1 Warm-up: the 1-D continuous wavelet transform

1.1 What is wavelet analysis?

Wavelet analysis is a particular time- or space-scale representation of signals that has found a wide range of applications in physics, signal processing and applied mathematics in the last few years. In order to get a feeling for it and to understand its success, we consider first the case of one-dimensional signals. Actually the discussion in this introductory chapter is mostly qualitative. All the mathematically relevant properties will be described precisely and proved systematically in the next chapter for the two-dimensional case, which is the proper subject of this book.

It is a fact that most real life signals are nonstationary (that is, their statistical properties change with time) and they usually cover a wide range of frequencies. Many signals contain transient components, whose appearance and disappearance are physically very significant. Also, characteristic frequencies may drift in time (e.g., in geophysical time series – one calls them pseudo-frequencies). In addition, there is often a direct correlation between the characteristic frequency of a given segment of the signal and the time duration of that segment. Low frequency pieces tend to last for a long interval, whereas high frequencies occur in general for a short moment only. Human speech signals are typical in this respect: vowels have a relatively low mean frequency and last quite a long time, whereas consonants contain a wide spectrum, up to very high frequencies, especially in the attack, but they are very short.

Clearly standard Fourier analysis is inadequate for treating such signals. Strictly speaking, it applies only to stationary signals, and it loses all information about the time localization of a given frequency component. In addition, it is very uneconomical. When the signal is almost flat, and thus uninteresting, one still has to sum an infinite alternating series to reproduce it. Worse yet, Fourier analysis is highly unstable with respect to perturbation, because of its *global* character. For instance, if one adds an extra term, with a very small amplitude, to a linear superposition of sine waves, the signal will barely be modified, but the Fourier spectrum will be completely perturbed. This does not happen if the signal is represented in terms of *localized* components. Indeed, as we shall see shortly, the basic idea of the wavelet transform is to decompose a signal *locally* into contributions living at different scales. This is a marked contrast with the Fourier components, which are sinusoidal waves repeating themselves indefinitely. As such, it is difficult to give them any physical reality. If a piece of audio signal is identically zero, it is because no sound is emitted, not because the Fourier components necessary to represent the zero signal interfere destructively. These components are a mathematical construction, rather than a genuine physical phenomenon. To quote J. Ville [364]:

> Si nous considérons en effet un morceau de musique ... et qu'une note, *la* par exemple, figure une fois dans le morceau, l'analyse harmonique [de Fourier] nous présentera la fréquence correspondante avec une certaine amplitude et une certaine phase, sans localiser le *la* dans le temps. Or, il est évident qu'au cours du morceau il est des instants où l'on n'entend pas le *la*. La représentation est néanmoins mathématiquement correcte, parce que les phases des notes voisines du *la* sont agencées de manière à détruire cette note par interférence lorsqu'on ne l'entend pas et à la renforcer, également par interférence, lorsqu'on l'entend; mais s'il y a dans cette conception une habileté qui honore l'analyse mathématique, il ne faut pas se dissimuler qu'il y a également une défiguration de la réalité: en effet, quand on n'entend pas le *la*, la raison véritable est que le *la* n'est pas émis.

That is,

> If we consider a piece of music ... and if a note, an *A* for instance, appears once in that piece, Fourier analysis will yield the corresponding frequency with a certain amplitude and a certain phase, without localizing the *A* in time. Clearly the *A* will not be heard at certain instants. Yet the representation is mathematically correct, because the phases of the neighboring notes conspire to suppress the *A* by interference when it is not heard and to enhance it, again by interference, when it is heard. However, although this conception shows a skillfulness that honors mathematical analysis, one should not hide the fact that it also distorts reality: indeed, when the *A* is not heard, the true reason is that the *A* is not emitted.

Another eloquent comment along the same line by L. de Broglie may be found, together with the one above, in [Fla93; p.9].

Facing these problems, signal analysts turn to *time–frequency* representations. The idea is that one needs *two* parameters: one, called a, characterizes the frequency, the other one, b, indicates the position in the signal. This concept of a time–frequency representation is in fact quite old and familiar. The most obvious example is simply a musical score (see Figure 1.1). Clearly, it is not sufficient to give the pitch of a given note, that is, the frequency to which it corresponds, it is also important to know when to play it (time information)!

