (Fig. 5.29).

Hereafter in the upper part on the left: the wavelet packets decomposition of the signal with the best basis (calculated with a Daubechies-12 filter). In the upper right part: Tree of the best basis of wavelet packets. In the lower left part: Heisenberg

boxes of wavelet atoms in the time frequency plane. In the lower right part: Image of the Heisenberg boxes of wavelet atoms:

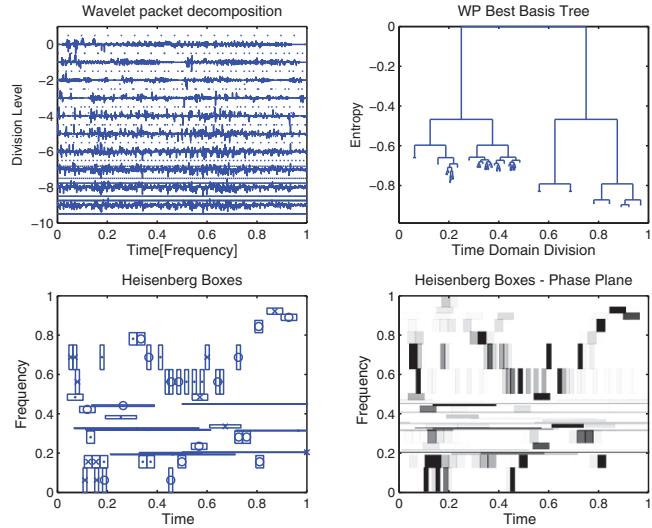


Illustration of Tree Structure and Best Basis of the Stock Exchange Index: Cac40

Tree of the wavelet packet space: The decomposed signal is the French stock exchange index (Cac40) for 2,048 observations of the daily growth rate.

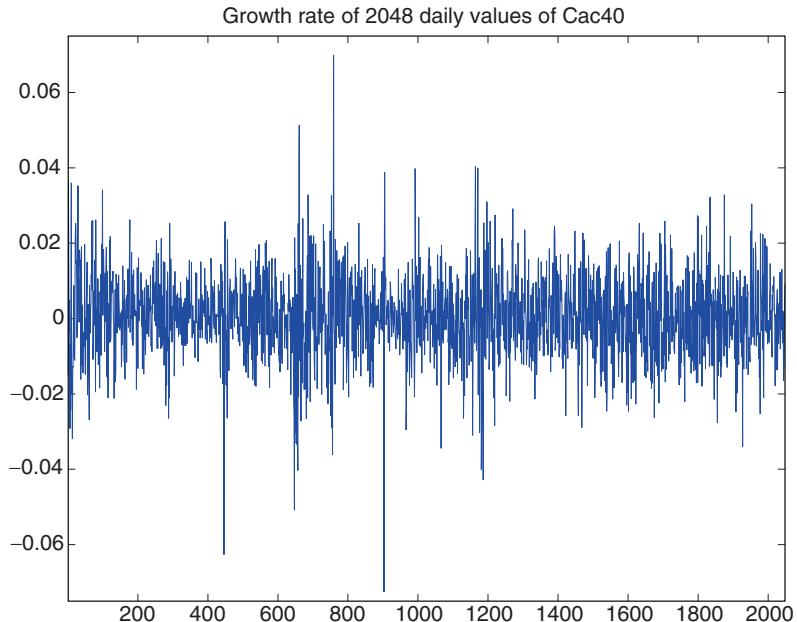
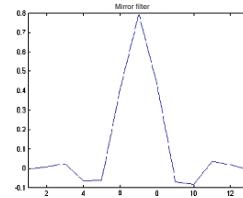
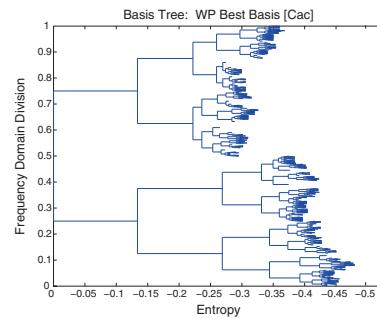


Fig. 5.30 Mirror filter: Coiflet**Fig. 5.31** Wavelet packet tree
(frequency domain division)

For the wavelet packet decomposition, a mirror filter (of the type Coiflet 3) is used, one gives its representation (Fig. 5.30). Then a convolution is operated between the signal and the filter. The calculation of the best basis (6.5285) gives the following tree with a depth level of $D = 11 = \log_2(2,048)$ (Fig. 5.31).

Tree of the sinusoid packet space: The representation of Heisenberg boxes in the time–frequency plane has, in this case, a weak visual interest and thus will not be presented (Fig. 5.32).

5.14.3 Donoho and Johnstone Nonlinear Estimation: Algorithm with Threshold

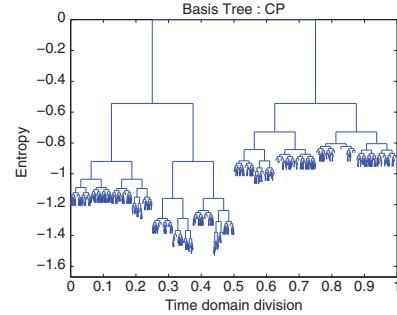
The problem of the estimation is to find a structure in a signal which apparently does not let it show through. Either a signal is really random without any structure, or the structure is hidden due to the presence of a noise which covers it. (It is obvious that the structures of nonlinear origin are more difficult to isolate.) A good part of the debate regarding the *estimation* is related to *the notion of threshold* between *noise* and *signal*. Thus, it is possible to conceive a signal as the addition of a structure and a noise:

$$X = f + \text{noise}. \quad (5.196)$$

The signal f is estimated by the transformation of the disturbed (i.e. noized) signal X by an operator D , one poses \tilde{F} *the estimator of f* :

$$\tilde{F} = DX. \quad (5.197)$$

Fig. 5.32 Fourier basis tree
(time domain division)



The average error provides the risk of the estimator:

$$r(D, f) = E \left\{ \|f - DX\|^2 \right\}. \quad (5.198)$$

The choice of a *nonlinear operator* D is *more effective* to decrease the risk than a *linear operator*. Donoho and Johnstone show that a nonlinear estimator close to the optimality is obtained by means of a *thresholding*:

$$\tilde{F} = DX = \sum_{m=0}^{N-1} \rho_T(\langle X, g_m \rangle) g_m. \quad (5.199)$$

$\rho_T(x)$ is a function with *threshold* T which is worth x for $|x| \geq T$ and 0 for $|x| < T$. The method proceeds to a *smoothing* which is a function of the threshold but also of the selected basis.

5.14.4 Nonlinear Estimators are More Efficient to Minimize the Bayesian Risk: Optimization by Minimax

The estimate of the operator D in the expression described previously $\tilde{F} = DX$ is a vast subject and obviously the objective here is not to be exhaustive. We will describe the current practices in the field, which use in particular the bayesian analysis. See also the (statistics) part II of the present book.

5.14.4.1 Bayesian Approach: A Posteriori Law Specification of the Parameter and Choice of the A Priori Law

A statistical model is a couple $(\mathcal{Y}, \mathbf{P})$, where \mathcal{Y} is the set of possible observations and \mathbf{P} is a family of probability laws on \mathcal{Y} . Such a model is known as dominated, if

all the laws of the family admit a density with respect to a same measure μ , which is called the dominating measure. Given a parametric (dominated) model:

$$[\mathcal{Y}, \mathbf{P} = \{P_\theta = l(y; \theta) \cdot \mu; \theta \in \Theta\}], \quad (5.200)$$

the Bayes approach provides the set Θ of the possible values of the parameter θ of a law Π , this law is known as the *a priori law (or prior law)*, and it is made as if the parameter were random of law Π . Moreover, when we possess the observations y , the “a priori” idea about the parameter is then modified. Indeed, we consider these observations, and have to replace the (*a priori*) marginal law of θ by the conditional law of θ knowing y , which is the “*a posteriori*” law. The transition of the *a priori* law to the *a posteriori* law is expressed by the *Bayes formula*:

$$\pi(\theta|y) = \frac{l(y; \theta)\pi(\theta)}{\int_{\Theta} l(y; \theta)\pi(\theta)v(d\theta)} = \pi(\theta) \frac{l(y; \theta)}{\int_{\Theta} l(y; \theta)}, \quad (5.201)$$

this formula supposes that the *a priori* law admits a density $\pi(\theta)$ with respect to a measure v . Consequently, the *a posteriori* law admits a density $\pi(\theta|y)$ with respect to this measure. The *a posteriori* law is obtained by multiplying the *a priori* density by the ratio of the conditional density of y knowing θ (likelihood of y knowing θ) to the marginal density of y (predictive density). In this approach, the most difficult is to choose the most adapted *a priori* law, even if obviously the calculation of the transition formula of the *a priori* law towards the *a posteriori* law is often also difficult.

In this type of Bayesian approach, there is thus an *a priori* law Π on Θ , and for a statistical test φ , it is possible to define the *bayesian risk* and to compare the tests. The bayesian risk is written:

$$r_{\Pi}(\varphi) = \int_{\Theta} R(\varphi, \theta)d\Pi(\theta). \quad (5.202)$$

The Bayes analysis benefits from the information contained in the signal to calculate an *a posteriori* law. (There is a kind of recursivity in the Bayes approach.) The estimate of the operator D , in the expression described previously $\tilde{F} = DX$, will be optimized if one can benefit from the information that the signal contains. Within the framework of the terminology used in time–frequency analysis, the *risk* is written:

$$r(D, \pi) = E_{\pi}\{r(D, F)\}, \quad (5.203)$$

this risk takes into account the empirical law, i.e. the *probability distribution* π of the signal. An optimal operator which is taken among all the possible estimators D , must provide the *minimum bayesian risk*, which is written:

$$r(\pi) = \inf_D r(D, F). \quad (5.204)$$

These techniques are difficult to use and the choice of the *a priori* law predetermines the results.

5.14.4.2 Minimax and Nonlinear Operators

The framework of the *Game* and *Decision Theory* allows to introduce simpler techniques, where we do not specify the probability distribution. Here are signals modelled as “particular elements of a set Θ ”. In order to control the risk, we evaluate its maximum for any signal belonging to the set Θ . The maximum risk is written:

$$r(D, \Theta) = \sup_{f \in \Theta} r(D, f), \quad (5.205)$$

And the *Minimax risk* is written:

$$r_n(\Theta) = \inf_D r(D, \Theta). \quad (5.206)$$

In a general way, in the field of the possible estimators for D , the nonlinear estimators prove to have a “minimax risk” lower than the linear estimators (except for particular cases of convexity).⁴⁴ Thus, a good operator D will be provided by a nonlinear estimate using the minimax to reach the weakest possible risk.

5.14.5 Approximation by the “Matching Pursuit”: A General Presentation

For many signals made up of complex structures, the *best bases* of local sinusoids or wavelet packets are not the most powerful approximations. As underlined previously, by the slackening of the orthogonality constraint, an approximation of the signal can be produced from M non-orthogonal vectors $\{g_{\gamma_m}\}_{0 \leq m \leq M}$, selected in a redundant dictionary $D = \{g_{\gamma}\}$. And the approximation of the signal was written:

$$f_M = \sum_{m=0}^{M-1} a_m g_{\gamma_m}. \quad (5.207)$$

For a signal length N , the number of vectors of a redundant dictionary is calculated by $P = N \log_2 N$. Thus, the dictionary is constituted of P vectors at the most, which contains at least N linearly independent vectors. These N linearly independent vectors are selected to build a *set of bases*. In addition, note that if a *set of non-orthogonal bases* is constructed, this set usually will be *larger than the set made up from orthogonal bases*. For a dictionary $D = \{g_p\}_{0 \leq p < P}$, whatsoever M (with $N < P$, $0 \leq m \leq M$, and $M \leq N$), the approximation is the result of a linear combination of the M vectors of the dictionary.⁴⁵ And the approximation is written

⁴⁴ For a discussion about the optimality of the minimax within the framework of the time–frequency analysis, we can refer to Mallat (1999, pp. 469–500).

⁴⁵ Ref. to *Theory* concerning the *linear programming*, the *minimax*, and the algorithm of the *simplex* is used in the time–frequency analysis and is also used by *economists*.

in a discrete way:

$$f_M = \sum_{m=0}^{M-1} a[p_m] g_{p_m}. \quad (5.208)$$

The calculations of f_M and the minimization of $\|f - f_M\|$ are heavy and difficult. Mallat and Zhang have developed the version of the Matching Pursuit algorithm which has the property to reduce the calculations by sacrificing the “optimality” to the “efficiency”. Starting from the initial redundant dictionary $D = \{g_\gamma\}$, the algorithm “pursuit”⁴⁶ allows to select N vectors to construct a basis $\{g_{\gamma_m}\}_{0 \leq m < N}$. The matching pursuit is connected with the *pursuit algorithms of projection* developed in *statistics*.

5.14.5.1 The “Matching Pursuit” Structure

Given the dictionary $D = \{g_\gamma(t)\}_{\gamma \in \Gamma}$ of P vectors (endowed with a unit norm such that $\|g_\gamma\| = 1$), which contains N linearly independent vectors defining a basis. Thus, a matching pursuit calculates a linear expansion of f , by successions of approximations of f , by means of orthogonal projections of the signal on elements of D . Given $g_{\gamma_0} \in D$, the vector f can be decomposed as follows:

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + Rf. \quad (5.209)$$

Rf is the *residual vector after approximation of f* in the direction of g_{γ_0} . And g_{γ_0} is orthogonal to Rf and is normalized at 1. Since Rf is orthogonal to g_{γ_0} :

$$\|f\|^2 = |\langle f, g_{\gamma_0} \rangle|^2 + \|Rf\|^2. \quad (5.210)$$

Minimize the residual vector $\|Rf\|$ comes down to choose g_{γ_0} in order to maximize $|\langle f, g_{\gamma_0} \rangle|$. From the point of view of calculations, the most effective is to find a sub-optimal vector g_{γ_0} :

$$|\langle f, g_{\gamma_0} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle f, g_\gamma \rangle|, \quad (5.211)$$

[with $0 \leq \alpha \leq 1$]. The Matching pursuit is an iterative algorithm which with successive stages decomposes the residue Rf of a previous projection.⁴⁷ If we are located at an unspecified moment of the iteration process and if the m th residue $R^m f$ has been calculated, then the following stage is to choose an element g_{γ_m} of the dictionary which approximates the $R^m f$ residue:

$$|\langle R^m f, g_{\gamma_m} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle R^m f, g_\gamma \rangle| \quad (5.212)$$

⁴⁶ Chen and Donoho’s algorithm (Chen and Donoho 1995).

⁴⁷ $R^0 f = f$.

the residue $R^m f$ is decomposed:

$$R^m f = \langle f, g_{\gamma_m} \rangle g_{\gamma_m} + R^{m+1} f, \quad (5.213)$$

Since $R^{m+1} f$ is orthogonal to g_{γ_m} , we can write:

$$\|R^m f\|^2 = |\langle R^m f, g_{\gamma_m} \rangle|^2 + \|R^{m+1} f\|^2. \quad (5.214)$$

We obtain by summation:

$$f = \sum_{m=0}^{M-1} \langle R^m f, g_{\gamma_m} \rangle g_{\gamma_m} + R^M f. \quad (5.215)$$

and then:

$$\|f\|^2 = \sum_{m=0}^{M-1} |\langle R^m f, g_{\gamma_m} \rangle|^2 + \|R^M f\|^2. \quad (5.216)$$

$\|R^m f\|$ converges (Mallat 1998, p.422) exponentially towards 0, when m tends towards the infinite:

$$\lim_{m \rightarrow \infty} \|R^m f\| = 0. \quad (5.217)$$

The approximation carried out by the matching pursuit is largely improved by the orthogonalization of the directions of projection. This is done by means of the Gram–Schmidt procedure (Pati et al. 1993). The *convergence* of the *orthogonal matching pursuit* is carried out with *a finite number of iterations*, whereas it is not the case with the *non-orthogonal matching pursuit*.

Orthogonal Matching Pursuit

In the previous procedure the vectors g_{γ_m} selected by the matching pursuit are a priori non-orthogonal to vectors earlier selected, which are written: $\{g_{\gamma_p}\}$ with $0 \leq p < m$. The previous matching pursuit selected the g_{γ_m} which verified:

$$|\langle R^m f, g_{\gamma_m} \rangle| \geq \alpha \sup_{\gamma \in \Gamma} |\langle R^m f, g_{\gamma} \rangle|. \quad (5.218)$$

In order to distinguish with the former procedure, we pose $\xi_0 = g_{\gamma_0}$. The Gram–Schmidt algorithm orthogonalizes the g_{γ_m} with the $\{g_{\gamma_p}\}_{0 \leq p < m}$ and we pose:

$$\xi_m = g_{\gamma_m} - \sum_{p=0}^{m-1} \frac{\langle g_{\gamma_m}, \xi_p \rangle}{\|\xi_p\|^2} \xi_p, \quad (5.219)$$

Previously $R^m f$ was projected on g_{γ_m} , but now, $R^m f$ is projected on ξ_m :

$$R^m f = \frac{\langle R^m f, \xi_m \rangle}{\|\xi_m\|^2} \xi_m + R^{m+1} f. \quad (5.220)$$

By summation on m , we obtain:

$$f = \sum_{m=0}^{k-1} \frac{\langle R^m f, \xi_m \rangle}{\|\xi_m\|^2} \xi_m + R^k f \quad (5.221)$$

$$= P_{V_k} f + R^k f. \quad (5.222)$$

The $\{\xi_m\}_{0 \leq m < k}$ generates the space V_k , and P_{V_k} is an orthogonal projector on V_k . For $m = k$, we have:

$$\langle R^m f, \xi_m \rangle = \langle R^m f, g_{\gamma_m} \rangle, \quad (5.223)$$

and it exists M such that $f \in V_M$, thus $R^M f = 0$. And for $k = M$ by combining the preceding expression with the equation of f , we have:

$$f = \sum_{m=0}^{k-1} \frac{\langle R^m f, g_{\gamma_m} \rangle}{\|\xi_m\|^2} \xi_m. \quad (5.224)$$

In the preceding section, we have introduced the convergence of the orthogonal matching pursuit, this convergence is done at the end of M iterations. The square of the norm of f is written as the following sum:

$$\|f\|^2 = \sum_{m=0}^{M-1} \frac{|\langle R^m f, g_{\gamma_m} \rangle|^2}{\|\xi_m\|^2}. \quad (5.225)$$

Finally, in order to develop f on the initial dictionary $\{g_{\gamma_m}\}_{0 \leq m < M}$, we write:

$$\xi_m = \sum_{p=0}^m b[p, m] g_{\gamma_p}$$

which gives:

$$f = \sum_{m=0}^{M-1} a[\gamma_p] g_{\gamma_p}, \quad (5.226)$$

with:

$$a[\gamma_p] = \sum_{m=p}^{M-1} b[p, m] \frac{\langle R^m f, g_{\gamma_m} \rangle}{\|\xi_m\|^2}. \quad (5.227)$$

Hereafter, we provide the representations in the time–frequency plane of both decompositions of a transitory signal (transient).

Figure 5.33 corresponds to the decomposition of the signal by the matching pursuit with dictionaries of windowed Fourier atoms. One will notice the “spread” of each atom in the time–frequency plane.

Figure 5.34 corresponds to the decomposition of the same signal by the matching pursuit with dictionaries of wavelet atoms. Observe the “spread” and the localization of atoms compared with Fig. 5.33.

Fig. 5.33 Heisenberg boxes of Fourier atoms

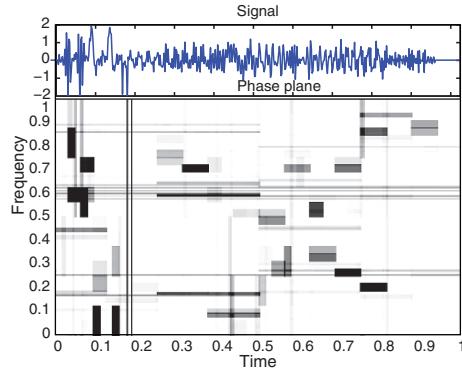
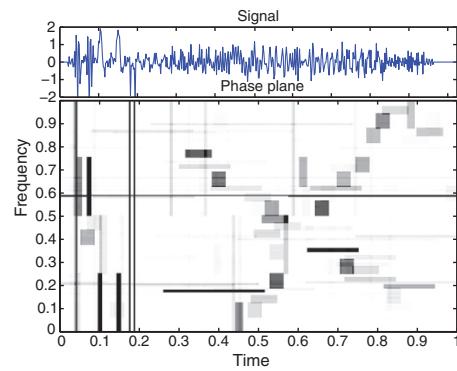


Fig. 5.34 Heisenberg boxes of wavelet atoms



Gabor Atom, Function and Dictionary

An improvement of the Matching Pursuit was made by Mallat and Zhang, as well as by Qian and Chen by means of the Gabor dictionary. This is by modulating, relocating and spreading (i.e. scaling) a Gaussian window which is used for its *qualities of energy (and frequency) distribution*.⁴⁸ The window is written:

$$g(t) = 2^{1/4} e^{-\pi t^2}, \quad (5.228)$$

in a discrete mode for a period denoted N and a parameter denoted α_j , this window is sampled and modulated on a scale 2^j , and becomes:

$$g_j[n] = \alpha_j \sum_{p=-\infty}^{+\infty} g\left(\frac{n-pN}{2^j}\right), \quad (5.229)$$

$\|g_j\| = 1$ is obtained by means of α_j . Then, the window is translated in time and frequency via an index denoted:

$$\gamma = (p, k, 2^j) \quad (5.230)$$

⁴⁸ See proof by the Heisenberg theorem.

with j belonging to $[0, \log_2 N]$ and (p, k) belonging to $[0, N - 1]^2$. We obtain a Gabor atom:

$$g_\gamma[n] = g_j[n - p] e^{\frac{2\pi i k n}{N}}, \quad (5.231)$$

We denote $D = \{g_\gamma\}$ the Gabor dictionary of these Gabor atoms. More precisely, the matching pursuit uses the following set of indexes:

$$\gamma_\pm = (p, \pm k, 2^j) \quad (5.232)$$

and provides groups of complex Gabor atoms: g_{γ_-} and g_{γ_+} (also called *polarized* atoms). In this case and for a real signal, the matching pursuit is a projection of $R^m f$ on:

$$g_\gamma^\Omega[n] = \alpha_{j, \Omega} g_j[n - p] \cos\left(\frac{2\pi i k n}{N} + \Omega\right). \quad (5.233)$$

The approximation of the signal is thus written as the following decomposition:

$$f = \sum_{m=0}^{+\infty} \left\langle R^m f, g_{\gamma_m}^{\Omega_m} \right\rangle g_{\gamma_m}^{\Omega_m}, \quad (5.234)$$

which is represented by an *energy distribution* resulting from the addition of the Wigner–Ville density of complex atoms:

$$P_M f[n, k] = \sum_{m=0}^{+\infty} \left| \left\langle R^m f, g_{\gamma_m}^{\Omega_m} \right\rangle \right|^2 \mathbf{D} g_{\gamma_m}[n, k]. \quad (5.235)$$

(This matching pursuit version will be presented in detail later on in another section.) Hereafter, one provides the representation of the decomposition in the time–frequency plane of the (same) transient signal by Matching Pursuit with Gabor atoms. It will be noticed the shape, the spread and the localization of atoms in the time–frequency plane (Fig. 5.35).

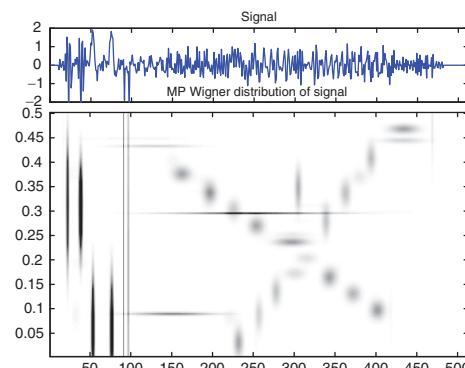


Fig. 5.35 Wigner–Ville distribution of Gabor atoms

5.14.6 Comparison of Best Bases and Matching Pursuits

The presentations of approximation methods of the Best Basis and Matching Pursuits previously made have been illustrated by means of the same signal that is of transient type (i.e. intermittent with some Diracs), this one now makes it possible to compare the results in Fig. 5.36.

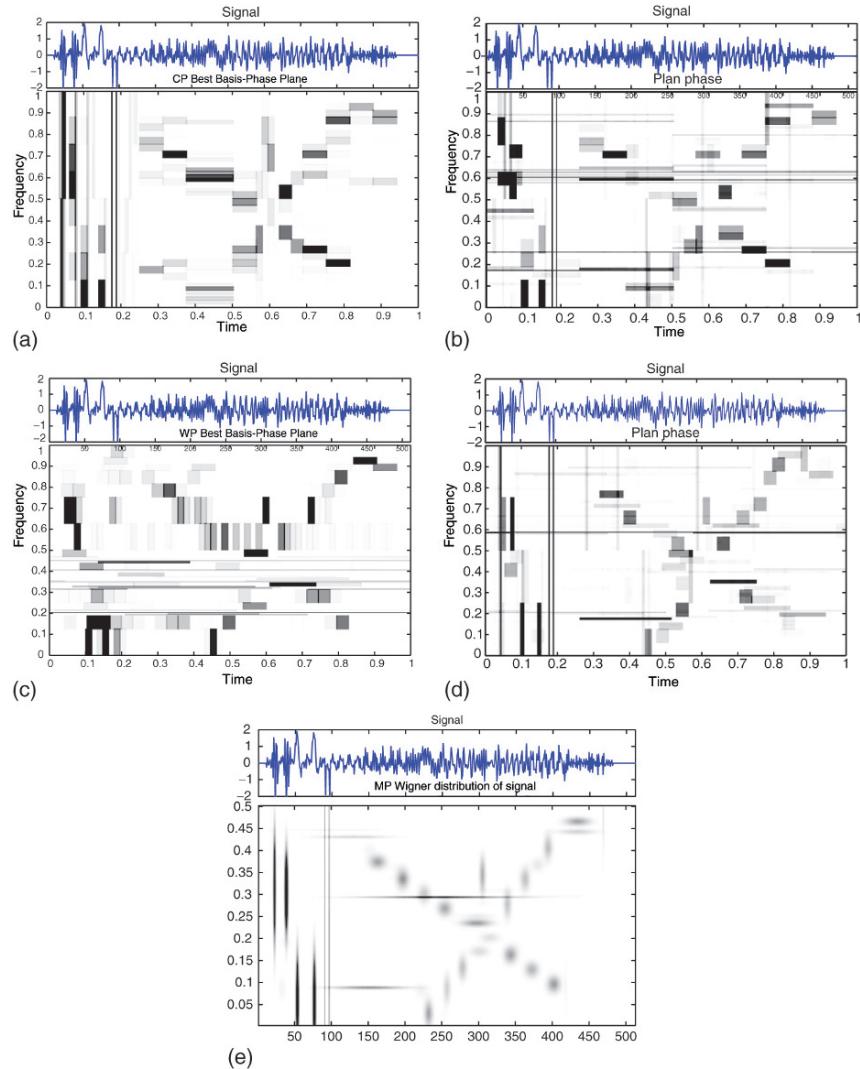


Fig. 5.36 (a) Best basis and Fourier, (b) matching pursuit and Fourier, (c) best basis and wavelets, (d) matching pursuit and wavelets, (e) matching pursuit and Gabor atoms

Although the parameter setting (choice of filter mirror for the wavelets, split level, etc.) in the construction of approximations, plays an important role and can change the quality of representations, it is easy to observe the evolution of approximation methods towards the latest method presented in Fig. 5.36e.

5.15 The Multiresolution Analysis Notion

It is obviously necessary to present the multiresolution analysis (MRA: Multiresolution analysis) whose first corollary is the discrete wavelet transform (DWT). Beforehand, we quickly introduce the notion of (quadratic) *conjugate mirror filter* which is used in the construction of fast algorithms of the wavelet transform.

5.15.1 (Quadratic) Conjugate Mirror Filter

This type of filter is used in various applications and gave rise to the development of a true theory. In particular for the construction of fast algorithms of calculations of the (orthogonal) wavelet transforms and the improvement of methods of signal reconstruction. It contributes also to the construction of the algorithms of the multiresolution analysis (MRA). Its role is important, it is perfectly presented and developed in the works of S. Mallat. The presentation of the history and the construction of these filters allows to approach the multiresolution analysis in an interesting way. It will be simply noted here that the fast algorithms of wavelet transforms calculations decompose the signals into *low-pass* and *high-pass components* which are sub-sampled in two subsets. The signal is filtered by a low-pass filter denoted $\bar{h}[p] = h[-p]$ and a high-pass filter $\bar{g}[p] = g[-p]$, these filters make it possible by convolution to sub-sample a signal f into two outputs:

$$a_1[p] = f * \bar{h}[2p] \quad \text{and} \quad d_1[p] = f * \bar{g}[2p]. \quad (5.236)$$

In fact, such a transformation decomposes each approximation $P_{\mathbf{V}_j}f$ into a coarse approximation $P_{\mathbf{V}_{j+1}}f$ and into wavelet coefficients $P_{\mathbf{W}_{j+1}}f$. During the reconstruction the $P_{\mathbf{V}_j}f$ are “re-found” from $P_{\mathbf{V}_{j+1}}f$ and $P_{\mathbf{W}_{j+1}}f$. It is known that:

$$\{\phi_{j,n}\}_{n \in \mathbb{Z}} \quad \text{and} \quad \{\psi_{j,n}\}_{n \in \mathbb{Z}} \quad (5.237)$$

are orthonormal bases of \mathbf{V}_j and \mathbf{W}_j . Furthermore, it is known that the projections on these bases are defined by:

$$a_j[n] = \langle f, \phi_{j,n} \rangle \quad \text{and} \quad d_j[n] = \langle f, \psi_{j,n} \rangle. \quad (5.238)$$

These coefficients are calculable by means of a cascade of discrete convolutions and sub-samplings. If we pose that $\bar{x}[n] = x[-n]$ and:

$$\hat{x}[n] = \begin{cases} x[p] & \text{if } n = 2p, \\ 0 & \text{if } n = 2p + 1, \end{cases} \quad (5.239)$$

then, it is known that *a signal is decomposable* into:

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_j[n] = a_j \star \bar{h}[2p], \quad (5.240)$$

$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g[n-2p]a_j[n] = a_j \star \bar{g}[2p], \quad (5.241)$$

and the *signal reconstruction* is written:

$$a_j[p] = \sum_{n=-\infty}^{+\infty} h[n-2p]a_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[n-2p]d_{j+1}[n] \quad (5.242)$$

$$= \hat{a}_{j+1} \star \bar{h}[p] + \hat{d}_{j+1} \star \bar{g}[p]. \quad (5.243)$$

The *mirror filters* satisfy the following (Mallat and Meyer) quadratic condition:

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega + \pi)|^2 = 2 \quad \text{and} \quad \hat{h}(0) = \sqrt{2}, \quad (5.244)$$

where ϕ is an integrable scaling function belonging to $L^2(\mathbb{R})$, where $h[n] = \langle 2^{-1/2}\phi(t/2), \phi(t-n) \rangle$ and \hat{h} is its Fourier transform $\forall \omega \in \mathbb{R}$. The mirror filters $h[n]$ are of different types, we will mention without presenting them the following types: “Haar”, “Beylkin”, “Coiflet”, “Daubechies”, “Symmlet”, “Vaidyanathan”, “Battle”.

5.15.2 Multiresolution Analysis

Definition 5.3 (Multiresolution). A sequence⁴⁹ $\{\mathbf{V}_j\}_{j \in \mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multiresolution approximation, if the six following properties are verified:

$$\forall (j, k) \in \mathbb{Z}^2, f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j, \quad (5.245)$$

$$\forall j \in \mathbb{Z}, \mathbf{V}_{j+1} \subset \mathbf{V}_j, \quad (5.246)$$

$$\forall j \in \mathbb{Z}, f(t) \in \mathbf{V}_j \Leftrightarrow f\left(\frac{t}{2}\right) \in \mathbf{V}_{j+1}, \quad (5.247)$$

$$\lim_{j \rightarrow +\infty} \mathbf{V}_j = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_j = \{0\}, \quad (5.248)$$

⁴⁹ \mathbb{Z} : Space of the (positive and negative) integers, $\mathbb{Z} = \{\dots, -3, -2, -1, 0, +1, +2, +3, \dots\}$ which is an extension of the space of the integers \mathbb{N} .

$$\lim_{j \rightarrow -\infty} \mathbf{V}_j = \text{Closure} \left(\bigcup_{j=-\infty}^{+\infty} \mathbf{V}_j \right) = \mathbf{L}^2(\mathbb{R}), \quad (5.249)$$

$$\text{There exists } \theta \text{ such that } \{\theta(t-n)\}_{n \in \mathbb{Z}} \text{ is Riesz basis of } \mathbf{V}_0. \quad (5.250)$$

These properties are interpreted in the following way:

- \mathbf{V}_{j+1} is the image of \mathbf{V}_j by a dilation of a factor 2: There exists a subjacent frequential grid in geometric progression.
- For any j , \mathbf{V}_{j+1} is a sub-space of \mathbf{V}_j .
- \mathbf{V}_j is invariant by translation of 2^j : There exists a subjacent temporal grid by step of 2^j .
- The intersection of \mathbf{V}_j is reduced to 0 in \mathbf{L}^2 : At a minimal resolution, one loses all the image. This hypothesis is formulated by convention, because it is always verified.
- The union of the \mathbf{V}_j is dense in \mathbf{L}^2 : At an infinite resolution, any signal is perfectly reproduced.
- There exists a function θ such that the whole translations of θ form a Riesz basis of \mathbf{V}_0 : Each resolution is generated by a basis of atoms translated of 2^j . A Riesz basis is a frame of independent vectors.

The dilated and translated functions $\phi_{j,k}(t)$:

$$\{\phi_{j,k}(t) = 2^{-j/2}\phi_0(2^{-j}t-k)\}_{k \in \mathbb{Z}} \quad (5.251)$$

constitute also a Riesz basis for \mathbf{V}_0 . It is fundamental to understand that to carry out a multiresolution analysis of a time series $x(t)$, this is successively to project the latter on each approximation subspaces V_j :

$$\text{Approx}_j(t) = \text{Proj}_{V_j}x(t) = \sum_k a_{j,k}\phi_{j,k}(t). \quad (5.252)$$

As presented previously $V_j \subset V_{j-1}$, consequently we know that the approximation: $\text{approx}_j(t)$ of the series $x(t)$ is an approximation coarser than $\text{approx}_{j-1}(t)$. Thus, another fundamental idea of the multiresolution analysis (MRA), consists in the analysis of the information loss which is indicated by the term: “*detail*_j”, it expresses the information loss related to an increasingly coarse approximation, as follows:

$$\text{Detail}_j(t) = \text{approx}_j(t) - \text{approx}_{j-1}(t). \quad (5.253)$$

The multiresolution analysis (MRA) computes the “*detail*_j” by projections of $x(t)$ on a collection of subspaces noted W_j , called *the wavelet subspaces*. Furthermore, the multiresolution analysis proves that there is a function ψ_0 , called *mother wavelet* (built as the scaling function ϕ_0) which is used to construct the functions:

$$\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t-k)\} \quad (5.254)$$

which constitute also a Riesz basis for W_j :

$$\text{Detail}_j(t) = \text{Proj}_{W_j}x(t) = \sum_k d_{j,k}\psi_{j,k}(t). \quad (5.255)$$

The principle of the multiresolution analysis is to transform the information contained in the time series $x(t)$ in a collection, which gathers on the one hand, an approximation with low resolution J , and on the other hand, the set of “*detail* _{j ” at different resolution $j = 1, \dots, J$ (noted $\sum_{j=1}^J \text{detail}_j$). Then it is possible to write:}

$$x(t) = \text{approx}_J + \sum_{j=1}^J \text{detail}_j(t) = \sum_k a_{J,k}\phi_{J,k}(t) + \sum_{j=1}^J \sum_k d_{j,k}\psi_{j,k}(t). \quad (5.256)$$

In order to construct approx_J which is a coarse representation at a low resolution of the series, it is necessary that the *scaling function* ϕ_0 (from which we calculate $\phi_{J,k}$ and $\sum_k a_{J,k}\phi_{J,k}(t)$) is a *low-pass filter*. As regards the terms “*detail* _{j ” which correspond to the differences $\text{approx}_j(t) - \text{approx}_{j-1}(t)$, the function ψ_0 , which enters in their construction, is a *band-pass filter*. The *band-pass filter* (which in spite of its source because it historically belongs to the Fourier analysis) corresponds also to a *wavelet* through its form.}

Scaling function: Within the general framework of the signal analysis, it is known that a *scaling function* is interpreted as the *impulse response* of a *low-pass filter*. From the gauge of a generic filter band-pass, we construct filters of the low-pass and high-pass type, which filter by means of their specific forms the Fourier coefficients, either from the high frequencies towards the low frequencies or from the low frequencies towards the high frequencies. If we note $\phi_s(t) = \frac{1}{\sqrt{s}}\phi\left(\frac{t}{s}\right)$ (and with $\bar{\phi}_s(t) = \phi_s^*(-t)$), then the *low frequency approximation* of a time series $x(t)$ at the scale s is written: $Lx_{u,s} = \left\langle x(t), \frac{1}{\sqrt{s}}\phi\left(\frac{t-u}{s}\right) \right\rangle = x \star \phi_s^*(u)$.

Definition of a scaling function: As presented previously, the approximation of a time series $x(t)$ at the resolution 2^j is defined as an orthogonal projection of the time series $x(t)$ on \mathbf{V}_j : $\text{Proj}_{\mathbf{V}_j}x(t)$. To calculate this projection, we have to be endowed with a basis of \mathbf{V}_j . The Multiresolution theory says that it is possible to construct an orthonormal basis from each space \mathbf{V}_j by *dilation* and *translation* of a simple function ϕ_0 called the *scaling function*. A *scaling function* can be defined coarsely as an aggregation of wavelets at scales higher than 1.

Approximation of coefficients: The projection of $x(t)$ on \mathbf{V}_j is obtained by a progressive dilation of the scale:

$$\text{Proj}_{\mathbf{V}_j}x = \sum_{k=-\infty}^{+\infty} \langle x, \phi_{j,k} \rangle \phi_{j,k}. \quad (5.257)$$

The inner products $a_j[k] = \langle x, \phi_{j,k} \rangle$ provide a discrete approximation at the scale 2^j , from which we can rewrite them as the result of a convolution: $a_j[k] = \int_{-\infty}^{+\infty} x(t)$

$\frac{1}{\sqrt{2^j}} \phi\left(\frac{t-k}{2^j}\right) dt$, or $a_j[k] = \int_{-\infty}^{+\infty} x(t) 2^{-j/2} \phi_0(2^{-j}t - k) dt$. The discrete approximation $a_j[k]$ is *low-pass filtering* of $x(t)$ sampled on the interval 2^j . We express the $a_{J,k}$ as follows:

$$a_{J,k} = \int x(t) 2^{-j/2} \phi_0(2^{-j}t - k) dt. \quad (5.258)$$

By analogy, we obtain the coefficients $d_{j,k}$ no more by means of a scaling function of the low-pass filter type, but by a band-pass filter, which also corresponds to a wavelet built from a mother wavelet ψ_0 (note this correspondence between filter band-pass resulting from the Fourier analysis and the wavelet analysis). The projection of $x(t)$ on \mathbf{W}_j is obtained by the filter of the wavelet function: $2^{-j/2} \psi_0(2^{-j}t - k)$. The inner products $\langle x, \psi_{j,k} \rangle$ provide an approximation at the scale 2^j . The coefficient at the scale 2^{-j} is written:

$$d_{j,k} = \int x(t) 2^{-j/2} \psi_0(2^{-j}t - k) dt. \quad (5.259)$$

The index j corresponds at the resolution level of the wavelet analysis, j is also called octave. The octave j is the logarithm of the scale 2^j , and k plays the role of time.⁵⁰ n is the length of the time series. The number of available coefficients at the octave j is denoted $n_j = 2^{-j}n$.

Discrete Wavelet Transform: The discrete non redundant transformation allows the transition from a Hilbert space $\mathbf{L}^2(\mathbb{R})$ constructed on the real numbers to the Hilbert space $\mathbf{l}^2(\mathbb{Z})$ constructed on the integer numbers \mathbb{Z} .⁵¹ Provided with a scaling function ϕ_0 and a mother wavelet ψ_0 , the discrete transform products from the time series $x(t)$ the set $\{ \{a_{J,k}\}_{k \in \mathbb{Z}}, \{d_{j,k}, j = 1, \dots, J\}_{k \in \mathbb{Z}} \}$. We observed previously that these coefficients are the result of the inner products $\langle x, \phi_{j,k} \rangle$ and $\langle x, \psi_{j,k} \rangle$ and where $\phi_{j,k}$ and $\psi_{j,k}$ are obtained by dilation and translation respectively of the gauge of a scaling function ϕ_0 and of a mother wavelet ψ_0 . The algorithm used to produce the discrete wavelet transform is the recursive pyramidal algorithm (APR).

General remark about the notations: It is noted that \mathbf{W}_j is a closed subspace generated by an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$ (i.e. here a family of wavelets at the scale 2^{-j}): $\{\psi_{j,k}(t) = 2^{-j/2} \psi_0(2^{-j}t - k)\}_{j \in \mathbb{Z}}$. We can write:

$$\mathbf{L}^2(\mathbb{R}) = \bigoplus_{j \in \mathbb{Z}} \mathbf{W}_j. \quad (5.260)$$

The analysis of a time series by scale level is carried out through the introduction of the approximation spaces noted \mathbf{V}_j , which are defined by:

$$\mathbf{V}_{l+1} = \bigoplus_{j \leq l} \mathbf{W}_j, \quad l \in \mathbb{Z}. \quad (5.261)$$

⁵⁰ The dyadic wavelet transforms are of large interest (they are discrete in scale but continuous in time) because they make it possible not to sub-sample the signals when we pass on to coarse scales and thus not to deteriorate a signal with low resolutions.