Let $s(x)$ be a finite energy signal, that is, a square integrable function $s \in L^2(\mathbb{R}, dx)$. In most cases, x will be a time variable and the (Fourier) conjugate quantity a frequency,

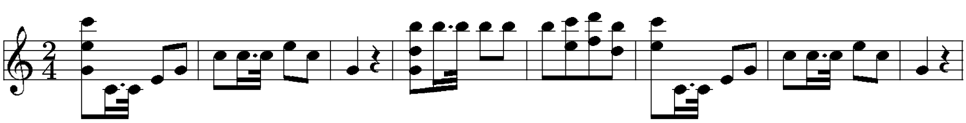

Fig. 1.1. A traditional time–frequency representation of a signal (from Mozart's Don Giovanni, Act 1).

but in general x simply represents position in the signal. Thus, following [Dau92], we prefer to keep a neutral notation (x, ξ) for the couple of conjugate variables, instead of the more familiar (t, ω). Accordingly, the Fourier transform of the signal s is defined by

$$\widehat{s}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, e^{-i\xi x} s(x). \tag{1.1}$$

If one requires the transform to be *linear*, a general time–frequency transform of the signal s will take the form:

$$s(x) \mapsto S(b, a) = \int_{-\infty}^{\infty} dx\, \overline{\psi_{b,a}(x)}\, s(x)\,, \tag{1.2}$$

where $\psi_{b,a}$ is the analyzing function. Within this class, two time–frequency transforms stand out as particularly simple and efficient: the windowed (or short time) Fourier transform (WFT) and the wavelet transform (WT). For both of them, the analyzing function $\psi_{b,a}$ is obtained by acting on a basic (or mother) function ψ, in particular, b is simply a time translation. The essential difference between the two is in the way the frequency parameter a is introduced:

(1) *Windowed Fourier transform:*

$$\psi_{b,a}(x) = e^{i(x-b)/a}\, \psi(x - b). \tag{1.3}$$

Here ψ is a window function and the a-dependence is a modulation ($1/a \sim$ frequency); the window has constant width, but the smaller a, the larger the number of oscillations in the window (see Figure 1.2 (left)).

(2) *Wavelet transform:*

$$\psi_{b,a}(x) = \frac{1}{\sqrt{a}}\, \psi\left(\frac{x-b}{a}\right). \tag{1.4}$$

The action of a on the function ψ (which must be oscillating, see below) is a dilation ($a > 1$) or a contraction ($a < 1$): the shape of the function is unchanged, it is simply spread out or squeezed (see Figure 1.2 (right)). In particular, the effective support of $\psi_{b,a}$ varies as a function of a.

The windowed Fourier transform was originally introduced by Gabor (actually in a discretized version), with the window function ψ taken as a Gaussian; for this reason, it is sometimes called the *Gabor transform*. With this choice, the function $\psi_{b,a}$ is simply a canonical (harmonic oscillator) coherent state [Kla85], as one sees immediately by writing $1/a = p$. Since the new variables are the time (position) b and the frequency $1/a$, the Gabor transform yields a genuine time–frequency representation of the signal. As for the wavelet transform, the variables are b and the scale a (or pitch in the case of music), hence we shall speak rather of a *time-scale* representation.

We may remark here that the resemblance between the windowed Fourier transform and the wavelet transform is not accidental. They are both particular instances of a large

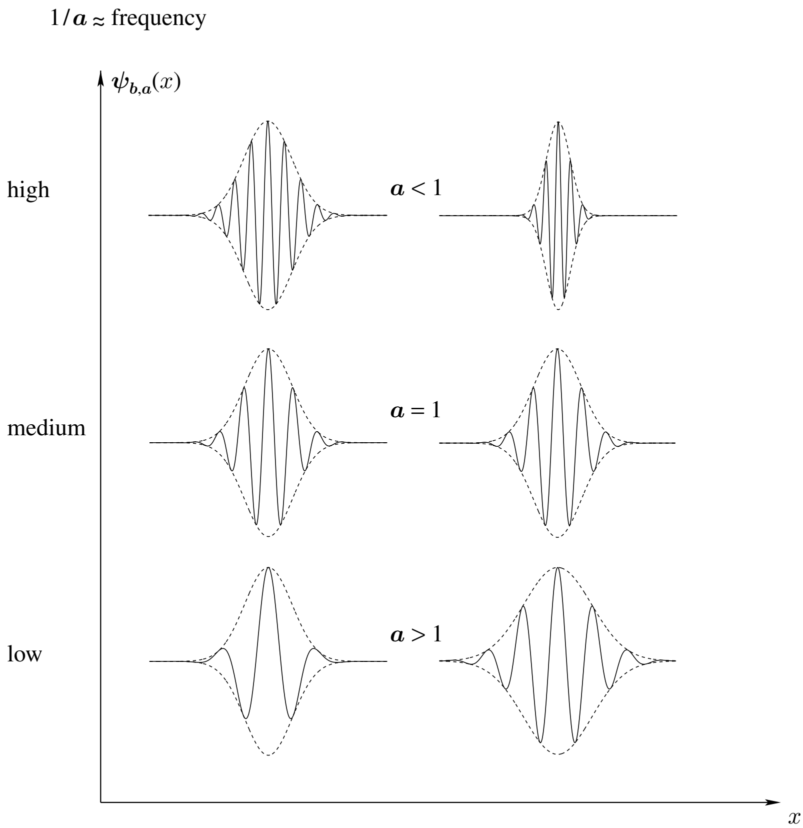

Fig. 1.2. The function $\psi_{b,a}(x)$ for different values of the scale parameter a, in the case of the windowed Fourier transform (left) and the wavelet transform (right). The quantity $1/a$, which corresponds to a frequency, increases from bottom to top.

class of integral transforms constructed by the formalism of coherent states [Ali00]. This general analysis, however, has a more mathematical flavor and is not needed in a first approach, although it clarifies and unifies the picture considerably. Therefore, we postpone it to Chapter 6, since we want to emphasize first the practical aspects of the wavelet transform.

One should note that the assumption of linearity is nontrivial, for there exists a whole class of quadratic or, more properly, sesquilinear time–frequency representations. The prototype is the so-called Wigner–Ville transform, introduced originally by E.P. Wigner [373] in quantum mechanics (in 1932!) and extended by J. Ville [364] to signal analysis:

$$W_s(b, \xi) = \int_{-\infty}^{+\infty} dx\, e^{-i\xi x}\, \overline{s(b - \frac{x}{2})}\, s(b + \frac{x}{2}), \quad \xi = 1/a. \tag{1.5}$$

Note that the signal $s(x)$ is usually a real function, but, in quantum mechanics, $s(x)$ represents a wave function, and is thus in general complex. This transform is entirely intrinsic to the signal, since it does not contain any extra function (wavelet, window)

Cambridge University Press
978-0-521-06519-1 - Two-Dimensional Wavelets and their Relatives
Jean-Pierre Antoine, Romain Murenzi, Pierre Vandergheynst and Syed Twareque Ali
Excerpt
More information

that inevitably influences the result. On the other hand, it is quadratic, which implies the appearance of interference terms whenever the signal is a superposition of two components. In order to minimize these as much as possible, one usually smoothes the Wigner–Ville transform with some function Φ, thus obtaining a whole class of quadratic transforms, called Cohen's class [109,Fla93], of the general form:

$$C_s(b, \xi) = \iint_{\mathbb{R}^2} db'\, d\xi'\; \Phi(b - b', \xi - \xi')\, W_s(b', \xi'). \tag{1.6}$$

An example is the so-called smoothened pseudo-Wigner–Ville distribution,

$$SPW_s(b, \xi) = \int_{-\infty}^{+\infty} db'\, g(b - b') \int_{-\infty}^{+\infty} dx\, h(x)\, e^{-i\xi x}\, \overline{s(b' - x/2)}\, s(b' + x/2), \tag{1.7}$$

corresponding to a factorizable kernel $\Phi(b, \xi) = (2\pi)^{-1/2} g(b)\, \widehat{h}(\xi)$, where $\widehat{h}$ denotes the Fourier transform of h. Further information about quadratic transforms may be found in [Fla93], and as a general survey for time–frequency methods, we refer to [Gro01].

1.2 The continuous wavelet transform

Actually one should distinguish two different versions of the wavelet transform, the *continuous* WT (CWT) and the *discrete* (or more properly, discrete time) WT (DWT) [Dau92,Hol95]. The CWT plays the same rôle as the Fourier transform and is mostly used for analysis and feature detection in signals, whereas the DWT is the analog of the Discrete Fourier Transform (see for instance [Bur98] or [326]) and is more appropriate for data compression and signal reconstruction. The situation may be caricatured by saying that the CWT is more natural to the physicist, while the DWT is more congenial to the signal analyst and the numericist. The continuous wavelet transform is the main topic of this book. Nevertheless, for the sake of comparison, we will give short overviews of the discrete WT, both in one and two dimensions.