⁵¹ Generic writing of the DWT: $Wf[n, a^j] = \sum_{n=0}^{N-1} f[m] \frac{1}{\sqrt{a^j}} \psi_j^*[m - n]$ with a wavelet $\psi_j[n] = \frac{1}{\sqrt{a^j}} \psi_j\left(\frac{n}{a^j}\right)$.

5.16 Singularity and Regularity of a Time Series: Self-Similarities, Multifractals and Wavelets

We had the opportunity to underline several times the property of wavelets to provide information simultaneously in time and frequency. This ability to locate in frequency the events is interesting. The *singularities* and the *irregularities* of a signal provide information as fundamental as the regularities or periodicities for example. We know that a (positive or negative) *peak very localized* in a *financial series* or in an *electrocardiogram contains fundamental information*. Mallat (1998, p. 163) textually highlights that “the local signal regularity is characterized by the decay of the wavelet transform amplitude across scales”. Moreover, he underlines that “non-isolated singularities appear in complex signals such as multifractals”. Multifractals are subjacent in numerous signals of natural origin. “The wavelet transform takes advantage of multifractals self-similarities” in order to calculate the distribution of their singularities. *Thus, the time-scale analysis (with its singularity spectrum) and the wavelets contribute to the knowledge of the properties of multifractals.*

In order to present the *singularity* and the *regularity* of a signal it is necessary before to point out *the Lipschitz conditions*, which were already presented in the part I. With this intention, as a preliminary, we present the Taylor approximation formula and the associated approximation error. Given the function f which is n times differentiable on a interval centered at x_0 : $[x_0 - h, x_0 + h]$, the Taylor formula is written:

$$p_{x_0}(t) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!} (t - x_0)^k. \quad (5.262)$$

And the approximation error $\varepsilon(t) = f(t) - p_{x_0}(t)$ verifies with $u \in [x_0 - h, x_0 + h]$:

$$\forall t \in [x_0 - h, x_0 + h], |\varepsilon(t)| \leq \frac{|t - x_0|}{n!} \sup |f^n(u)|, \quad (5.263)$$

The n th order differentiability of f around x_0 provides the upper bound of the error $\varepsilon(t)$ when $t \rightarrow x_0$. At this stage, *the Lipschitz exponent allows to specify this upper bound.*

5.16.1 Lipschitz Exponent (or Hölder Exponent): Measurement of Regularity and Singularity by Means of the Hölder Functions $\alpha(t)$

The *Fourier analysis* makes it possible to characterize the *global regularity* of a function and the *wavelet transform* makes it possible to analyze the *pointwise regularity*. This distinction exists also in the *definition of the regularity in the Lipschitz sense* (also called Hölder regularity), we speak of *pointwise or uniform regularity*. This *Lipschitz or Hölder regularity concept* was extended recently to more robust approaches such as the *regularity 2-microlocal* (not presented here). Moreover, there

are multiple manners of carrying out a *fractal analysis* of a signal, *the calculation of the pointwise regularity is one of them*, and the *multifractal analysis* is another one also. In the first case, we associate with a signal $f(t)$ another signal $\alpha(t)$, which is called the Hölder function of f , which measures the regularity of f at each point t . The *regularity* of f can be evaluated by means of different methods, the *pointwise Lipschitz (or Hölder) exponent* α of f at x_0 , is one of those methods, another one is the *local exponent* which is often written $\alpha_L(x_0)$. But let us present the initial definition of the regularity in the Lipschitz sense.

Definition 5.4 (Regularity in the Lipschitz sense).

- A function f is pointwise Lipschitz $\alpha \geq 0$ at x_0 , if there exist $K > 0$ and a polynomial p_{x_0} of degree n , where n is equal to the largest integer $n \leq \alpha$ such that:

$$\forall t \in \mathbb{R}, |f(t) - p_{x_0}(t)| \leq K|t - x_0|^\alpha. \quad (5.264)$$

- A function f is uniformly Lipschitz- α over $[a, b]$, if it satisfies the preceding condition for all $x_0 \in [a, b]$, with a constant K which is independent of x_0 .
- The Lipschitz regularity of the function f at x_0 or over $[a, b]$ is the sup of the α such that f is Lipschitz- α .

If f is n time continuously differentiable (with n equal to the largest integer) in a neighborhood of x_0 , then the polynomial p_{x_0} is the Taylor approximation of f at x_0 .⁵²

It is remarkable to note that “it is possible to construct multifractal functions with non-isolated singularities”, for which the signal has a different Lipschitz regularity at each point. However, the uniform Lipschitz exponents provide a more global measure of regularity on the entire interval. With a signal f uniformly Lipschitz $\alpha > n$ in the neighborhood of x_0 , it is necessary to verify that f is n times continuously differentiable. Otherwise, i.e. for a f non-differentiable, α will characterize the type of singularity. We will note simply that if the condition: f is n times continuously differentiable is verified for $0 \leq \alpha < 1$, then $p_{x_0}(t) = f(x_0)$ and the Lipschitz condition $\forall t \in \mathbb{R}, |f(t) - p_{x_0}(t)| \leq K|t - x_0|^\alpha$ becomes:

$$\forall t \in \mathbb{R}, |f(t) - f(x_0)| \leq K|t - x_0|^\alpha. \quad (5.265)$$

The main idea could be summarized as follows: *For the pointwise regularity we associate a signal $f(t)$ with another signal $\alpha(t)$ which is the Hölder function of f and which measures the regularity of f at each point t .*

This *pointwise regularity* is evaluated by means of:

1. The *pointwise Lipschitz exponent* α of f at x_0 , which can be defined as follows:

$$\alpha(x_0) = \limsup_{\rho \rightarrow 0} \{\alpha : \exists K > 0, |f(t) - f(x_0)| \leq K|t - x_0|^\alpha, |t - x_0| < \rho\}$$

if α is non-integer and if f is non-differentiable. Otherwise, it is necessary to replace in the expression above the $f(x_0)$ term by a polynomial $p_{x_0}(t)$.

⁵² For each x_0 the polynomial is unique.

2. The *local exponent* $\alpha_L(x_0)$ such that:

$$\alpha_L(x_0) = \limsup_{\rho \rightarrow 0} \{ \alpha : \exists K > 0, |f(t) - f(y)| \leq K|t - y|^\alpha, |t - x_0| < \rho, |y - x_0| < \rho \}.$$

It will be noted that α and α_L have different properties and usually do not correspond,⁵³ we give the following example, if we consider the signal f as a function such that: $f(t) = |t|^\alpha \sin(1/|t|^\gamma)$, then $\alpha(0) = \alpha$ and $\alpha_L(0) = \alpha/(1+\gamma)$.

Generally the value of $\alpha(t)$ give us the following indications about the *regularity* and the *continuity* of f :

- If $\alpha(t)$ is small, then the function (or signal) f is *irregular* at t .
- If $\alpha(t) > 1$: f is *at least differentiable once* at t .
- If $\alpha(t) < 0$: there is a *discontinuity* on f .

The Lipschitz (or Hölder) exponent allows to *intuitively* represent the *regularity concept* and to *characterize a time-series* by means of their Hölder regularities. It is used in particular in the *signal processing to study the “turbulence phenomena”*. These methods are of a strong *interest* when the *irregularities* of a signal contain *important information*, as it is frequent besides.

5.16.2 *n* Wavelet Vanishing Moments and Multiscale Differential Operator of Order *n*

In order to study and measure the local regularity of a function or a signal, it is fundamental to use the wavelet vanishing moments. (Recall: A wavelet ψ has N vanishing moments if $\int t^k \psi(t) dt = 0$, $0 \leq k < N$, and $\int t^N \psi(t) dt \neq 0$.) Indeed, the wavelet transform of a signal which has n vanishing moments is interpreted as a multiscale differential operator of order n .⁵⁴

The Lipschitz condition presented previously ($\forall t \in \mathbb{R}$, $|f(t) - p_{x_0}(t)| \leq K|t - x_0|^\alpha$) makes it possible to approximate f by a polynomial p_{x_0} in the neighborhood of x_0 : $f(t) = p_{x_0}(t) + \varepsilon(t)$ with $|\varepsilon(t)| \leq K|t - x_0|^\alpha$. However, a *wavelet transform estimates the exponent* α ignoring the polynomial p_{x_0} . In order to do this, a wavelet is selected which has $n > \alpha$ vanishing moments: $\int t^k \psi(t) dt = 0$, $0 \leq k < n$. A wavelet with n vanishing moments is orthogonal to the polynomials of degree $n-1$. And because $n > \alpha$, the polynomial p_{x_0} has a degree at most equal to $n-1$. If one carries out a change of variable you $t' = (t - u)/s$, we verify the following transform:

$$Wp_{x_0}(u, s) = \int p_{x_0}(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt = 0. \quad (5.266)$$

because $f = p_{x_0} + \varepsilon$: $Wp_{x_0}(u, s) = W\varepsilon(u, s)$.

⁵³ **Stable by differentiation:** One of the properties which is respected by α_L (and which is not respected by α) is the property *to be stable by differentiation* $\alpha_L(f', x_0) = \alpha_L(f, x_0) - 1$.

⁵⁴ See in the appendix the *definition of the differentiable operators in Banach spaces*.

The transformation of a signal by a wavelet which has n vanishing moments is interpreted as a (multiscale) differential operator of order n . There is a relation between the differentiability of f and the decay of its wavelet transform at the fine scales. And it is possible to write that a wavelet with n vanishing moments can be written as the derivative of order n of a function θ .

Theorem 5.5 (Vanishing moments of wavelet with a fast decay). A wavelet ψ with a fast decay has n vanishing moments if and only if there exists θ with a fast decay such that:

$$\psi(t) = (-1)^n (d^n \theta(t)/dt^n). \quad (5.267)$$

Consequently:

$$Wf(u, s) = s^n \frac{d^n}{dt^n} (f * \bar{\theta}_s)(u), \quad \text{with } \bar{\theta}_s(t) = (\theta(-t/s)/\sqrt{s}). \quad (5.268)$$

And ψ have no more than n vanishing moments if and only if $\int_{-\infty}^{+\infty} \theta(t) dt \neq 0$.

5.16.3 Regularity Measures by Wavelets

In the prolongation of what was presented above, the decay of the *amplitude* of the wavelet transform along the scales is associated with the uniform and pointwise Lipschitz regularities of a signal. To measure this decay comes down to observe the structure of the signal while varying the scale. If we choose a n times differentiable wavelet with n vanishing moments, it comes: $\exists C_\gamma$, with $\gamma \in \mathbb{N}$, $\forall t \in \mathbb{R}$, $|\psi^{(k)}(t)| \leq C_\gamma / (1 + |t|^\gamma)$.

Theorem 5.6 (Lipschitz uniform). If $f \in L^2(\mathbb{R})$ is uniformly Lipschitz $\alpha \leq n$ over $[a, b]$, then there exists $A > 0$ such that:

$$\forall (u, s) \in [a, b] \times \mathbb{R}^+, |Wf(u, s)| \leq As^{\alpha+(1/2)}. \quad (5.269)$$

Reciprocally, let us suppose that f is bounded and that $Wf(u, s)$ satisfy the condition above for an $\alpha < n$ which is not an integer, then f is uniformly Lipschitz α over $[a + \varepsilon, b - \varepsilon]$ for any $\varepsilon > 0$.

Theorem 5.7 (Jaffard). If $f \in L^2(\mathbb{R})$ is Lipschitz $\alpha \leq n$ at x_0 , then there exists A such that: $\forall (u, s) \in \mathbb{R} \times \mathbb{R}^+$,

$$|Wf(u, s)| \leq As^{\alpha+(1/2)}(1 + |(u - x_0)/s|^\alpha). \quad (5.270)$$

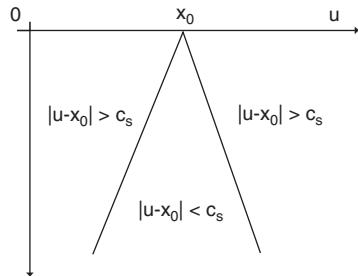
Reciprocally, if $\alpha < n$ is not an integer and there exist A and $\alpha' < \alpha$ such that $\forall (u, s) \in \mathbb{R} \times \mathbb{R}^+$,

$$|Wf(u, s)| \leq As^{\alpha+(1/2)}(1 + |(u - x_0)/s|^{\alpha'}), \quad (5.271)$$

then f is Lipschitz α at x_0 .

This condition establishes a relationship between the pointwise regularity of a signal and the decay of the modulus of its wavelet transform. To give an illustration of the condition above we can represent the cone of influence of a point x_0 . If we consider a mother wavelet with compact support $[-C, C]$, we obtain by modulations of scales $\psi_{(u,s)} = \psi((t-u)/s)/\sqrt{s}$, and the modulations on the compact support provide $\psi((t-u)/s) : [u - Cs, u + Cs]$.

The cone of influence of x_0 is written $|u - x_0| \leq Cs$. If u is in the cone of influence of x_0 , there is the wavelet transform $Wf(u, s) = \langle f, \psi_{u,s} \rangle$ which depends on the value of f in the neighborhood of x_0 . We observe the cone of influence of x_0 in the *scale-plane* (u, s) .



Since $|u - x_0| / s \leq C$, the conditions of the theorem above is written thus:

$$|Wf(u, s)| \leq A's^{\alpha+(1/2)} \quad (5.272)$$

which corresponds to the *first theorem of a uniformly Lipschitz function f*.

The two preceding theorems show that the Lipschitz regularity of a function f at x_0 depends on the decay of the modulus $|Wf(u, s)|$ in the neighborhood of x_0 .

5.16.4 Detection of Singularities: The Maxima of the Modulus of Wavelet Transform are Associated with the Singularities

The theorem of Hwang and Mallat shows that there is a maximum at fine scales when a signal contains singularities. Indeed, it is shown that *there cannot be singularity without local maximum of the modulus of the wavelet transform at fine scales*. The theorem is thus interested in the modulus of the wavelet transform and obviously in the abscissa to which the singularity is identified on the signal. In general, we detect a *succession of modulus maxima converging towards the singularity*. The notion of *modulus maximum* $|Wf(u, s)|$ is described by means of the derivative at u of the transform:

$$\frac{\partial Wf(u, s)}{\partial u} = 0. \quad (5.273)$$

with s given, we search a local maximum at u .⁵⁵ (Remember that ψ is a wavelet, f is the signal-function and θ is a function.)

Theorem 5.8 (Hwang, Mallat). Suppose that ψ is of the class C^n with a compact support and that $\psi = (-1)^n \theta^{(n)}$ with $\int_{-\infty}^{+\infty} \theta(t) dt \neq 0$. Let $f \in L^1[a, b]$. If there exists $s_0 > 0$ such that the modulus $|Wf(u, s)|$ has no local maximum for $u \in [a, b]$ and $s < s_0$, then f is uniformly Lipschitz n on $[a + \varepsilon, b - \varepsilon]$ for any $\varepsilon > 0$.

Thus, the function-signal f can be singular at a point x_0 , which means not Lipschitz-1, if there exists a sequence of wavelet maxima points (u_α, s_α) that converges towards x_0 at the fine scales,⁵⁶ i.e. if $\alpha \in \mathbb{N}$:

$$\lim_{\alpha \rightarrow +\infty} (u_\alpha, s_\alpha) = (x_0, 0). \quad (5.274)$$

The decay rate of maxima on the curves indicates the order of the isolated singularities.⁵⁷ In general, we make appear the logarithm of the modulus of the wavelet transform $\log_2 |Wf(u, s)|$ in a graph in the plane $[\log_2(s), \log_2 |Wf(u, s)|]$. That means that it is possible to express the modulus maxima according to the scale in a log-log plane and the slope obtained provides the estimated order of singularity.

5.16.4.1 Examples of Local Maxima for the Continuous Wavelet Transform of the Stock Exchange Index: Cac40

The calculation of the local modulus maxima of the continuous wavelet transform for 2,048 daily values of the *Cac40 index*, then for its daily growth rate (Fig. 5.37).

The modulus maxima correspond to the “ridges curves” of images. Figure 5.38 represents in a log-log plane, the amplitude along the ridges of the continuous wavelet transform for a selection of six ridges.

5.16.5 Self-Similarities, Wavelets and Fractals

The most familiar way to approach the concept of *self-similarity* or more exactly the concept of *self-affinity*, that it is possible to observe for example *in the high frequencies of the financial series* or *in the Internet traffic*, can be presented from

⁵⁵ In detail, the wavelet transform is rewritten as a multiscale operator of order n ($\psi(t) = (-1)^n (d^n \theta(t)/dt^n)$; ψ with n vanishing moments) as presented previously: $Wf(u, s) = s^n \frac{d^n}{du^n} (f * \overline{\theta}_s)(u)$. The multiscale modulus maxima are used to analyze the discontinuities of a signal (see S. Mallat).

⁵⁶ Stephane Mallat underlines that they are modulus maxima of the transform. Whereas the instantaneous frequencies are detected while following the maxima of the normalized scalogram $(\xi/n)P_W f(u, \xi)$.

⁵⁷ $\log_2 |Wf(u, s)| \leq \log_2 A + (\alpha + 1) \log_2 s$.

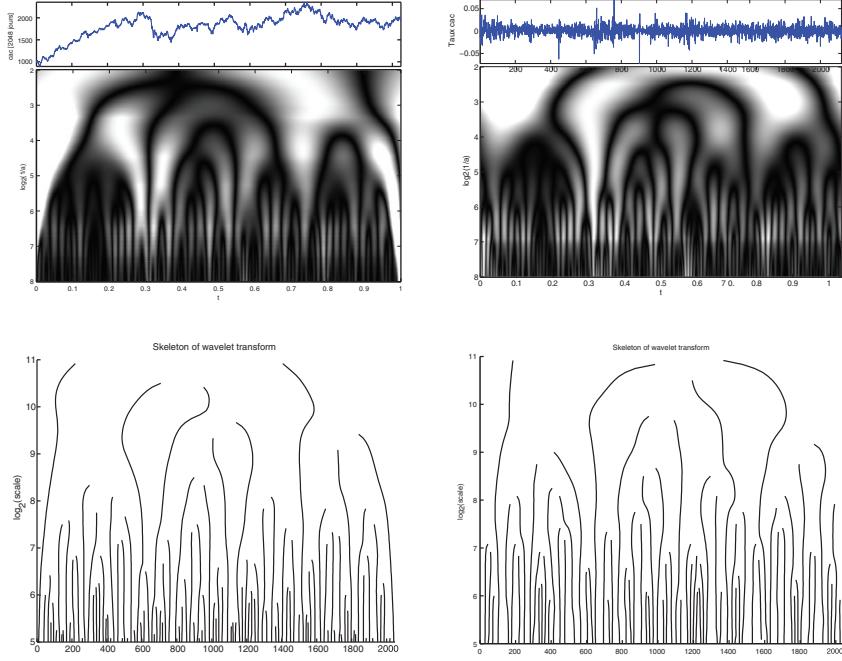


Fig. 5.37 Local maxima of the cwt

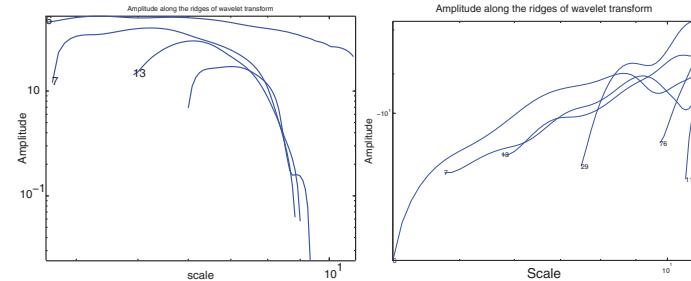


Fig. 5.38 Amplitude along the ridges

the *Hurst exponent* in the following way,⁵⁸ for a stochastic process $x(t), t \in \mathbb{R}^+$, $\exists H > 0$:

$$x(t) = a^{-H} x(at) \quad \text{for any } a > 0. \quad (5.275)$$

The *topological approach* of the *definition of self-similarities* states that a set $S \in \mathbb{R}^n$ is *self-similar* (or *self-affine*), if the union of the disjoint subsets S_1, \dots, S_m can be

⁵⁸ Where the equality is understood as a equality by distribution $\stackrel{d}{=}$.

obtained from S by scales of translation and rotation. The examples of sets of this type are numerous, such the *Cantor set* presented in another section. This concept must be associated with the concepts of the *fractal dimension*, *Hausdorff dimension or capacity dimension*, which is a simplification of the Haussdorff dimension. It is possible to describe the generic concept of capacity dimension in the following way. Let us suppose a set of points P, in a *space of dimension d*, then if we imagine for example a *first return map* which *intersects the trajectory* of an unspecified dynamics; consequently the dynamics is lying on this plane. We imagine then a cube (or a “hypercube” of unspecified dimension) denoted ε , and we measure the number $N(\varepsilon)$ of ε which is necessary to *cover the set of points P* (explaining thus its other name of *box counting*). Then, D the capacity dimension of the point set P is equal to:

$$D = \lim_{\varepsilon \rightarrow 0} [\log N(\varepsilon) / \log(1/\varepsilon)]. \quad (5.276)$$

We can illustrate the subject while choosing for the set P a *single point*, consequently the number of ε necessary to cover P, is $N(\varepsilon) = 1$, and the dimension D is equal to 0. In the case where P is a *segment* of which the length by convention is equal to 1, then $N(\varepsilon) = 1/\varepsilon$ and $D = 1$. And in the case where P is a simple *plane*, by convention of side equal to 1, then $N(\varepsilon) = 1/\varepsilon^2$ and $D = 2$, and so on.