The two versions of the WT are based on the same transformation formula, which reads, from (1.2) and (1.4):

$$S(b, a) = |a|^{-1/2} \int_{-\infty}^{\infty} dx\; \overline{\psi\left(\frac{x - b}{a}\right)}\, s(x), \tag{1.8}$$

where $a \neq 0$ is a scale parameter and $b \in \mathbb{R}$ a translation parameter (one often imposes only $a > 0$, which is more natural, but makes formulas slightly more complicated; see Chapter 6). Equivalently, in terms of Fourier transforms:

$$S(b, a) = |a|^{1/2} \int_{-\infty}^{\infty} d\xi\; \overline{\widehat{\psi}(a\xi)}\, \widehat{s}(\xi)\, e^{i\xi b}. \tag{1.9}$$

In these relations, s is a square integrable function, representing a finite energy signal, and the function ψ, the analyzing wavelet, is assumed to be well localized *both* in the space (or time) domain and in the frequency domain. In addition ψ must satisfy the following admissibility condition, which guarantees the invertibility of the WT:

$$c_\psi \equiv 2\pi \int_{-\infty}^{\infty} d\xi \, \frac{|\widehat{\psi}(\xi)|^2}{|\xi|} < \infty. \tag{1.10}$$

In most cases, this condition may be reduced to the (only slightly weaker) requirement that ψ has zero mean:

$$\widehat{\psi}(0) = 0 \iff \int_{-\infty}^{\infty} dx \, \psi(x) = 0. \tag{1.11}$$

Intuitively, it expresses the fact that a wavelet must be an oscillating function, real or complex ("little wave"). This is often thought to be the origin of the term "wavelet", but it is *not* the case historically. Indeed the word was widely in use in the geophysics community, with quite a different meaning, when it was introduced by Grossmann and Morlet [205,206] in the present sense, under the name "wavelets of constant shape" – but, of course, this lengthy nomenclature did not survive the very first founding paper!

The wavelet ψ is said to be *progressive* if its Fourier transform $\widehat{\psi}(\xi)$ is real and vanishes identically for $\xi \leqslant 0$. (In the signal processing community, a signal with this property is called *analytic*, following the terminology introduced by J. Ville [364].) In addition, ψ is often required to have a certain number of *vanishing moments*:

$$\int_{-\infty}^{\infty} dx \, x^n \, \psi(x) = 0, \;\; n = 0, 1, \ldots N. \tag{1.12}$$

This property improves the efficiency of ψ at detecting singularities in the signal, since it is then blind to polynomials up to order N, which constitute the smoothest part of the signal.

Notice that, instead of (1.8), which defines the WT as the scalar product of the signal s with the transformed wavelet $\psi_{b,a}$, $S(b, a)$ may also be seen as the convolution of s with the scaled, flipped and conjugated wavelet $\psi_a^\#(x) = |a|^{-1/2}\, \overline{\psi(-x/a)}$:

$$S(b, a) = (\psi_a^\# * s)(b) = \int_{-\infty}^{\infty} dx \, \psi_a^\#(b - x)\, s(x). \tag{1.13}$$

In other words, the CWT acts as a *filter* with a function of zero mean.

This property is crucial, for the main virtues of the CWT follow from it, combined with the support properties of ψ. Indeed, we must assume that ψ and $\widehat{\psi}$ are as well localized as possible, but respecting, of course, the Fourier uncertainty principle. This means that, up to minute corrections, the product of the lengths of the supports of ψ and $\widehat{\psi}$ is bounded from below by a fixed constant, usually taken as 1/2. Equivalently, the product of the variances of the distributions $|\psi|^2$ and $|\widehat{\psi}|^2$ is bounded from below. More precisely, one defines the centers of gravity (which may in fact be normalized to zero by a suitable redefinition of the coordinates):

$$x_0 = \int_{-\infty}^{\infty} dx\; x\, |\psi(x)|^2, \quad \xi_0 = \int_{-\infty}^{\infty} d\xi\; \xi\, |\widehat{\psi}(\xi)|^2, \tag{1.14}$$

and the corresponding variances

$$(\Delta x)^2 = \|\psi\|^{-2} \int_{-\infty}^{\infty} dx\; (x - x_0)^2\, |\psi(x)|^2; \tag{1.15}$$

$$(\Delta \xi)^2 = \|\psi\|^{-2} \int_{-\infty}^{\infty} d\xi\; (\xi - \xi_0)^2\, |\widehat{\psi}(\xi)|^2. \tag{1.16}$$

Then the Fourier uncertainty theorem [Fla93] says that

$$\Delta x\, \Delta\xi \geqslant \frac{1}{2}\,. \tag{1.17}$$

Under these assumptions, the transformed wavelets $\psi_{b,a}$ and $\widehat{\psi_{b,a}}$ are also well localized. Therefore, the WT $s \mapsto S$ performs a *local filtering*, both in time (b) and in scale (a). The transform $S(b, a)$ is nonnegligible only when the wavelet $\psi_{b,a}$ matches the signal, that is, the WT selects the part of the signal, if any, that lives around the time b and the scale a.

In addition, if $\widehat{\psi}$ has a numerical support (bandwidth) of width $\Delta\xi$, then $\widehat{\psi_{b,a}}$ has a numerical support of width $\Delta\xi/|a|$. Thus, remembering that $1/a$ behaves like a frequency, we conclude that the WT works at constant *relative* bandwidth, that is, $\Delta\xi/\xi =$ constant. This implies that it is very efficient at high frequency, i.e., small scales, in particular for the detection of singularities in the signal. By comparison, in the case of the Gabor transform, the support of $\widehat{\psi_{b,a}}$ keeps the same width $\Delta\xi$ for all a, that is, the WFT works at constant bandwidth, $\Delta\xi =$ constant. This difference in behavior is often the key factor in deciding whether one should choose the WFT or the WT in a given physical problem.

Another crucial fact is that the transformation $s(x) \mapsto S(b, a)$ may be inverted exactly, which yields a reconstruction formula (this is only the simplest one, others are possible, for instance using different wavelets for the decomposition and the reconstruction):

$$s(x) = c_\psi^{-1} \int_{-\infty}^{\infty} db \int_{-\infty}^{\infty} \frac{da}{a^2}\; \psi_{b,a}(x)\, S(b, a), \tag{1.18}$$

where the normalization constant c_ψ is given in (1.10) (incidentally, this relation shows why the admissibility condition $c_\psi < \infty$ is required for the transformation to be invertible). This means that the WT provides a decomposition of the signal as a linear superposition of the wavelets $\psi_{b,a}$ with coefficients $S(b, a)$. Notice that the natural measure on the parameter space (a, b) is $da\, db/a^2$, and it is invariant not only under time translation, but also under dilation. This fact is important, for it suggests that these geometric transformations play an essential rôle in the CWT.

One should emphasize here that the choice of the normalization factor $|a|^{-1/2}$ in (1.4) or (1.8) is not essential. This choice makes the transform unitary: $\|\psi_{b,a}\|_2 = \|\psi\|_2$ and also $\|S\|_2 = \|s\|_2$, where $\|\cdot\|_2$ denotes the L^2 norm in the appropriate variables (the squared norm is interpreted as the total energy of the signal). In practice, one often uses instead a factor a^{-1}, which has the advantage of giving more weight to the small scales, i.e., the high frequency part (which contains the singularities of the signal, if any). Thus, defining

$$\psi_{(b,a)} = \frac{1}{|a|}\,\psi\left(\frac{x-b}{a}\right), \tag{1.19}$$

we obtain the so-called L^1-normalized transform:

$$\check{S}(b,a) = \langle \psi_{(b,a)}|s\rangle \equiv |a|^{-1}\int_{-\infty}^{\infty} dx\; \overline{\psi\left(\frac{x-b}{a}\right)}\, s(x), \tag{1.20}$$

which preserves the L^1-norm of the signal, as follows immediately from the corresponding convolution formula

$$\check{S}(b,a) = (\psi_a^{\#} * s)(b), \tag{1.21}$$

where $\psi_a^{\#}(x) = |a|^{-1}\,\overline{\psi(-x/a)}$. Thus indeed $\|\psi_a^{\#}\|_1 = \|\psi\|_1$ and $\|\check{S}\|_1 = \|s\|_1$, where $\|\cdot\|_1$ denotes the L^1-norm in the corresponding variables.

1.2.1 Examples

In order to fix ideas, we exhibit here two simple examples of wavelets, both in the time domain and in the frequency domain.

(1) *The Mexican hat wavelet*

This wavelet is simply the second derivative of a Gaussian:

$$\psi_{\rm H}(x) = (1-x^2)\,\exp(-\tfrac{1}{2}x^2),\;\; \widehat{\psi}_{\rm H}(\xi) = \xi^2\,\exp(-\tfrac{1}{2}\xi^2). \tag{1.22}$$

(2) *The Morlet wavelet*

This wavelet is essentially a plane wave within a Gaussian window:

$$\psi_{\rm M}(x) = \exp(ik_o x)\,\exp(-\tfrac{1}{2}x^2) + c(x),\;\; \widehat{\psi}_{\rm M}(\xi) = \exp(-\tfrac{1}{2}(\xi-\xi_o)^2) + \widehat{c}(\xi). \tag{1.23}$$

Here the correction term c must be added in order to satisfy the admissibility condition (1.11), but in practice one will arrange that this term be numerically negligible ($\leqslant 10^{-4}$) and thus can be omitted (it suffices to choose the basic frequency $|\xi_o|$ large enough, typically $|\xi_o| > 5.5$).