If we uses again the topological approach of the self-similarity, and if we replace the cubes (or hypercubes) of the “box-counting” approach by “radial structures” (that we could presage above) by the action of the rotation scale. Then we can consider our set S (bounded on \mathbb{R}^n) and we can *count the minimum number* $N(s)$ of radial structures of radius s which it is necessary to cover the set S. If S is a set of dimension D of finite size (such that $D = 1, D = 2, D = 3$) then,

$$N(s) \sim s^{-D}, \quad (5.277)$$

consequently, we have: $D = -\lim_{s \rightarrow 0} (\log N(s) / \log(1/s))$. Thus, the *capacity dimension* D, or *fractal dimension*, of our set is defined in a generic way by:

$$D = -\liminf_{s \rightarrow 0} \frac{\log N(s)}{\log s}. \quad (5.278)$$

And the measure of S is written: $\limsup_{s \rightarrow 0} N(s) s^D$.

Self-similarity: Beyond the definition of selfsimilar processes $x(t) = a^{-H}x(at)$ given above, the wavelet analysis of signals and self-affine functions, offers a very interesting perspective to introduce the subject. Let us pose a continuous function f with a compact support S. This function is selfsimilar, if there exist disjoint subsets S_1, \dots, S_m such as the representation of f on each subset S_i is an affine transformation of f. Thus, we have for any $t \in S_i$ an affine transformation by a translation a_i , a weight b_i , a scale d_i , with a constant c_i :

$$f(t) = b_i f(d_i(t - a_i)) + c_i. \quad (5.279)$$

It is said that the affine invariance of f on S_i produces an affine invariance for any wavelet whose support belongs to S_i . Moreover, it is said that if a function is self-similar, or self-affine, its wavelet transform is also self-similar. Moreover, *the self-similarity of the wavelet transform means that the positions and the modulus maxima are also self-similar.*

If an affine transformation ℓ of f is selected, it is written:

$$\ell(t) = bf(d(t-a)) + c, \quad (5.280)$$

the wavelet transform of ℓ is written:

$$W\ell(u, s) = \int_{-\infty}^{+\infty} \ell(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt. \quad (5.281)$$

Since a wavelet had an integral equal to zero $\int_{-\infty}^{+\infty} \psi(t) dt = 0$, by a change of variable $t' = d(t-a)$, it comes:

$$W\ell(u, s) = \frac{b}{\sqrt{d}} Wf(d(u-a), ds). \quad (5.282)$$

If we take a wavelet with non-infinite support, i.e. compact, for example on $[-C, C]$, the affine invariance of f on $S_i = [m_i, n_i]$ provides an affine invariance for any wavelet whose support is included in S_i . For any $s < (n_i - m_i)/C$ and for any $u \in [m_i + Cs, n_i - Cs]$:

$$Wf(u, s) = b_i(d_i^{-1/2}) Wf(d_i(u-a_i), d_i s). \quad (5.283)$$

5.16.6 Spectrum of Singularity: Multifractals, Fractional Brownian Motions and Wavelets

Definition 5.5 (Singularity spectrum). Let S_α be a set of points $t \in \mathbb{R}$ where the Lipschitz regularity of f is equal to α . The spectrum of singularity $D(\alpha)$ of f is the fractal dimension of S_α .

The singularity spectrum measures the global distribution of singularities having different Lipschitz regularities. And the distribution of singularities in a multifractal signal is used for the study of its properties. It is in the *complex phenomena of turbulences*, in particular in fluid dynamics and in geophysics that this type of work was born. The spectrum of singularity makes it possible to obtain the *proportion of Lipschitz- α singularities* which are present at any scale s . Indeed, the capacity dimension or fractal: $D = -\lim_{s \rightarrow 0} \inf(\ln N(s)/\ln(s))$, is based on a principle which consists in proceeding to a disjoint cover of the support of f by “objects” (i.e. radial structures, intervals, or hypercubes) of length s , of which the number on the set S_α is:

$$N_\alpha(s) \sim s^{-D(\alpha)}. \quad (5.284)$$

A multifractal function or signal is known as “homogeneous”, if all the singularities have the same Lipschitz exponent α_0 . That means that the support of $D(\alpha)$ is reduced to α_0 . In order to illustrate the subject we will present the case of fractional Brownian motions which are homogeneous multifractals. It was shown previously that the multifractals have non-isolated singularities, and thus it is impossible to obtain their pointwise Lipschitz regularity. However, it is possible to analyze the singularity spectrum of multifractals by means of wavelet transforms of the local maxima. It is known that for a signal f , if the pointwise Lipschitz regularity $\alpha_0(< n)$ at x_0 (n is the vanishing moments of the wavelet), then the wavelet transform $Wf(u, s)$ has a sequence of modulus maxima $|Wf(u, s)|$ at the fine scale that converges towards x_0 .

The maxima at the scale s can be taken as a “cover of the singular support” of f by means of the wavelet transform at the scale s . At these maxima locations we have:

$$|Wf(u, s)| \sim s^{\alpha_0 + (1/2)}. \quad (5.285)$$

The locations of all these modulus local maxima of transforms at the scale s are represented by the sequence $u_\theta(s)$ (with $\theta \in \mathbb{Z}$: integers). By means of a partition⁵⁹ we can proceed to the sum of these modulus maximum at a power $q \in \mathbb{R}$, this sum is written: $\sum_\theta |Wf(u, s)|^q$. And the length between two consecutive maxima has a lower limit for a $\Delta > 0$, such that $|u_{\theta+1} - u_\theta| > (s\Delta)$. If the length between two maxima was shorter, i.e. equal or lower than the limit length $s\Delta$, the sum $\sum_\theta |Wf(u, s)|^q$ will include only the maxima of larger amplitude to “avoid the redundancies”. If we pose the exponent $\tau(q)$ which evaluates the decay of the sum, we have:

$$\tau(q) = \liminf_{s \rightarrow 0} \frac{\log \sum_\theta |Wf(u, s)|^q}{\log s}, \quad (5.286)$$

which provides:

$$\sum_\theta |Wf(u, s)|^q \sim s^{\tau(q)}. \quad (5.287)$$

Under several conditions concerning the wavelet and its vanishing moments, $\tau(q)$ is the Legendre transformation⁶⁰ of the singularity spectrum $D(\alpha)$.

Theorem 5.9 (Self-similar signal, Arneodo, Bacry, Jaffard, Muzy). Let $\Lambda = [\alpha_{\min}, \alpha_{\max}]$ be the support of $D(\alpha)$ and ψ is a wavelet with $n > \alpha_{\max}$ vanishing moments. If f is a selfsimilar function or signal, then we obtain (the Legendre transformation of the spectrum of singularity $D(\alpha)$):

$$\tau(q) = \min_{\alpha \in \Lambda} (q(\alpha + 1/2) - D(\alpha)). \quad (5.288)$$

⁵⁹ **Definition (Partition).** If $\bigcup_{i \in I} A_i = G$ and if all the A_i are supposed to be different from \emptyset and are pairwise disjoint, then the set $\{A_i\}$ is called a partition of G .

⁶⁰ **Legendre transformation:** A mathematical procedure in which one replaces a function of several variables with a new function which depends on partial derivatives of the original function with respect to some of the original independent variables. Also known as Legendre contact transformation.

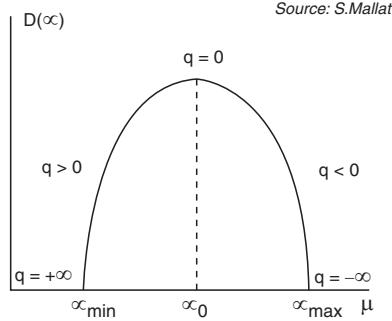


Fig. 5.39 Convex spectrum with $D(\alpha_0) = \max_{\alpha \in \Lambda} D(\alpha) = -\tau(0)$

It is noted that the scaling exponent $\tau(q)$ is the Legendre transformation $\tau(q) = \min_{\alpha \in \Lambda}(q(\alpha + 1/2) - D(\alpha))$ and is a convex function of q . This Legendre transformation is invertible (to recover the singularity spectrum $D(\alpha)$) if and only if $D(\alpha)$ is convex, thus:

$$D(\alpha) = \min_{q \in \mathbb{R}}(q(\alpha + 1/2) - \tau(q)). \quad (5.289)$$

Then it is possible to write that “the $D(\alpha)$ spectrum of a selfsimilar signal is convex”. Moreover, this observation can be applied to the majority of multifractals and fractional Brownian motions, under the condition of the convexity of the singularity spectrum. Refer to the works of A. Arneodo, E. Bacry, J.F. Muzy, S. Jaffard, Y. Meyer and S. Mallat about the singularity spectrum and its convexity. The Brownian motions are regarded as (Gaussian) selfsimilar processes and it is interesting to analyze their (wavelet and Fourier) singularity spectra (Fig. 5.39).

5.16.6.1 Power Spectrum and Wavelet Transform: Brownian Motion and Fractal Noise

The power spectrum is introduced to show that the differences or increments in a Brownian motion are stationary. But, let us note about the fractional Brownian motions that they:

- Are *non-stationary* and with *Gaussian increase* (i.e. *Gaussian increment*)
- Have *power spectra with fast decay*, (in spite of the difficulty in producing the spectrum because of non-stationarity)
- Are singular almost everywhere with the same Lipschitz regularity at all points

Definition 5.6 (Fractional Brownian motions). A fractional Brownian motion $x(t)$ of Hurst exponent $0 < H < 1$, is a process with zero-average Gaussian increase (increments), such that with $x(0) = 0$:

$$E[|x(t) - x(t - \tau)|^2] = \sigma^2 |\tau|^{2H}. \quad (5.290)$$

It follows that a fractional Brownian motion “fBm” is singular almost everywhere with a pointwise Lipschitz regularity $\alpha = H$. There exists an additional property, which explains that the singularity of fBm decreases when H increases, and conversely when H decreases, the singularity increases. Moreover if $\tau = t$ is posed, it follows:

$$E[|x(t)|^2] = \sigma^2 |t|^{2H}. \quad (5.291)$$

Furthermore, if $\tau = t - u$, is posed, we obtain then:

$$E[x(t) \cdot x(u)] = \frac{\sigma^2}{2} (|t|^{2H} + |u|^{2H} - |t-u|^{2H}). \quad (5.292)$$

Remember *the autocovariance definition* for a *random process* and a *stationary process*:

Definition 5.7 (Autocovariance function of a random process). An autocovariance function of a random process x_t with finite variance, is written: $\gamma_k = Cov[x(t), x(t+k)] = E([x(t) - E(x(t))][x(t+k) - E(x(t+k))])$.

Definition 5.8 (Autocovariance function of a stationary process). An autocovariance function of a stationary process x_t with finite variance, is written:

$$\begin{aligned} \gamma_0 &= Cov[x(t), x(t)] = E([x(t) - E(x(t))]^2) = Var(x(t)) = \sigma_x^2 \geq 0, \\ |\gamma_k| &\leq \gamma_0 \text{ and } \gamma_k = \gamma_{-k}. \end{aligned}$$

If we consider again the property $E[x(t) \cdot x(u)] = (\sigma^2/2) \cdot (|t|^{2H} + |u|^{2H} - |t-u|^{2H})$, it is said that the covariance of fBm does not depend solely on the argument $\tau = t - u$, which shows that a fBm is quite non-stationary.

Let us recall the definition of self-similar processes: $x(t) \stackrel{d}{=} a^{-H} x(at)$ or $x(at) \stackrel{d}{=} a^H x(t)$ (where $\stackrel{d}{=}$ means: Equal in distribution). It is by introducing a scale parameter on t which is noted a , that we verify the self-similarity characteristic. This way of making which consists in introducing a scale parameter on t , is familiar to the practitioners of the wavelet transform, although in general there are translations on t by u and modulations of the scale by $1/s$. In the case of the self-similarity property we note that the modulation of the scale over time is done with $1/s = a$. Beyond this remark, from the preceding expression: $E[x(t) \cdot x(u)] = (\sigma^2/2) \cdot (|t|^{2H} + |u|^{2H} - |t-u|^{2H})$, it follows:

$$E[x(at) \cdot x(au)] = E[a^H x(t) \cdot a^H x(u)], \quad (5.293)$$

and $x(at)$, $a^H x(t)$ are Gaussian with identical covariances and averages, and it is possible to write:

$$x(at) \stackrel{d}{=} a^H x(t). \quad (5.294)$$

As explained previously, one of the characteristics of the fBm is that they have power spectra with fast decay, in spite of the difficulty to produce the spectrum because of

non-stationariness. We circumvent the difficulty by *incrementing*,⁶¹ thus the power spectrum is calculated on the increments of fBm which are stationary.

Power Spectrum of a Fractional Brownian Motion and of a Fractal Noise

If we pose, on the one hand, a fractional Brownian motion $fBm(t)$ and on the other hand $\Delta_\tau(t) = \delta(t) - \delta(t - \tau)$, the process resulting from the increments:

$$I[fBm(t)] = fBm(t) * \Delta_\tau(t) = fBm(t) - fBm(t - \tau), \quad (5.295)$$

is a stationary process whose spectrum is written:

$$\widehat{S}_{I[fBm]}(\lambda) = \frac{\sigma_H^2}{|\lambda|^{2H+1}} |\widehat{\Delta_\tau}(\lambda)|^2. \quad (5.296)$$

where λ is the frequency, $\widehat{\Delta_\tau}(\lambda)$ is the Fourier transform of $\Delta_\tau(t)$. And $(fBm(t) * \Delta_\tau(t))$ is a continuous convolution. $\delta(t)$ a distribution of dirac.⁶² From the lines which precede, it arises that *the increments of a (non-stationary) fractional Brownian motion are stationary*. And we can extend the subject by writing:

$$\widehat{S}_{fBm}(\lambda) = \frac{\widehat{S}_{I[fBm]}(\lambda)}{|\widehat{\Delta_\tau}(\lambda)|^2} = \frac{\sigma_H^2}{|\lambda|^{2H+1}}. \quad (5.297)$$

The increments $I[fBm(t)]$ are stationary because in the expression of the power spectrum $\widehat{S}_{I[fBm]} = (\sigma_H^2 / |\lambda|^{2H+1}) |\widehat{\Delta_\tau}(\lambda)|^2$ the multiplication by $|\widehat{\Delta_\tau}(\lambda)|^2$ allows to “remove” the energy explosion of low frequency. And it is important to highlight that the non-stationarity of fBm occurs in the energy burst at low frequency. Note that $|\widehat{\Delta_\tau}(\lambda)|^2 = O(\lambda^2)$. (Recall: For two different discrete signals $f_1[n]$ and $f_2[n]$, to write $f_1[n] \sim f_2[n]$ is equivalent to write that $f_1[n] = O(f_2[n])$ and $f_2[n] = O(f_1[n])$. The function $O(\cdot)$ means “of order #”, indeed one can write that $f_1[n] = O(f_2[n])$, and there exists K such that $f_1[n] \leq K \cdot f_2[n]$).⁶³ *The power spectrum is used to prove that the increments of a fBm are stationary.*

⁶¹ In statistics we evaluate the degree of differentiation τ th to measure the quality of the stationarization.

⁶² *Dirac distribution:* These distributions are important in particular when one passes from the continuous to the discrete, and during the transition between functions and real discrete series. (They make it possible to relieve convergence problems.) A Dirac delta function $\delta(t)$ has a support reduced to $t = 0$, i.e. $\delta(0)$. And if one associates with a Dirac any function $f(t)$ its value in $t = 0$: $\int_{-\infty}^{+\infty} \delta(t) f(t) dt = f(0)$.

⁶³ In a similar way one defines the order $o(\cdot)$ of two discrete signals or function with $f_1[n] = o(f_2[n])$ by $\lim_{n \rightarrow \infty} (f_1[n]/f_2[n]) = 0$.

In a similar way to the Brownian motions, the power spectrum of a fractal noise $x(t)$ is decreasing and is written:

$$\widehat{S}_{x(t)}(\lambda) = \frac{\sigma_H^2}{|\lambda|^{2H+1}}. \quad (5.298)$$

This type of processes although generally non-Gaussian which can contain singularities of various types, have decreasing power spectra.

The Wavelet Transform of a Fractional Brownian Motion Is a Gaussian Stationary Process

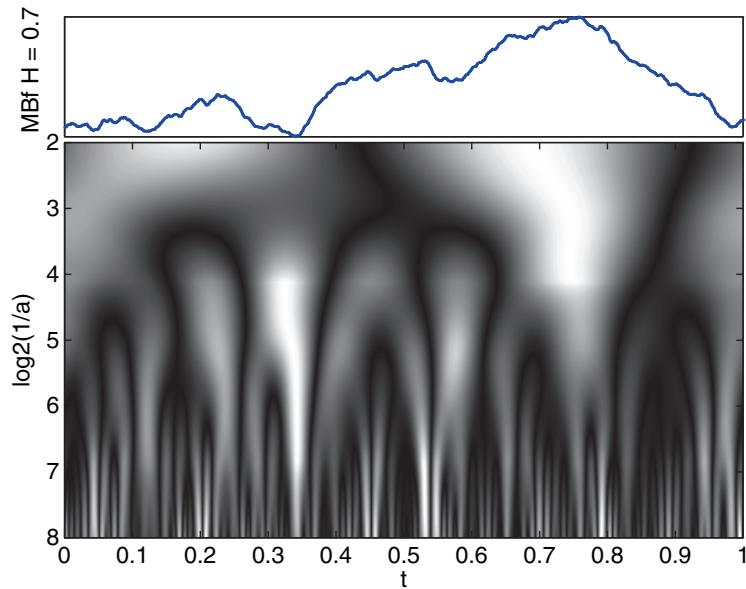
The transformation is written:

$$W_{\text{fBm}}(u, s) = fBm \star \overline{\psi}_s(u). \quad (5.299)$$

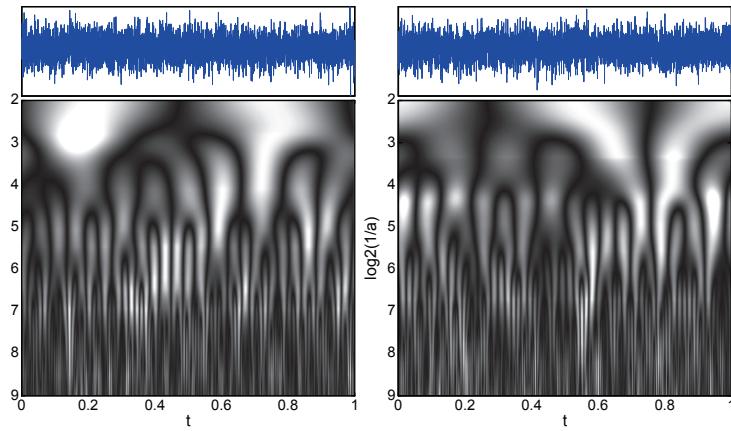
Around $\lambda = 0$, the modulus of the Fourier transform of a wavelet, which has at least one vanishing moment, is of order λ : $|\widehat{\psi}_s(\lambda)| = O(\lambda)$. *The purpose is to highlight the fact that the wavelet transform of a fBm on each scale is a Gaussian stationary process.*

$$\widehat{S}_{W_{\text{fBm}}}(\lambda) = s |\widehat{\psi}_s(s\lambda)|^2 \frac{\sigma_H^2}{|\lambda|^{2H+1}} = s^{2H+2} \widehat{S}_{W_{\text{fBm}}}(s\lambda). \quad (5.300)$$

The gaussianity and the self-similarity of the fBm show that the wavelet transform is selfsimilar on the scale: $W_{\text{fBm}}(u, s) = s^{H+1/2} W_{\text{fBm}}(u/s, 1)$. The figure hereafter illustrates the wavelet transform of a fractional Brownian motion for $H = 0.7$.



Hereafter two different white noises generated in an identical way and their wavelet transforms.



5.16.6.2 The Wavelet Transform of an Artificial Signal

The signal which is used as support here is successively made up of a Gauss curve, its derivative, then, its second derivative, a triangular, a sinusoid, a Dirac, a Morlet wavelet, a staircase, a random series and an increasing then decreasing oscillating structure ($n = 512$). Here is a continuous transforms by the Morlet wavelets and by the derivative[Gauss] wavelet in the time-scale plane (Figs. 5.40 and 5.41).

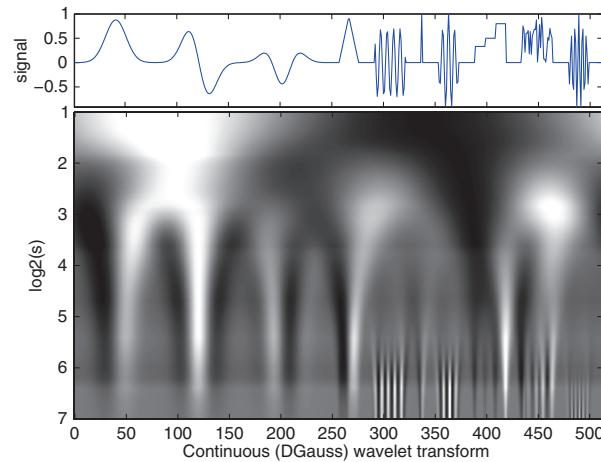


Fig. 5.40 Derivative[Gauss]-wavelet transform

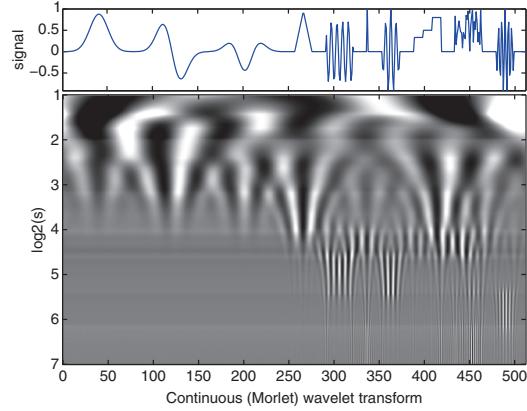


Fig. 5.41 Morlet-wavelet transform

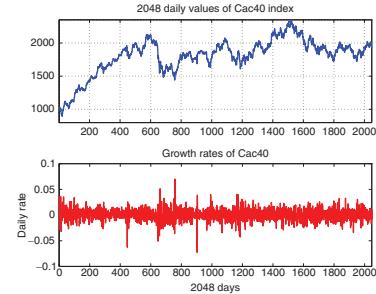


Fig. 5.42 2,048 daily values and their growth rates

5.17 The Continuous Wavelet Transform

5.17.1 Application to a Stock Exchange Index: Cac40

The decomposed series is a sample of the *Cac40* index corresponding to 2,048 daily values from January 1988 during more than 8 years (Fig. 5.42).

Let us recall that *the continuous wavelet transform* of a function (or a *signal*) f at the scale s and at the position u is calculated by correlating f with a *wavelet* (or with a *wavelet atom*):

$$Wf(u, s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt, \quad (\text{a})$$

where ψ^* is the *complex conjugate* of ψ in \mathbb{C} . There is *equivalence* between the expression above and the following equation which is written as a product of convolution:

$$Wf(u, s) = f * \bar{\psi}_s(u) \quad (\text{b})$$

with $\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^*(-\frac{t}{s})$. The Fourier transform of $\bar{\psi}_s(t)$ is:

$$\hat{\bar{\psi}}_s(\omega) = \sqrt{s} \hat{\psi}^*(s\omega) \quad (5.301)$$

with $\hat{\psi}(0) = \int_{-\infty}^{+\infty} \psi(t) dt = 0$, where $\hat{\psi}$ is the *transfer function*⁶⁴ of a *band-pass filter* of frequencies. Thus, *in practice the algorithms calculate the continuous wavelet transform by means of band-pass filters*.

The filtering is carried out in the space of the Fourier coefficients of the signal and in the space of the filter. The coefficients are successively filtered by the filter modulations. Then, we proceed to an inverse transform to return to the phase space of the signal at the chosen frequency scales, from the coarser to the finest. In this case, the length of the series which is the subject of the transformation is equal to 2,048, that means that it is a dyadic dimension ($n = 2^j$), i.e. $n = 2^j = 2^{11} = 2,048$. The finest scale of the decomposition of the signal is written: $\log_2(n) - 5 = \log_2(2,048) - 5 = \log_2(2,048) - \log_2(32) = 11 - 5 = 6$, and the coarsest scale is 2. We illustrate hereafter the method with a Gauss pseudo-wavelet at a fine scale, we observe in the lower left part of Fig. 5.43 the coefficients filtered in the Fourier space.

The result of the signal transformation at different scales is a matrix of vectors, it is represented in Fig. 5.44.

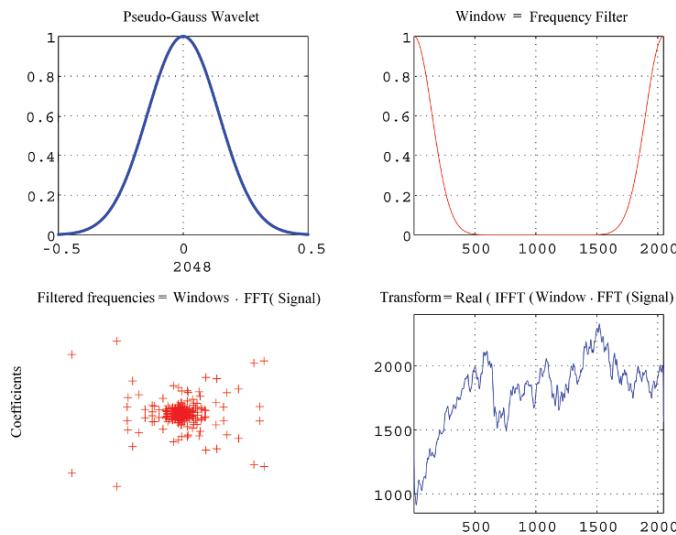


Fig. 5.43 Stages of the transformation by filtering

⁶⁴ Transfer function: The engineering terminology for a use of Fourier transforms. By breaking up a wave pulse into its frequency spectrum: $f_v = F(v)e^{2\pi i vt}$, the entire signal can be written as a sum of contributions from each frequency, $f(t) = \int_{-\infty}^{+\infty} f_v dv = \int_{-\infty}^{+\infty} F(v)e^{2\pi i vt} dv$.

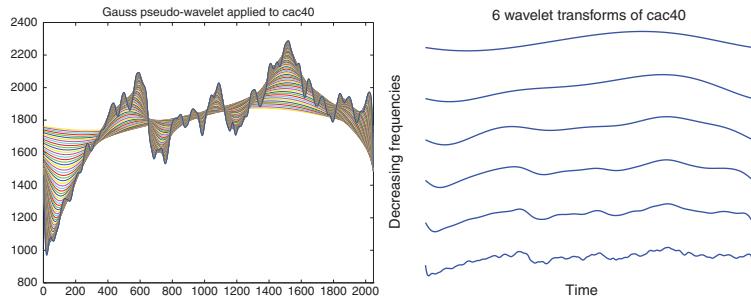


Fig. 5.44 Matrix of transforms (*left*). Range (scale) of transforms (*right*)

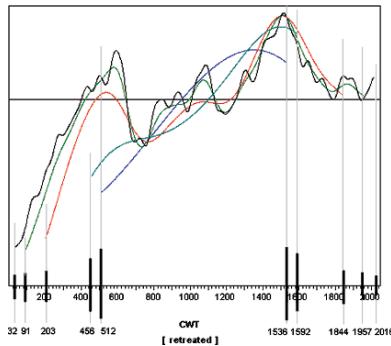


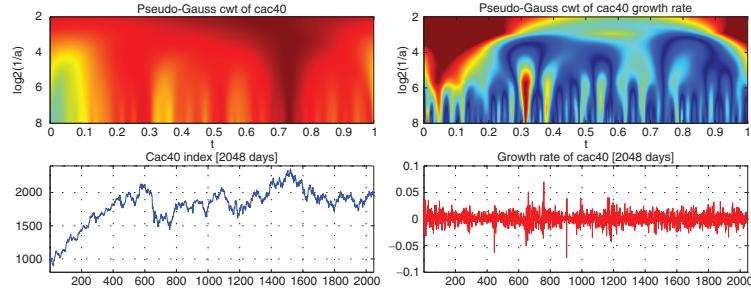
Fig. 5.45 Range of transforms used for the time-scale representation

The transform which considers at the same time the low frequencies and the high frequencies is the closest to the initial signal.

5.17.1.1 Images of Transforms in the Time-Scale Plane

The graphical representation of the continuous wavelet transform can be done in a time-scale plane with in abscissa u the time unit and in ordinate $\log_2(s)$ the frequency scale. The axis of ordinates exploits the form of wavelet transforms while going from the lower part for the transforms at the high frequencies towards the upper part for the transforms at the low frequencies. (The colors or the scale of gray represents the amplitude.) Thus, the time, the frequency and the amplitude are depicted by means of this graphical representation. The scale denoted $\log_2(1/a)$ corresponds to a division of the frequency, with a minimum and a maximum respectively equal to 2 and $\log_2(n) - 5$. Without providing the detail of the construction of the wavelet transform matrix resulting from the algorithm used to build this image in the time-scale plane, we show simply through Fig. 5.45 that the divergent elements of the matrix at the extremities have been removed.

The amplitude of the transform is expressed by a scale of colors or a gray scale (if the color is not used). Here are the results of a transformation by a Gauss-window (i.e. a Gauss pseudo-wavelet) of the *cac40* (picture on the left) and of its growth rate (picture on the right):



It is known that filtering by a Gauss-window (i.e. a Gauss pseudo-wavelet) does not stationarize the signal that contains a trend. Thus, the resulting image does not offer much information because the amplitudes are not centered and normalized. However, this Gauss filtering applied to the growth rate of the Cac40, offers more information. Indeed, the use of the growth rate (i.e. “differentiation”) stationarizes the signal. Thus, the changes of amplitudes and the frequencies become more visible. In the time-scale planes, the amplitude differences at the limits of involved frequencies are depicted by means of a gray scale (or colors). In the left image the green and blue colors represent the Cac40 depressions, which are visible in particular at the beginning of the series. In the right image, although the *analyzing function* is the same one, the transformation is applied to the growth rate and provides more information. We observe the dark color (or red) in the upper left part and in the upper right part of the image, they depict the depressions (described previously) at the beginning and also at the end of the signal, which are not visible any more besides by means of a direct reading on the growth rate series. We now clearly observe the high frequencies which are in the lower part of the plane (and which are depicted by a blue scale) which express at the same time different amplitudes. Let us remark for example, at the middle of the frequency scale and around $t = 0.3$ and $t = 0.75$, a dark color (blue) with ramifications in the lower part of the image in the highest frequencies. This is the representation of the two signal rises, visible on the initial signal, but that we do not distinguish any more on the growth rate; they involve many frequencies but not all frequencies.

Image of the Mexican Hat-wavelet transform (or Sombrero): Hereafter, we show the rough matrix of wavelet transforms and different three-dimensional perspectives (3D) of this same matrix, with the contours projected on the bottom of the reliefs of the image. Moreover, a two-dimensional image of the continuous Sombrero-wavelet transform (scale: “color jet”) is shown. The transformations by true wavelets center at zero the amplitudes and stationarize the signal, which is not the case with the Gauss pseudo-window (Figs. 5.46 and 5.47).

In the 2D image, the darkest colors (reds) express depressions of the signal. It is easy to identify on the left a depression which involves almost all the frequencies.

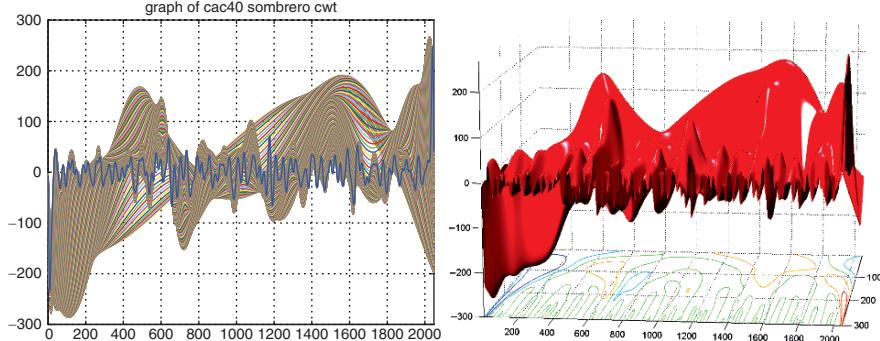


Fig. 5.46 Matrix of transforms (*left*). Perspective and contours (*right*)

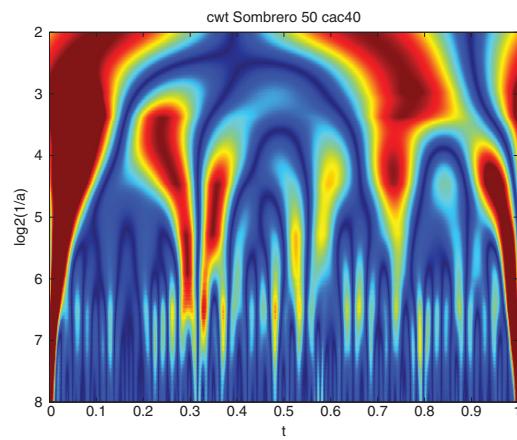
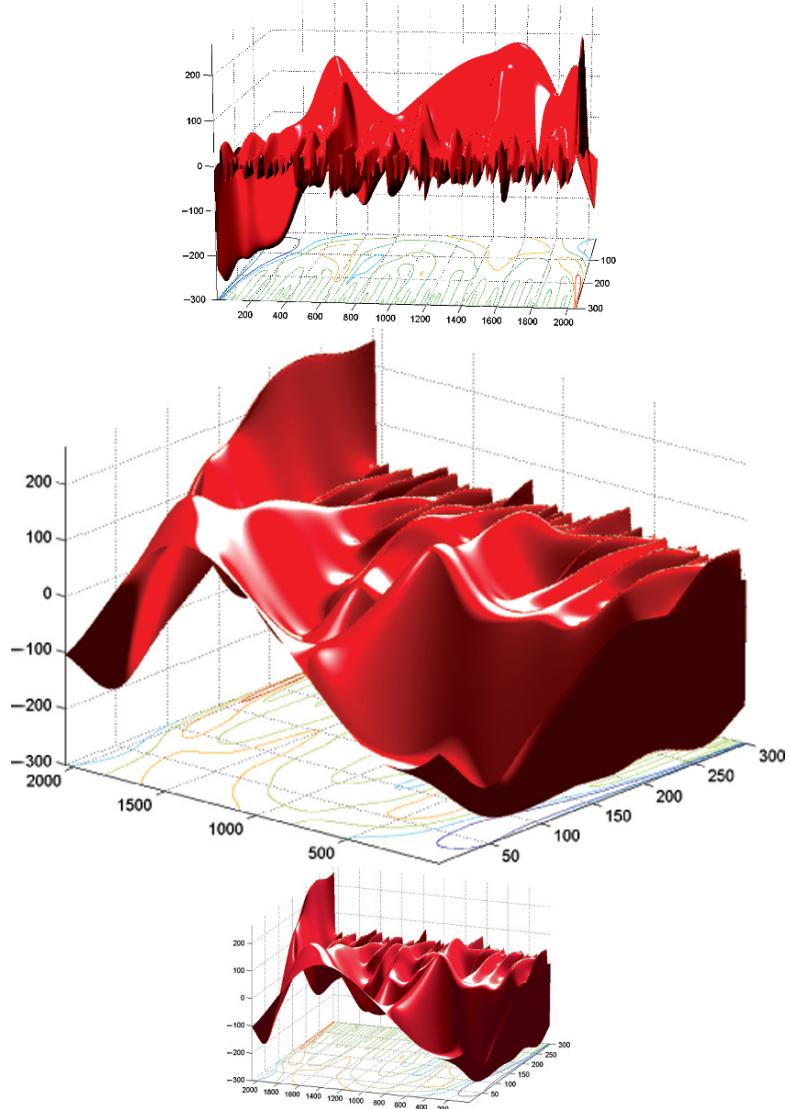
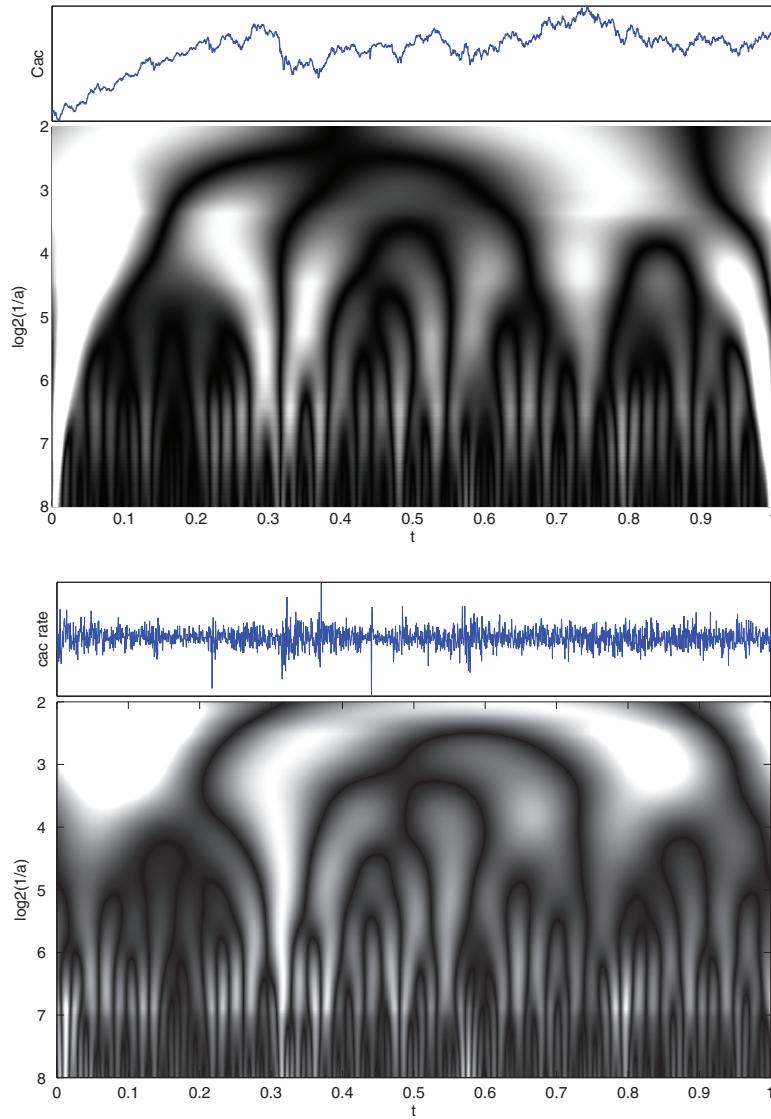


Fig. 5.47 Image of a CWT of the index Cac40 by sombrero wavelet (Jet version)

We identify smaller depressions that imply less frequencies. This is the case around $t = 0.75$ and around $t = 0.3$, for which the involved frequencies are different. We clearly observe at $t = 0.3$ that the covered frequencies are located on the scale between 3 and 7. Whereas for $t = 0.75$ the covered frequencies are located between 2 and 6. The high frequencies in the lower part of the image are numerous. Certain low frequencies in the upper part of the image are spread in the form of tree structures. It is the case for example of a (blue) dark tree structure at $t = 0.35$ which is divided into three very distinct parts, then is subdivided again until the highest frequencies in the lower part of the plane. That means that the observed “object”, if it is indeed made of all the frequencies, does not cover however all the time axis but only sub-segments of this one.

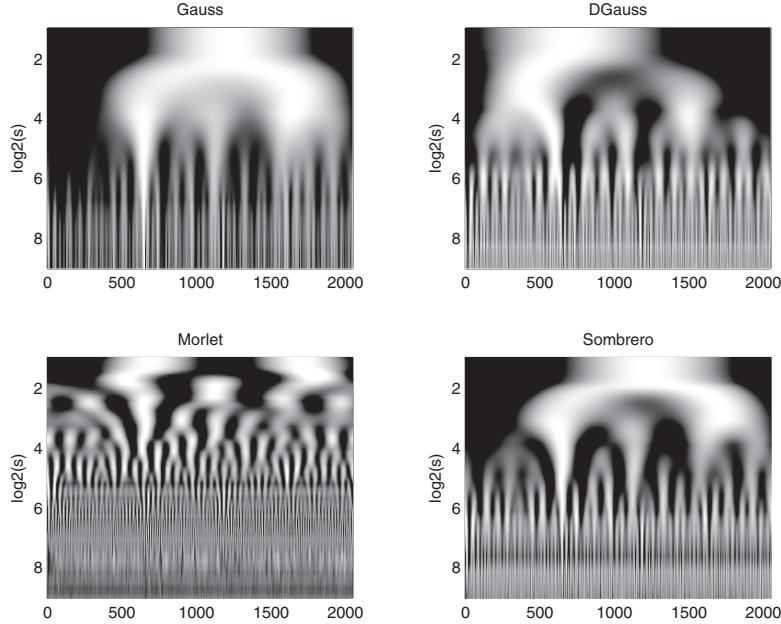


The option for the two images which follow is to use a gray scale to represent the amplitudes. The wavelet is a sombrero and the signal is the same French stock index as previously:



Compared images of transforms: The decomposed signal is the growth rate of the Cac40. The four types of decompositions were gathered in a same figure. We show a decomposition by means of a Gauss pseudo-wavelet, then by a true wavelet that

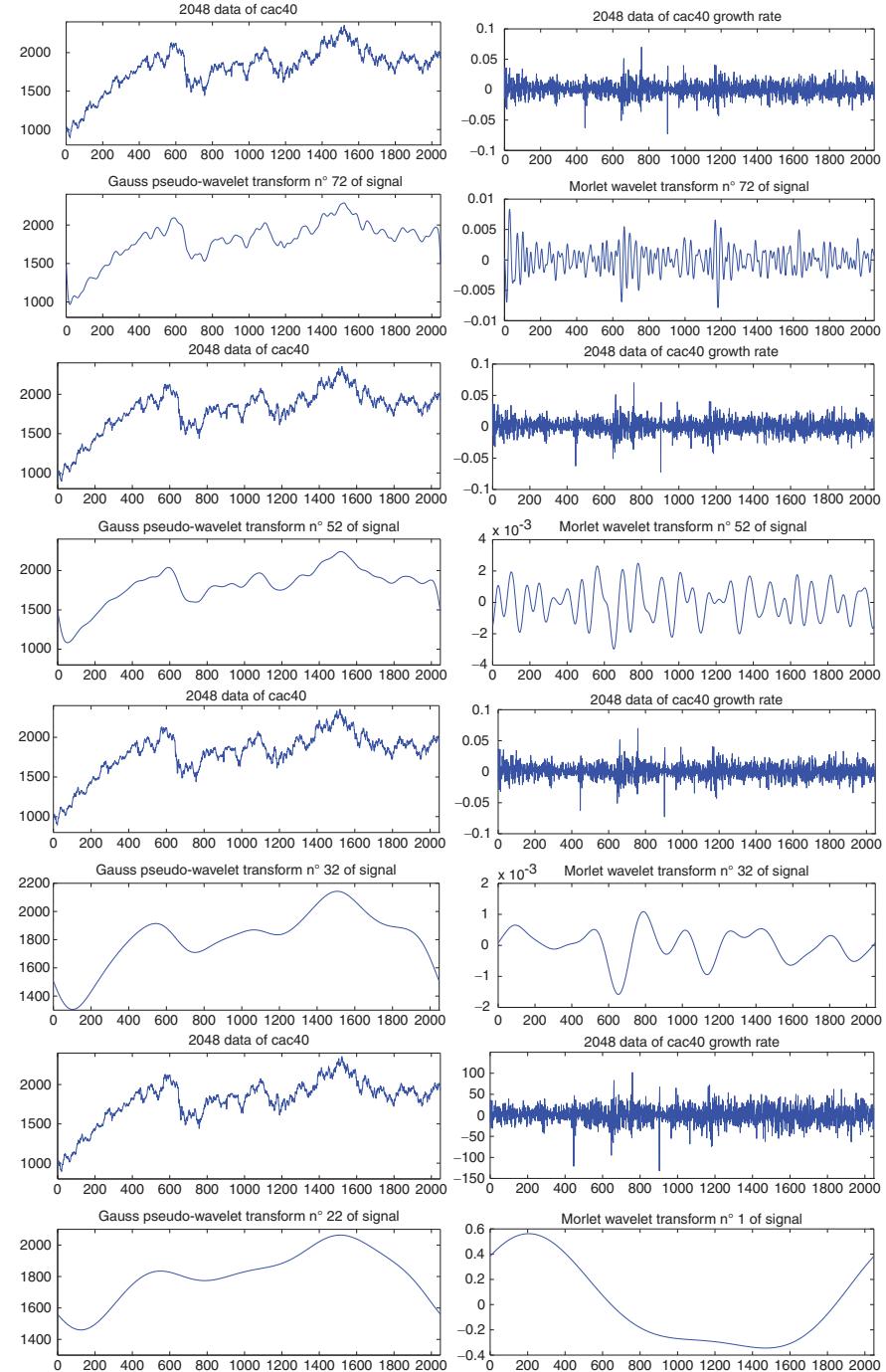
is the derivative of the Gauss pseudo-wavelet, then by a Morlet wavelet and finally by a sombrero wavelet. Note that the representations are rather different, but the fundamental subjacent structures can be still distinguished. The differences in the representation result from the shape of the analyzing wavelet.



5.17.1.2 Details of the Transform Matrix

We show the elements of the transform matrix by means of two types of wavelets: the Gauss pseudo-wavelet and the Morlet wavelet. Note that the results highlight different structures of the signal. This results from the nature of the *analyzing waveform* but also from the fact that we initially analyze the rough signal (i.e. non-stationary) and then, its growth rate (i.e. stationary).

Each group of graphs represents the stock index (Cac40) or its *growth rate* and *its transform* at arbitrary octaves; The simple objective is to illustrate the evolution of the shape of transforms at different scales, from the finest towards the coarsest. We will notice how much for the highest octave (i.e. for the finest transform) the adaptation to the shape of the signal is strong, as an “interpolation”, an “approximation”, or an “estimation”. Moreover, it is remarkable to note how much for the lowest octave, *the transform exhibits the shape of a wave whose cycle is spread over the entire length of the signal*.



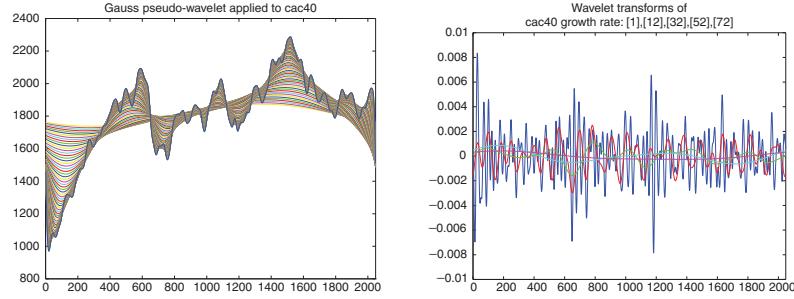


Fig. 5.48 Gauss pseudo-wavelet transforms (*left*). Morlet wavelet transforms (*right*)

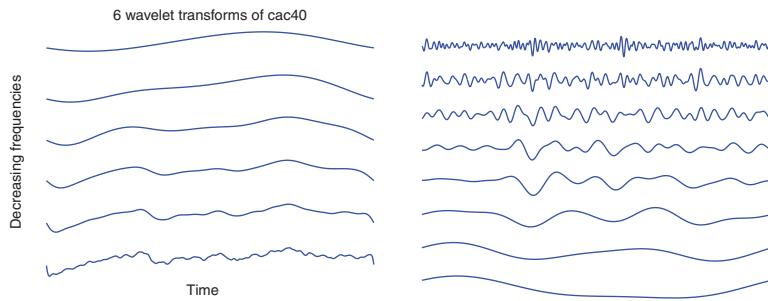


Fig. 5.49 Gauss pseudo-wavelet transforms (*left*). Morlet wavelet transforms (*right*)

Fig. 5.48 shows superimposed graphs of different transforms at different octaves.

Fig. 5.49 shows some transforms without superimposition. Note the decreasing scale of transforms in left-picture, and the increasing scale in the second.

5.18 Wigner–Ville Density: Representation of the Fourier and Wavelet Atoms in the Time–Frequency Plane

We will not describe the history of the Wigner–Ville density or distribution in spite of its great interest, which concerns physics and thermodynamics. As explained previously the concept of time–frequency atom allows to have a common notion to Fourier and wavelet analyses. The Wigner–Ville density is calculated by correlating the signal with a translation in time and frequency of itself. It is possible to do the same observation about time–frequency atoms. The Fourier transforms or wavelet transforms are also a correlation between the signal and its translation in time and frequency. Even if it is not the reason of its creation, since at the beginning the interest was the *representation of the instantaneous frequencies, nevertheless the Wigner–Ville density allows a type of representation in a time–frequency plane of time–frequency atoms, without resolution loss and without energy loss* (ref. to

Moyal’s theorem⁶⁵⁾. The limit of resolution corresponds to the limit of time–frequency atoms themselves. The Wigner–Ville density or distribution $\mathbf{D}_{W,V}(u, \xi)$ (which is also a projector P_V on V) is written as follows:

$$\mathbf{D}_{W,V}f(u, \xi) = P_V f(u, \xi) = \int_{-\infty}^{+\infty} f\left(u + \frac{\tau}{2}\right) f^*\left(u - \frac{\tau}{2}\right) e^{-i\pi\xi} d\tau. \quad (5.302)$$

We can also write it in the frequency field by means of the Parseval formula:

$$\mathbf{D}_{W,V}f(u, \xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \widehat{f}\left(\xi + \frac{\gamma}{2}\right) \widehat{f}^*\left(\xi - \frac{\gamma}{2}\right) e^{-i\gamma u} d\gamma. \quad (5.303)$$

Remark 5.1. While leaving the framework of this section temporarily, it seems interesting to provide a simple illustration of the difference between the Wigner distribution and the wavelet transform in the time–frequency planes. An elementary Dirac function makes it possible to observe in Fig. 5.50 that the Wigner–Ville distribution does not spread the localization of the Dirac function, whereas the wavelet transform simultaneously spreads the Dirac function in time and frequency in the time-scale plane.

The Wigner–Ville distribution is recognized to be an important instrument of the time–frequency analysis. However, its main *critique* is to produce *interferences* because of *quadratic terms* (or *cross terms*) in its construction. These terms can be highlighted, by creating for example a signal built by means of two “sub-signals”,

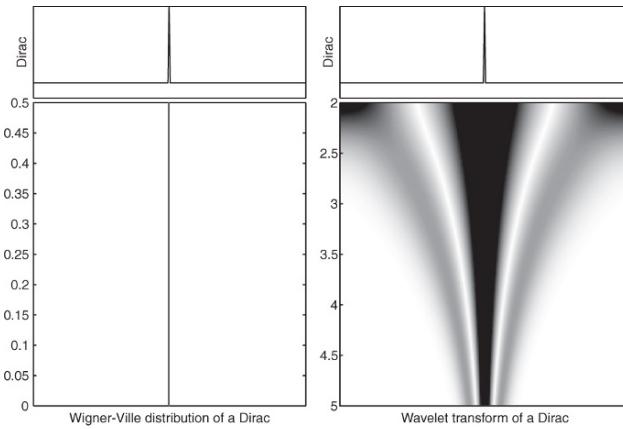


Fig. 5.50 Comparison (in time–frequency and time-scale planes) between the Wigner–Ville distribution and wavelet transformation of a Dirac function

⁶⁵ **Theorem (Moyal).** For any f and g in $L^2(\mathbb{R})$:

$$\left| \int_{-\infty}^{+\infty} f(t) g^*(t) dt \right|^2 = \frac{1}{2\pi} \int \int P_V f(u, \xi) P_V g(u, \xi) du d\xi.$$

For a demonstration see Moyal or Mallat.

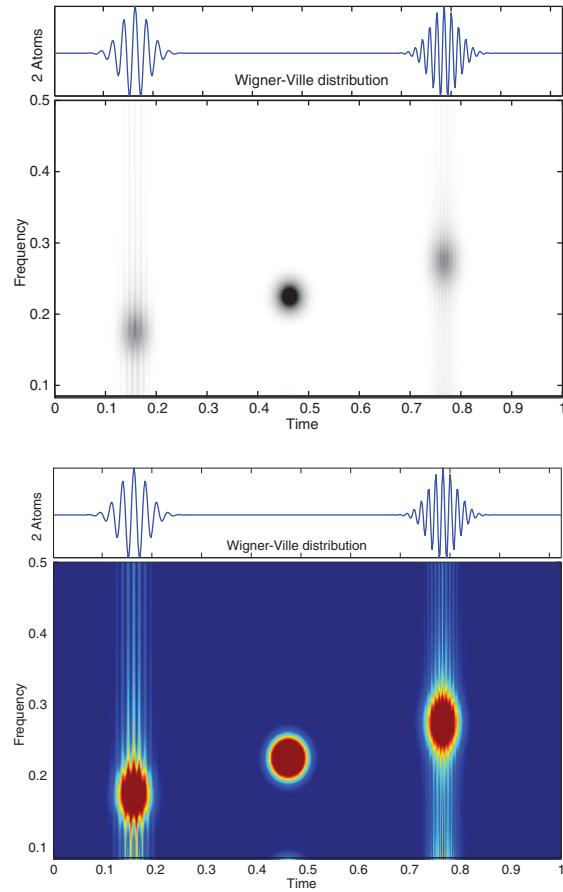
the quadratic terms of the combination of the two sub-signals *reveal non-zero values which correspond to interferences*: $f = f_a + f_b$:

$$P_V f(u, \xi) = P_V f_a(u, \xi) + P_V f_b(u, \xi) + P_V [f_a, f_b](u, \xi) + P_V [f_b; f_a](u, \xi) \quad (5.304)$$

the *interference terms* are:

$$I(u, \xi) = P_V [f_a, f_b](u, \xi) + P_V [f_b; f_a](u, \xi). \quad (5.305)$$

The figure which follows represents a signal composed of two Gabor atoms of different frequencies. We will observe the localization in time and frequency of the energy of each atom, but also the interference at the center (either by means of a gray scale, or a color scale according to the representation mode).



5.18.1 Cohen's Class Distributions and Kernels of Convolution

The *interference terms*, which are of *oscillatory nature* and with more or less complex structure, *can be “removed” or “attenuated” by the weighting action of kernels* θ (also called windows) in the Wigner–Ville distribution, which give in particular positive densities but nevertheless induced a resolution loss. Indeed, the interference terms contain positive and negative oscillations. They can be “removed” by means of the kernels θ . Then, it comes:

$$P_\theta f(u, \xi) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P_V f(\bar{u}, \bar{\xi}) \theta(u, \bar{u}, \xi, \bar{\xi}) d\bar{u} d\bar{\xi}. \quad (5.306)$$

The invariance⁶⁶ by linear translation is a fundamental property of the time–frequency analysis.⁶⁷ And the theory of the time–frequency analysis shows that “the invariant operators of linear translations are convolution products” (Mallat 1998, p. 116). The energy distributions are translated of a quantity equivalent to the translation, because the *energy conservation*⁶⁸ properties are respected (ref. to Moyal’s theorem). The kernel above can thus be written as a structure of translation in time and modulation in frequency, provided with the invariance properties:

$$\theta(u, \bar{u}, \xi, \bar{\xi}) = \theta(u - \bar{u}, \xi - \bar{\xi}), \quad (5.307)$$

which allow by convolution with the Wigner–Ville distribution to obtain a weighted or smoothed distribution, such as:

$$P_\theta f(u, \xi) = P_V f \star \theta(u, \xi) = \int \int P_V f(\bar{u}, \bar{\xi}) \theta(u - \bar{u}, \xi - \bar{\xi}) d\bar{u} d\bar{\xi}. \quad (5.308)$$

This type of distribution is known as of the “Cohen’s class”. The general expression above of the distribution is provided with a double integral on time and on frequency, and the kernels impact obviously the initial distribution on spreads corresponding to their own support in time and frequency. The types of kernels are rather numerous and condition obviously the result of the convolution. (In spite of the interest of the subject, we will not report it here.)

⁶⁶ Invariance by linear translation:

$g(t) = Lf(t) \Rightarrow g(t - \tau) = Lf_\tau(t).$

⁶⁷ Invariance property of the Wigner–Ville distribution:

$f(t) = g(t - u_0) \Rightarrow P_V f(u, \xi) = P_V g(u - u_0, \xi),$

$f(t) = \exp(i\xi_0 t)g(t) \Rightarrow P_V f(u, \xi) = P_V g(u, \xi - \xi_0).$

⁶⁸ Energy conservation equation: $\|f\|^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\widehat{f}(\lambda)|^2 d\lambda.$

Property of the Wigner–Ville distribution $P_V f(u, \xi)$: For any f , one has the following properties:

$\int_{-\infty}^{+\infty} P_V f(u, \xi) du = |\widehat{f}(\xi)|^2$ and $\frac{1}{2\pi} \int_{-\infty}^{+\infty} P_V f(u, \xi) d\xi = |f(u)|^2.$

These properties make it possible to regard the Wigner–Ville distribution as a “time–frequency energy density”.

When the distribution is computed again and the *interference terms removed* or more exactly *attenuated*, we can observe the image in the time–frequency plane of atoms (see Figs. 5.51 and 5.52).

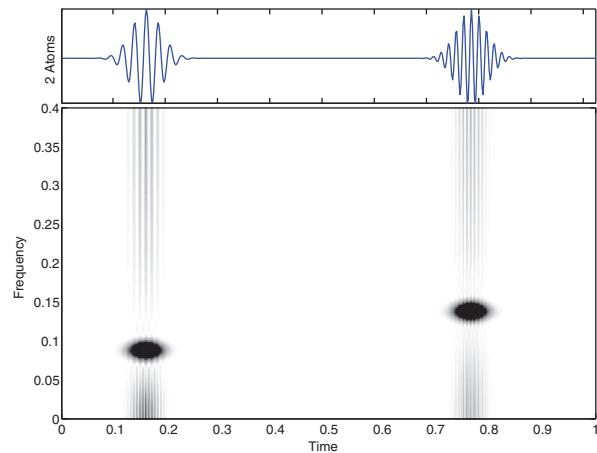


Fig. 5.51 Cohen’s class distribution of two Gabor atoms (B&W)

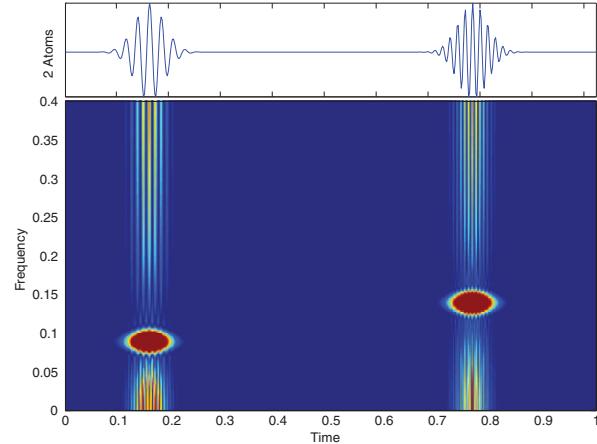


Fig. 5.52 Cohen’s class distribution of two Gabor atoms (Jet)

5.18.1.1 Distribution of an Artificial Signal

We create an artificial signal, it is successively composed of a Gauss-curve, then, its derivative, its second derivative, a triangular function, a sinusoid, a Dirac function, a Morlet wavelet, a staircase, a random series and an increasing oscillating structure then decreasing ($n = 512$). Figures 5.53 and 5.54 show the Wigner–Ville distribution and the Cohen’s class distribution in the time–frequency plane.

Hereafter, we present the Cohen’s class distribution for the same signal. Note that the representation in the time frequency plane is refined and the interferences are attenuated.

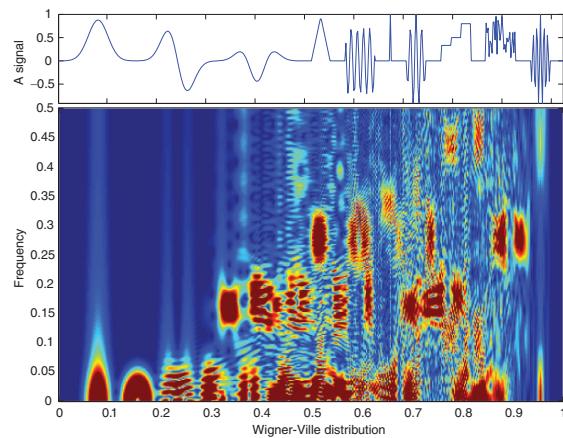


Fig. 5.53 Wigner–Ville distribution

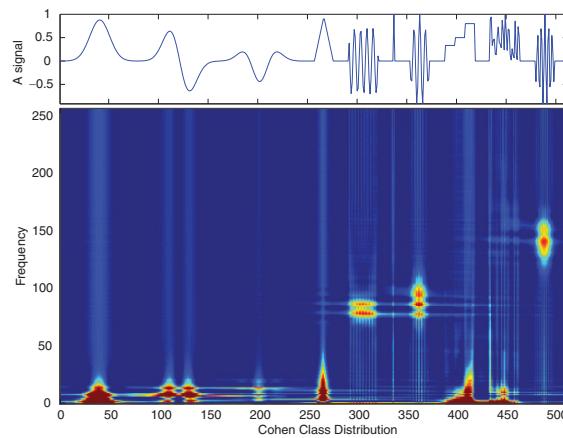


Fig. 5.54 Cohen’s class distribution

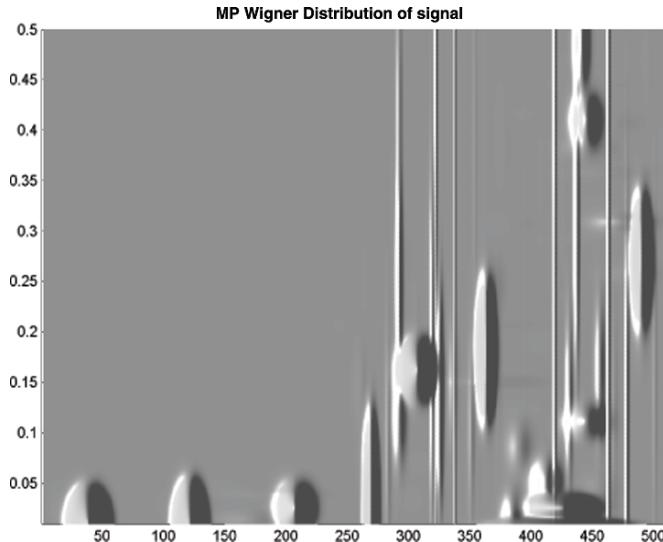


Fig. 5.55 Wigner distribution of the atomic decomposition by the MP

Wigner Distribution of the Decomposition by the “Matching Pursuit” of This Artificial Signal

As explained previously, the Wigner distribution can be used to represent in the time–frequency plane the decomposition by the “Matching Pursuit” with dictionaries of time–frequency atoms of this artificial signal. The signal is almost the same, as in the preceding section, except for the “random part” which is a new random sample. We will observe in Fig. 5.55 the result of the matching pursuit approximation in the time–frequency plane that it will be possible to compare with the previous Wigner–Ville distribution or with the Cohen’s class distribution for the artificial signal.

5.19 Introduction to the Polyspectral Analysis (for the Nonlinearities)

It is not possible to approach the domains of spectral analysis, time–frequency analysis or signal processing without providing a report about polyspectral analysis. New estimators of the normalized *polyspectrum* of the *order 3* (*bicoherency, bispectrum*) and of the normalized polyspectrum of the *order 4* (*trispectrum*) have been developed. Their *normalization terms take into account the fact that we search to identify the “phase relations” between the involved frequencies*. The evaluation of the performances of estimators are currently in process, in particular *about the nonlinear dynamical systems described by the Zakharov equations*. This type of

study is to be supplemented by the *evaluations of the fractal dimension of turbulent processes*. Such approaches are carried out by the physics laboratories for the signal processing of the stellar plasma type for example, but also by some economic laboratories about the social data analysis. We refer to some studies supported by the OECD concerning economical and social data which exploit the polyspectral analysis but also wavelet analysis (Amin et al. 1997).

5.19.1 Polyspectral Analysis Definition for Random Processes with Zero-Average

In order to define the polyspectral analysis or the spectral analysis of a high order (i.e. order higher than two), we use an extension of the autocorrelation notion with multiple lags (delays), from the definition of the *cumulant* (or moments) of order n of a random process with zero-average (Mendel 1991; Pilgram et al. 1997),

$$C_n(\tau_1, \tau_2, \dots, \tau_{n-1}), \quad (5.309)$$

the *cumulants of order 2* for the time series $x(t)$ is written:

$$C_2(\tau_1) = E[x(t)x(t + \tau_1)], \quad (5.310)$$

where $E[\cdot]$ is the expectation. Thus, the second order cumulant of $x(t)$ is the autocovariance function of $x(t)$. The third order cumulant is:

$$C_3(\tau_1, \tau_2) = E[x(t)x(t + \tau_1)x(t + \tau_2)] \quad (5.311)$$

and the fourth order cumulant is written:

$$\begin{aligned} C_4(\tau_1, \tau_2, \tau_3) &= E[x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3)] - C_2(\tau_1)C_2(\tau_2 - \tau_3) \\ &\quad - C_2(\tau_2)C_2(\tau_3 - \tau_1) - C_2(\tau_3)C_2(\tau_1 - \tau_2). \end{aligned} \quad (5.312)$$

If $x(t)$ is a *Gaussian random process*, all the cumulants of an order higher than 2 are equal to 0. This property is useful for example during a *cardiorespiratory analysis*, because it makes it possible to distinguish the *non-Gaussian components* (e.g. *a deterministic oscillation as a respiratory arrhythmia*) in a *Gaussian background noise*, independently of the noise spectrum shape, by the calculation of cumulants of order higher than 2. But, because the third order cumulants are equal to 0 for the Gaussian processes, but also for the processes which have a symmetrical distribution, then we use in such cases the fourth order cumulants. *The Fourier transform of cumulants are called spectra of high order, or polyspectra.* The Fourier transform of second order cumulants (corresponding to the autocovariance) is obviously the *power spectrum*. The Fourier transform of $C_3(\tau_1, \tau_2)$ is a *bispectrum*.

5.19.2 Polyspectra and Nonlinearities

The spectra of a high order are supposed to provide a tool to study the nonlinearities. *All the cumulants of order higher than two are equal to zero for a Gaussian process*, thus the polyspectra are particularly useful to study the deviations (i.e. gaps) in comparison with the Gaussian behaviors. But, this creates problems, because the intuitive interpretation of cumulants is not spontaneous. The spectral distribution of the second order is linked with the autocorrelation function by the Fourier transform,

$$S(\lambda) = \int_{-\infty}^{+\infty} C(\tau) e^{-i2\pi\lambda\tau} d\tau, \quad (5.313)$$

$$C(\tau) = \int_{-\infty}^{+\infty} S(\lambda) e^{+i2\pi\lambda\tau} d\lambda, \quad (5.314)$$

λ is the frequency and τ the lag. And we have in particular σ^2 the variance of the process written as follows:

$$\sigma^2 = C(0) = \int_{-\infty}^{+\infty} S(\lambda) d\lambda. \quad (5.315)$$

To calculate the bispectrum, we pose the “tri-variance” $T(\tau_1, \tau_2)$, written:

$$T(\tau_1, \tau_2) = E[(x(t) - \mu) \cdot (x(t + \tau_1) - \mu) \cdot (x(t + \tau_2) - \mu)], \quad (5.316)$$

to simplify the writings we poses μ as the average of the process $x(t)$, and we obtain the *bispectrum*:

$$B(\lambda_1, \lambda_2) = \int \int T(\tau_1, \tau_2) e^{-i2\pi(\lambda_1\tau_1 + \lambda_2\tau_2)} d\tau_1 d\tau_2. \quad (5.317)$$

Its inverse Fourier transform is written:

$$T(\tau_1, \tau_2) = \int \int B(\lambda_1, \lambda_2) e^{+i2\pi(\lambda_1\tau_1 + \lambda_2\tau_2)} d\lambda_1 d\lambda_2. \quad (5.318)$$

From the definition above, $T(0, 0)$ is σ^3 times the *skewness* (asymmetry⁶⁹), then the preceding equation becomes:

$$\sigma^3 \times \text{skewness} = \int \int B(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2, \quad (5.319)$$

thus, $B(\lambda_1, \lambda_2)$ is interpreted as a function which shows how the “skewness” is linked with the frequency pairs.

It is known that the *power spectra*, i.e. the correlations of order 2, are blind as regards the phases. In addition, it is said that the Gaussian random processes can

⁶⁹ In statistics, the *Skewness* test (i.e. asymmetry) is written: $\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^3 / \sigma^3$. And the *Kurtosis* test (i.e. flatness) is written: $\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^4 / \sigma^4$.

be completely specified by the knowledge of their statistics of the first and second order. There exist many practical situations, where we have to look at correlations of a higher order, i.e. cumulants of order higher than 2 of a signal, in order to extract the information concerning the phase, i.e. the presence of nonlinearities or the deviation in relation to the “Gaussianity” of the signal, etc. The polyspectra (order >2) of Gaussian processes are equal to zero. Thus, it is said in theory that the polyspectra analysis are domains with high Noise–Signal Ratio, where the identification of the system and the reconstruction of the signal can done and provide results. The non-Gaussian processes are better apprehended or identified by the polyspectra. However, the polyspectra have tardily received many critiques, in particular because the amount of data necessary to produce estimates of weak-variance and the number of involved calculations are very high.

Examples of Polyspectrum.

First example: Bispectrum and third order cumulants are calculated for an ARMA (2,1) process with $\text{AR} = [1, -1.5, 0.8]$ and $\text{MA} = [1, -2]$. Figures 5.56 and 5.57 depict respectively, Bispectrum, Cumulants in the plane, Cumulants in 3D and a sample of Trispectrum (contour plots).

Second example: The analyzed time series is segmented into several overlapping records. The signal corresponds to artificial data for the *Quadratic-Phase Coupling problem*. Signal is constructed from four unity amplitude harmonics with frequencies 0.1, 0.15, 0.25 and 0.40 Hz. We add to the signal a white Gaussian noise with a variance of 1.5. The harmonics at (0.1, 0.15, 0.25) are *frequency-coupled* and *phase-coupled*, however the harmonics at (0.15, 0.25, 0.40) are frequency-coupled, but not phase-coupled. Hereafter the bispectrum amplitude is shown in a contour plot. The sharpness of peaks highlights the quadratic phase coupling (see Fig. 5.58a).

Third example: Figure 5.58b shows the Bispectrum of sunspots (Annual sunspots for years 1700–1987). The sunspot time series is positive and the bispectrum is calculated by means of differences. *Last examples:* The first time series is the logistic model close to the chaotic regime. The second time series is a Bi-linear model ($\text{Signal} = y_1 \cdot y_2 : y_1 = \sin(2\pi \cdot f_1 \cdot t + \varphi_1)$ and $y_2 = (0.1) \cdot \sin(2\pi \cdot f_2 \cdot t + \varphi_2)$, $f_1 = 60 \text{ Hz}$, $f_2 = 4 \text{ Hz}$, and φ_1, φ_2 are randomly chosen. In this case $\varphi_1 = 0.7$ and $\varphi_2 = 0.3 \text{ radians}$). Each Fig. 5.59a and Fig. 5.59b depicts Averaged signal, Third order cumulant, Bispectrum magnitude and Bispectrum phase.

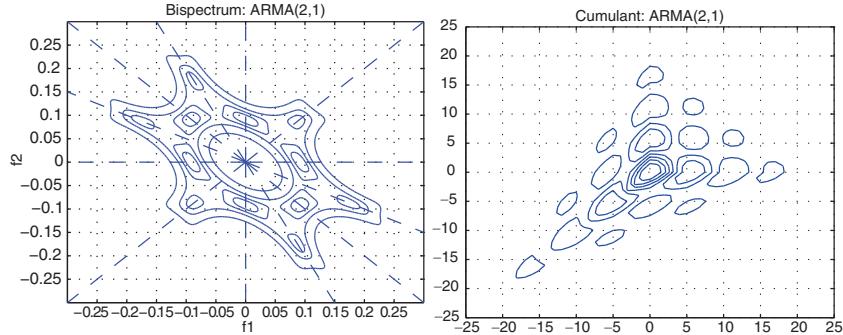


Fig. 5.56 ARMA(2,1) Bispectrum (left), ARMA(2,1) Cumulants (right)

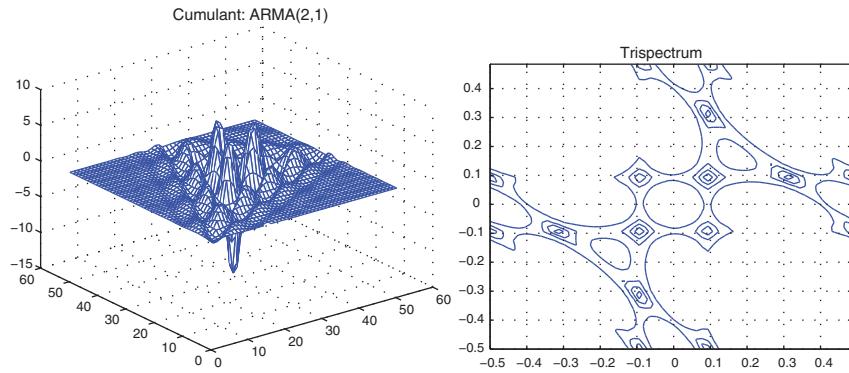


Fig. 5.57 3D ARMA(2,1) Cumulants (left), ARMA(2,1) Trispectrum (right)

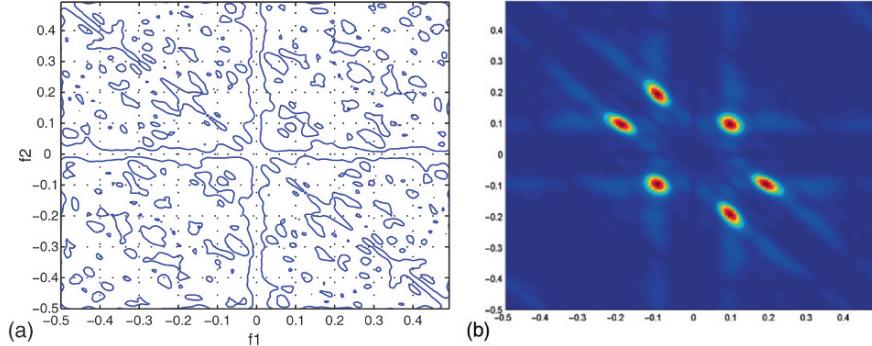


Fig. 5.58 (a) Cross-Bispectrum of signal, (b) Bispectrum of Sunspots

5.20 Polyspectral and Wavelet Bicoherences

5.20.1 A New Tool of Turbulence Analysis: Wavelet Bicoherence

Recently, a new tool for the analysis of turbulence has been introduced and investigated, it is the *wavelet bicoherence* (see van Milligen et al. 1995a). This tool

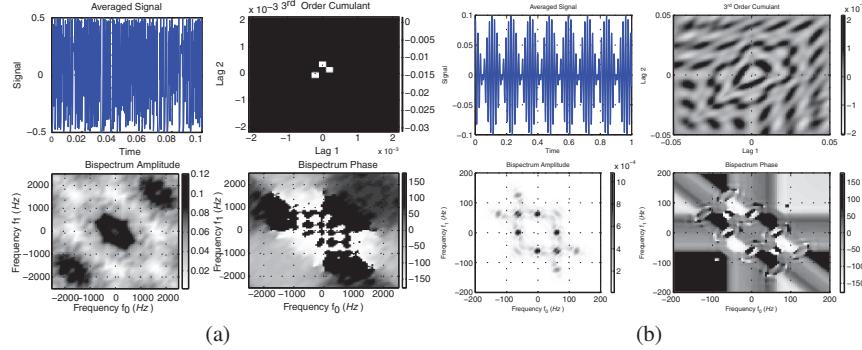


Fig. 5.59 (a) Logistic equation, (b) Bi-linear process

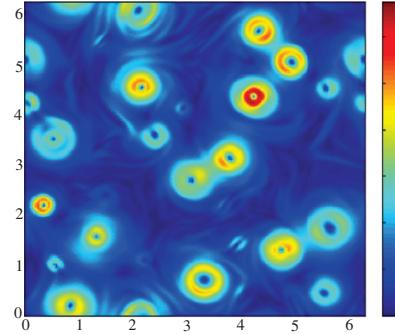
is able to detect the *phase coupling*, i.e. the *nonlinear interactions of the lowest “quadratic” order with time resolution*. Its potential is important and it was applied in specific works (see footnote⁷⁰) to numerical models of chaos and turbulence and also applied to *real measures*. In the first case, the van der Pol model of chaos, this tool detected the *coupling interaction between two coupled van der Pol oscillators*. In the second case, the *drift wave turbulence model* concerning the plasma physics, it detected a *highly localized “coherent structure”*. In the case of real measures, i.e. for the analysis of reflectometry measures concerning fusion plasmas, it detected *temporal intermittency* and a strong increase in nonlinear phase coupling which coincide with the Low-to-High⁷¹ confinement mode transition. Three arguments plead for a new tool about the turbulent phenomenon analysis: (1) *First*, the tools of *chaos theory* (as fractal dimensions, Lyapunov exponent, etc.) are not always easily applicable to real phenomena, in particular if the noise level is high. Moreover, information recovered by these methods is not always adapted to physical understanding. Indeed, a low fractal dimension measured in a real phenomenon is helpful information, but a high dimension, in particular higher than 5 (for example often observed in fusion plasmas) does not offer interesting solution. (2) *Second*, the applications of the traditional analysis (i.e. which involve long-time averages of moments of data) of *standard spectral analysis* are limited concerning chaotic or turbulent real phenomena. Indeed, if we consider the transition from quasi-periodicity to chaos in theoretical dynamical systems, as we have explained in the first part of this book, there exist several possible routes to chaos which can be summarized as follows: *period-doubling, crises and intermittency*. In these three main routes (Ott 1993), the transition to chaos are abrupt. Then, an explanation of the chaotic regime by means of a *superposition of a large number of harmonic modes* (i.e. oscillators which correspond to the Fourier analysis⁷²) does not seem suitable. Indeed,

⁷⁰ van Milligen et al. 1995b. Asociación EURATOM-CIEMAT, Madrid, Spain. B. Carreras. Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A. L. García. Universidad Carlos III, Madrid, Spain.

⁷¹ Usually written by physicists: (L/H).

⁷² See the Navier–Stokes equations sections in the present book.

Fig. 5.60 “Coherent structures” in turbulent phenomenon (ref: Haller)



for some systems of which the equations do not have a complete set of solutions and for which therefore the observed behavior cannot be decomposed into eigenmodes. Therefore, *a decomposition into Fourier harmonics (i.e. modes), which is appropriate for the global description of turbulence or chaos by means of (for example) the decay of the spectrum in frequency, produces a confused picture at a finer scale, because it is not an optimal method in particular for the expansion* (van Milligen et al. 1995b, p. 2). (3) *Third*, numerous and important arguments (both numerical simulations and observation turbulence) lead to state that *turbulence usually is an intermittent phenomenon*, that means that *it is localized in time and often in space also. Such an observation is in opposition to the Fourier analysis*, indeed, *the Fourier analysis supposes the “homogeneity” in the phenomenon, and thus does not seem suitable. Moreover, this concept of non-homogeneity in the turbulence phenomena is closely related to the concept of Coherent structures*⁷³ (see for example, Lagrangian Coherent Structures in 2D Turbulence, Haller’s research, Department of Mechanical Engineering, MIT).

Figure 5.60 depicts how *Coherent structures (and their possible interactions) could be distributed in a turbulence phenomenon picture*. These types of phenomena plead for the use of the wavelet analysis. As explained previously, the wavelet analysis can be taken as an extension of the Fourier analysis. Even if Fourier and wavelet analyses in many cases lead to similar conclusion, however *wavelet analysis has a fundamental additional property* which is the *time resolution*. (Even if we use the windowed Fourier transform or short time Fourier transform, because the defects and deficiencies of the Fourier methods persist.) It is important to highlight that many successful applications (Farge 1992; Hudgins et al. 1993; Meneveau 1991) of the wavelet analysis concern turbulent phenomena. At this stage, during the wavelet transformation the important purpose is to “extract the information relevant to non-linear interactions”. In connection with what precedes, the higher-order spectra in the Fourier analysis were applied with interesting results (Meneveau 1991; Kim and Powers 1978). The paper (van Milligen et al. 1995b) that we evoke presently goes further, indeed, the concept of *the first higher-order spectrum*, i.e. the *bispectrum*

⁷³ “While the existence of these structures is clear from visual observations, their mathematical description is far more difficult” Haller and Iacono (2003).

was generalized to the wavelet analysis. The bicoherence, i.e. the normalized bispectrum, is a measure of the amount of “phase coupling” which occurs in a signal. The phase coupling occurs when two frequencies θ_1 and θ_2 are simultaneously present in a signal along with their sum (or difference) and the sum of the phases φ of these frequency components remains constant. The bicoherence measures this quantity and is a function of two frequencies θ_1 and θ_2 which is close to 1 when the signal contains three frequencies θ_1 , θ_2 and θ that satisfy the relation $\theta_1 + \theta_2 = \theta$ and $\varphi_1 + \varphi_2 = \varphi + \text{constant}$ (otherwise it is close to 0). When the signal, which is analyzed, shows “structures”, it is highly probable that “some phase coupling occurs”. Therefore, the generalization of the bispectrum to wavelet analysis can be used to be able to “detect temporal variations in phase coupling”, i.e. “intermittent behavior” or “structures localized in time”.

In the paper evoked presently, this method is applied to the output of a nonlinear chaotic dynamical system (i.e. van der Pol), then to a model of drift wave turbulence relevant to plasma physics, and finally to real measurements, i.e. to the analysis of reflectometry measures made in (thermonuclear) fusion plasmas. The only one application which will be briefly presented in this section is the nonlinear van der Pol model, in the case of a system of two coupled van der Pol oscillators in the periodic and chaotic states.

5.20.1.1 Wavelet Bicoherence: Study of “Abrupt Frequency Change”, “Phase Coupling” and Singularities

The objective of these works was to analyze the abrupt changes of frequencies, pulses, or singularities very localized in time (as in turbulent phenomena). *These types of problems cannot be analyzed and resolved by the Fourier analysis which does not have resolution in time*, because this type of problem involves an integral over time as well as on frequencies. With the Fourier analysis, abrupt variations and singularities localized in time are spread among all the signal decomposition, and the temporal information is eliminated during the reconstruction of a signal (such remark is valid also for the windowed Fourier transform to rebuild the singularities or abrupt variations). Wavelet analysis is based on (non-continuous) “*oscillating functions*” (*with compact supports*) which decay rapidly in the course of time, rather than sines and cosines in the Fourier analysis which do not have such a decay. Indeed, a wavelet transform is a function of the frequency and time. Let us present briefly these important works. If ψ indicates a wavelet, consider a *wavelet family* which can be written:

$$\psi_a(t) = \frac{1}{a^p} \psi\left(\frac{t}{a}\right), \quad (5.320)$$

where a is a scale parameter and the factor p is a normalization choice, the authors selected $p = 0.5$ (the argument of their choice is to say that the L^2 -norm of the wavelet is independent of a). We know that the wavelet admissibility condition is written:

$$C_\psi = \int_{-\infty}^{+\infty} |\hat{\psi}(\omega)|^2 |\omega|^{-1} d\omega < \infty. \quad (5.321)$$

The wavelet transform of a function $f(t)$ is written:

$$W_f(a, \tau) = \int f(t) \psi_a(t - \tau) dt, \quad (5.322)$$

where τ is a translation parameter. And the inverse wavelet transform is written for $p = 0.5$:

$$f(t) = \frac{1}{C_\psi} \int \int W_f(a, \tau) \psi_a^*(\tau - t) d\tau \frac{da}{a^2}. \quad (5.323)$$

In the *wavelet analysis* we know that *time-resolved spectra* (*i.e. time resolution in spectra*) can be preserved. The *wavelet cross spectrum* is written:

$$C_{fg}^W = \int_T W_f^*(a, \tau) W_g(a, \tau) d\tau, \quad (5.324)$$

where T is a finite time interval and $f(t)$ and $g(t)$ are two time series. It is introduced a *delayed* wavelet cross spectrum:

$$C_{fg}^W(a, \Delta\tau) = \int_T W_f^*(a, \tau) W_g(a, \tau + \Delta\tau) d\tau \quad (5.325)$$

which is a *quantity* used to *detect “structures”* between two separated observations. A normalized delayed wavelet cross coherence is written:

$$\gamma_{fg}^W(a, \Delta\tau) = \frac{\left| \int_T W_f^*(a, \tau) W_g(a, \tau + \Delta\tau) d\tau \right|}{\left(P_f^W(a) P_g^W(a) \right)^{1/2}} \quad (5.326)$$

which is contained in the interval $[0, 1]$. The *wavelet power spectrum* can be written in term of the usual Fourier power spectra of the wavelet:

$$P_f^W(a) = \frac{1}{2\pi} \int P_{\psi_a}(\omega) P_f(\omega) d\omega \quad (5.327)$$

with $P_f(\omega)$ calculated on the interval T . The *wavelet*⁷⁴ *auto-power spectrum* can be written as follows: $P_f^W(a) = C_{ff}^W(a)$. The first higher-order spectrum along the same lines is introduced and the *wavelet cross-bispectrum* is defined by the authors of the paper in the following way:

$$B_{fg}^W(a_1, a_2) = \int_T W_f^*(a, \tau) W_g(a_1, \tau) W_g(a_2, \tau) d\tau \quad (5.328)$$

with $1/a = 1/a_1 + 1/a_2$ which is the frequency sum-rule. This *wavelet cross-bispectrum* measures the amount of *phase coupling* (inside T) which occurs between

⁷⁴ According to the term used in the authors' article.

wavelet components a_1 and a_1 of $g(t)$ and wavelet component a of $f(t)$ (in such a way that the sum-rule is verified). It is supposed that $\omega = 2\pi/a$, *wavelet cross-bispectrum* can be understood as the coupling between waveforms or more exactly between waves of frequencies, i.e. wavelets of frequencies such that $\omega = \omega_1 + \omega_2$.⁷⁵ In a similar way, the wavelet⁷⁶ auto-bispectrum is written:

$$B^W(a_1, a_2) = B_{ff}^W(a_1, a_2). \quad (5.329)$$

It is possible to write that the “squared wavelet cross bicoherence” is the “normalized squared cross-bispectrum”:

$$(b_{fg}^W(a_1, a_2))^2 = \frac{|B_{fg}^W(a_1, a_2)|^2}{\left(\int |W_g(a_1, \tau)W_g(a_2, \tau)|^2 d\tau \right) P_f^W(a)} \quad (5.330)$$

contained between the interval $[0, 1]$. Besides, the *squared wavelet auto bicoherence* is written:

$$(b^W(a_1, a_2))^2 = (b_{ff}^W(a_1, a_2))^2. \quad (5.331)$$

The square bicoherence $(b^W(a_1, a_2))^2$ is usually plotted in a (ω_1, ω_2) -plane rather than (a_1, a_2) -plane.⁷⁷ At this stage, the *total bicoherence* is written:

$$(b^W)^2 = \frac{1}{S} \sum \sum (b^W(a_1, a_2))^2, \quad (5.332)$$

where S the number of terms in the summation insures the inclusion in the interval $[0, 1]$. At this stage, the choice of one wavelet can be introduced. Usually the choice is often restricted to a wavelet which has Fourier transforms showing a single prominent peak. In this respect, a theoretical wavelet is constructed by the authors which is written as follows: $\psi_a(t) = \frac{1}{\sqrt{a}} \exp [i \frac{2\pi t}{a} - \frac{1}{2}(t/ad)^2]$ the choice of d provides the exponential decay of the wavelet, which is adapted to the time–frequency resolution.

5.20.1.2 Application to the Coupled van der Pol Oscillators

The coupled *van der Pol oscillators* can be written as follows:

$$\frac{dx_i}{dt} = y_i, \quad (5.333)$$

$$\frac{dy_i}{dt} = ((a_i - (x_i + b_j x_j)^2) y_i - (x_i + b_j x_j)), \quad (5.334)$$

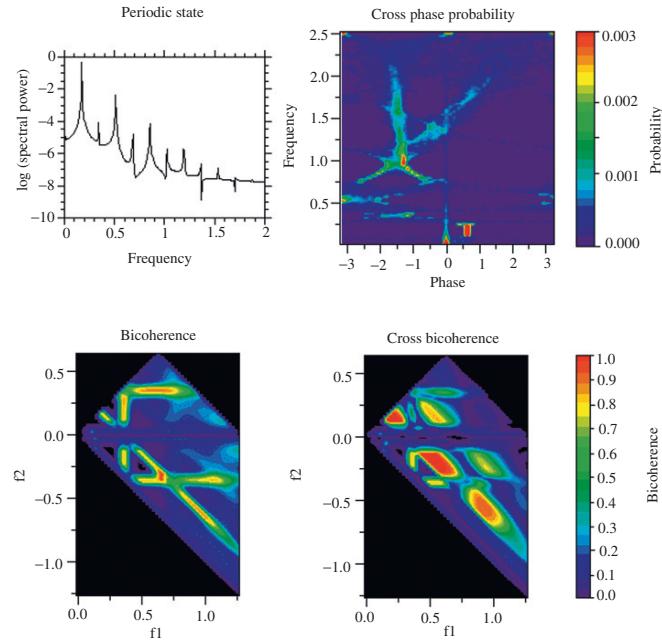
⁷⁵ See article about this postulate.

⁷⁶ According to the term used in the authors' article.

⁷⁷ Remark: $\omega_1 + \omega_2 = \omega$, this sum must be smaller than the Nyquist frequency (half sampling frequency). Moreover, it is postulated to plot that $\omega_1 \geq \omega_2$.

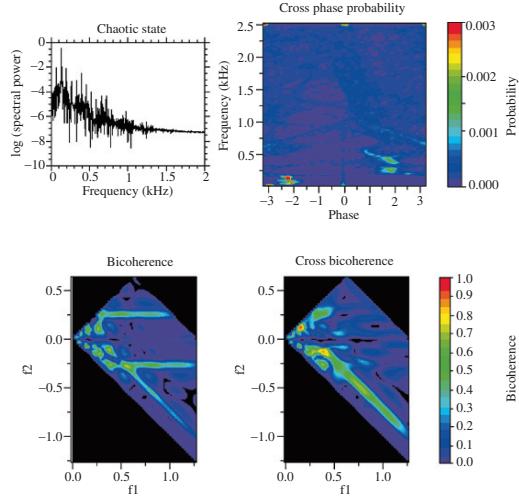
where $i = 1, j = 2$ correspond to the first oscillator and $i = 2, j = 1$ is the second. a_i correspond to the limit-cycles of oscillators and b_j correspond to the *nonlinear couplings* between oscillators. Two cases are studied, the first one is the *periodic state* and the second is the *chaotic state*.

Periodic state: The successive pictures below, show the *Fourier spectrum* (top-left), the cross phase probability (top-right), the *wavelet bicoherence* (bottom-left), the *wavelet cross bicoherence* (bottom-right). The Fourier spectrum shows some rare peaks of frequency and their associated harmonics. The bicoherence (bottom-left) shows *rectilinear horizontal and diagonal ridges* corresponding to a frequency at 0.34, which can be identified in the spectrum as the second peak. So, the two dominant peaks (i.e. the first and the third) in the Fourier spectrum respectively at 0.17 and at 0.5 couple (i.e. there is coupling) with their difference frequency at 0.34. Similarly, the difference in frequency between the second and the fourth peak, between the fourth and the sixth peak, ... is always 0.34. Idem for the odd sequence of peaks. *The even peaks are the consequence of the coupling interaction between the two oscillators, whereas the odd peaks are the harmonics of the limit cycle of oscillators* (ref. to van Milligen et al. 1995b).

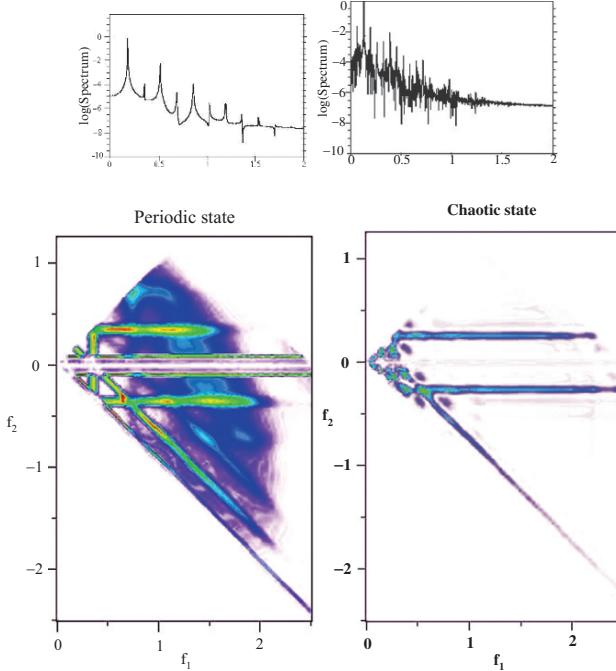


Chaotic state: Fourier spectrum (top-left) still shows some of former peaks and other of the periodic state which are shifted in frequency. *New peaks occur because of period-doublings* (on the route to chaos). Spectrum shows chaos on a broad band. The bicoherence (bottom-left) shows (weaker) *rectilinear horizontal and diagonal ridges* which appear at a frequency close to 0.25 corresponding to the 4th peak in the Fourier spectrum, which is related to the frequency at 0.34 highlighted in the

periodic-state spectrum. A modification of the parameter of the coupled model to obtain the chaotic regime induces a global frequency shift.



Extracting the *wavelet bicoherences* of the periodic and chaotic states from the preceding pictures, we *highlight* the comparison between two states:



The comparison above shows that the images have similar strong main ridges, and except for these ridges the images are different. Indeed, the vertical line which

appeared around the frequency “0.34” (showing the *coupling* in the periodic-state bicoherence) is divided and partitioned into several small distinct parts. This remark is in connection with the period-doublings appeared in the chaotic-state. New couplings can be also observed in lower frequencies (around 0.2 and 0.13, see article for comments).

5.20.2 Compared Bicoherences: Fourier and Wavelet

In order to compare the bicoherences between wavelet and Fourier, the analyzed measures are taken at the “Advanced Toroidal facility” (ATF, Lyon et al. 1986) with “Langmuir” probes and with (known) strong Fourier bicoherence (Hidalgo et al. 1994).at 1 MHz. The *Fourier spectrum* of this signal (not shown) is *turbulent* and does not have preeminent peaks or modes (i.e. oscillations, periods or cycles). *Figure 5.61* show the Fourier and wavelet bicoherences.⁷⁸ The wavelet picture shows less detail and the global aspects of the pictures are similar.

In order to demonstrate that the wavelet bicoherence shows more clearly elements which are less well detected by means of the Fourier spectrum, the FFT is computed on the raw data, the phase information is scrambled, and an inverse FFT is computed to obtain a new data series. We strongly recommend the reader to consult this interesting article to have comments, technical demonstrations and results.

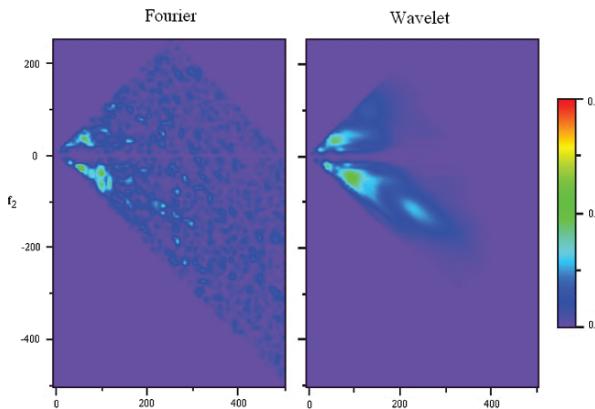


Fig. 5.61 Bicoherences (ref. to van Milligen et al, 1995a, 1995b)

⁷⁸ Computed on a frequency grid with 64 grid points from 0 to 500 kHz from a data section running from 1 to 16 ms.

5.21 Arguments in Favor of Wavelet Analysis Compared to Fourier Analysis

5.21.1 Signal Deformation by Diffusion of Peaks, Discontinuities and Errors in the Fourier Transform

In a Fourier transform the information about time is not lost because we can rebuild the signal from a transform, however it is hidden under the phases, i.e. the same sines and cosine can represent moments very different of the signal because they are shifted in phase to amplify or to cancel themselves. Consequently, the Fourier analysis is inadequate and unsuited to the signals which change abruptly and in an unpredictable way. However, such changes contain very interesting information. In theory, it is possible to extract information about time by computing the phases from the Fourier coefficients. In practice, to calculate them precisely is impossible. The fact that the *information at one moment of the signal is widespread among all the frequencies of the transform is a major problem*. A local characteristic of the signal as a peak or a discontinuity becomes a global characteristic of the transform. In connection with the peaks, we know for example that a narrow and high rectangular function has a very wide Fourier transform, as it is easy to observe in Fig. 5.62.

As mentioned before, a discontinuity for example, is represented by a superposition of all the possible frequencies. In the same way, as the white color is made of all the colors of the spectrum; It is not possible to deduce from such a superposition that the signal is discontinuous and moreover to locate this discontinuity. Consequently for all these reasons, in *the Fourier space the lack of information about time makes a Fourier transform very sensitive and vulnerable to the errors*. Indeed, *an error can corrupt all the Fourier transform. The (true or erroneous) information in a part of the signal spreads throughout the transform; thus, the errors of phases deform the signal and can produce a signal very different from the original*.

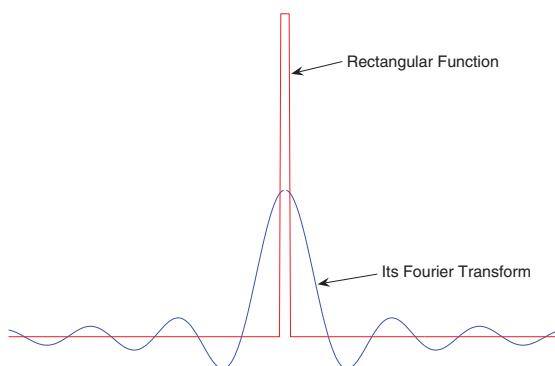


Fig. 5.62 The Fourier transform of a peak is extensive

5.21.2 Wavelets are Better Adapted to the Signal by Respecting Discontinuities and Peaks, Because they Identify the Variations

Because of the preceding observations, the wavelets have been conceived. *They avoid the diffusion and interference phenomena, allowing also the decomposition of the signal and its reconstruction. In the Fourier analysis, the window is fixed and the number of oscillations varies. A small window is blind for the low frequencies which are too large to be captured. In a too large window, the information about the fast or abrupt variations is drowned in the information concerning the entire interval contained in the window.* In the wavelet analysis, we analyze the signal at different scales, because we stretch and compress a mother wavelet. The broad wavelets provide an approximate image of the signal while the narrow wavelets allow to zoom in the details.⁷⁹ We reconstruct a signal from its wavelet transform, by adding wavelets of various sizes, just as for the Fourier transform we rebuild the signal by adding sines and cosines. In principle, the calculation of coefficients is done in the same way, we multiply the signal and the analyzing function, and we compute the integral of the product.⁸⁰ However, in practice we use different fast algorithms. (We have already explained that a wavelet⁸¹ was built starting from the “Gaussian function”, but this one does not satisfy the strict conditions for the construction of a wavelet, and such a wavelet was used by Gabor in the windowed Fourier analysis.) The stretching and compression of wavelets are the true inventions which made evolve the method. The wavelets are automatically adapted to the different components of the signal: they use a narrow window to look at the *transitory* components and a broad window for the *components of long duration and low frequency*.

It is essential to remember that, unlike the Fourier analysis, the information about frequencies is only approximate, indeed a sine and a cosine have a precise frequency, this is not the case for a wavelet. This is the characteristic of wavelets, due to their “constant shapes”: The resolution, the scale and the frequency vary at the same time.

The set of coefficients at all the scales provide a good image of the signal. Unlike Fourier series, here, the wavelet coefficients translate, in simple, precise and exact ways, the properties of these functions, i.e. the properties which correspond to transitory periods, for example the ruptures, discontinuities. It is said that the *wavelets detect and encode only the variations*. A wavelet coefficient measures the correlation between the wavelet (its peaks and its hollows) and the corresponding interval of the signal and they allow to closely look at the details of a signal. A constant interval of the signal provides a coefficient equal to zero, because the wavelet by definition has a zero-integral. Thus, the integral of the product of the wavelet by the signal is equal to zero for a constant signal. *The wavelet analysis is a manner of expressing the sensitivity to the variations.*

⁷⁹ A technique which is sometimes called the “mathematic microscope”.

⁸⁰ $\int(f * g)dt$.

⁸¹ Created by Yves Morlet.

When we use the continuous wavelet transform, any function can be named “wavelet”, on the condition that it has a zero integral. The Fourier transform decomposes a signal according to its frequencies; the wavelet transform decomposes a signal into its components at different scales. In both cases, we calculate the integrals: we multiply the signal by the analyzing function (sinusoids or wavelets) and we integrate. It is essential to explain that the transformation is robust, i.e. a small change in the wavelet representation induces a comparable change of size in the signal; *a small error or modification is not amplified in an exaggerated way.*

5.21.3 Wavelets are Adapted to Non-Stationary Signals

The wavelet analysis is thus appropriate to the non-stationary signals which present discontinuities and peaks. The Fourier analysis is appropriate to regular periodic signals and to stationary signals. If the signal is regular over a long duration and simply oscillates, it is not logical to analyze it with small wavelets, which capture only some oscillations. Moreover, the determination of the frequency by wavelets is imprecise in high frequencies.

5.21.4 Signal Energy is Constant in the Wavelet Transform

Grossmann and Morlet proved that when we represent a signal by means of the wavelet transform, the signal “energy” does not change. This *energy* corresponds to *the average value of the square of the amplitude*, which is different from the energy concept in physics.

5.21.5 Wavelets Facilitate the Signal “Denoizing”

The wavelets provide also a method to extract the signal from the white noise, which exists at all the frequencies. The procedure is simple, we transform the signal by means of wavelets, we eliminate for all the resolutions the coefficients lower than a “threshold value” and we rebuild the signal with the remaining values. This method requires few information about the signal. Previously we were supposed to guess the type of regularity of the signal.

5.21.6 Wavelets are Less Selective in Frequency than the Fourier Transform

In the Fourier analysis, the analyzing function is a sinusoid of precise frequency and when we multiply it by the signal, the obtained coefficient only refers to this

frequency. On the other hand, a wavelet is composed of a mixture of frequencies (which is indicated by its own Fourier transform). The wavelet coefficients refer to this mixture of frequencies.

5.21.7 The Hybrid Transformations Allow an Optimal Adaptation to Transitory Complex Signals

Starting from the analytic observation concerning wavelet and Fourier transforms, the creation of “hybrid” tools turned out a necessity for an optimal adaptation to the structure of transitory signals, i.e. signals composed of different sub-structures. *The wavelet packets*, for example, are the product of a *wavelet* with an *oscillating function*. The *wavelet* detects the abrupt change, whereas the *oscillations reveal the regular variations*. This technique is used for example to study the *turbulence phenomena*.

Moreover, the algorithm of the “best basis (Coifman)” identifies the signal and in so doing directs its analysis towards:

- *Fourier transforms* for the *periodic “patterns”*
- *Wavelets* for the *irregular signals with strong short variations*
- *Wavelet packets* mentioned above, or towards the *Malvar wavelets* which are an evolution of wavelet packets (i.e. an attack, a plate and a decreasing of the analyzing function)

The algorithm selects, for each signal, the basis which will encode it with an optimal concision.

Finally, the *Pursuit* algorithm is better adapted than the “best basis” to the decomposition of non-stationary signals. This algorithm finds for each part of the signal the wave which resembles the most to it. Instead of seeking the optimal representation for the entire signal, we seek the optimal representation for each characteristic of the signal. To this end, the algorithm browses a wave and wavelet dictionary and selects the most resembling to the part of the gradually analyzed signal. The window of the algorithm is a *Gaussian* of variable size which respects the non-stationarities, modified by sinusoids of different frequencies. Starting from this structure it is possible to build an infinity of bases. The dictionary can be very large.