These two wavelets have very different properties and, naturally, they will be used in quite different situations. Typically, the Mexican hat is sensitive to singularities in the signal, and it yields a genuine time-scale analysis. On the other hand, since it is complex, the Morlet wavelet will catch the phase of the signal, hence will be sensitive to frequencies, and will lead to a time-frequency analysis, somewhat closer to a Gabor

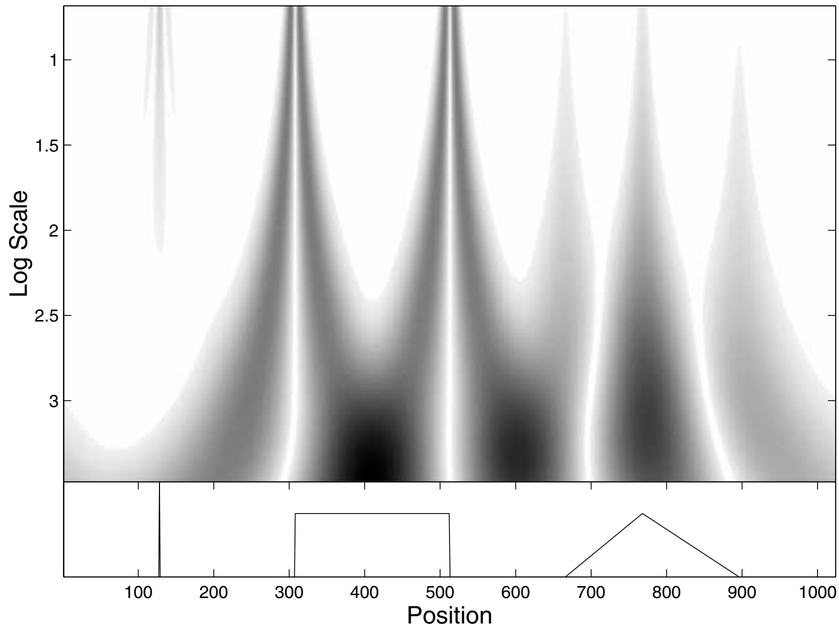

Fig. 1.3. Wavelet analysis with a Mexican hat wavelet of the discontinuous signal *bumps* (shown in the bottom panel).

analysis. In both cases, additional flexibility is obtained by adding a width parameter to the Gaussian (see (3.8) in the equivalent 2-D situation).

As an illustration of the performance of the CWT as a singularity scanner, we first show in Figure 1.3 the analysis with a Mexican hat wavelet of a discontinuous signal, called *bumps* and consisting of three pieces, a δ function, a boxcar function and a tent function. Clearly the wavelet locates all discontinuities in the signal and in its successive derivatives well. However, if one wants to discriminate between the various types of singularities, one has to invoke the concept of vanishing moment, defined in (1.12). Let us consider the successive derivatives of a Gaussian:

$$\psi_{\mathrm{H}}^{(n)}(x) = -\frac{d^n}{dx^n}\, \exp(-\tfrac{1}{2}x^2). \tag{1.24}$$

For increasing n, these wavelets have more and more vanishing moments, and are thus sensitive to increasingly sharper details. As an example, we consider a continuous signal obtained by glueing together an arc of parabola (the so-called function x_+^2) and a linear piece and we analyze it successively with the first three derivatives of a Gaussian, $\psi_{\mathrm{H}}^{(n)}(x)$, $n = 1, 2, 3$. The result is shown in Figure 1.4. In (a), the first-order wavelet $\psi_{\mathrm{H}}^{(1)}$ has only one vanishing moment, hence it sees the full content of the two pieces of the signal. In (b), the second-order wavelet $\psi_{\mathrm{H}}^{(2)}(x) \equiv \psi_{\mathrm{H}}$ does not see the linear part anymore, only the singularities at the two ends, but still sees the quadratic piece on the left (in technical terms, one would say that this wavelet is blind to a linear trend in the signal). In (c), finally, the third-order wavelet correctly erases both pieces of the

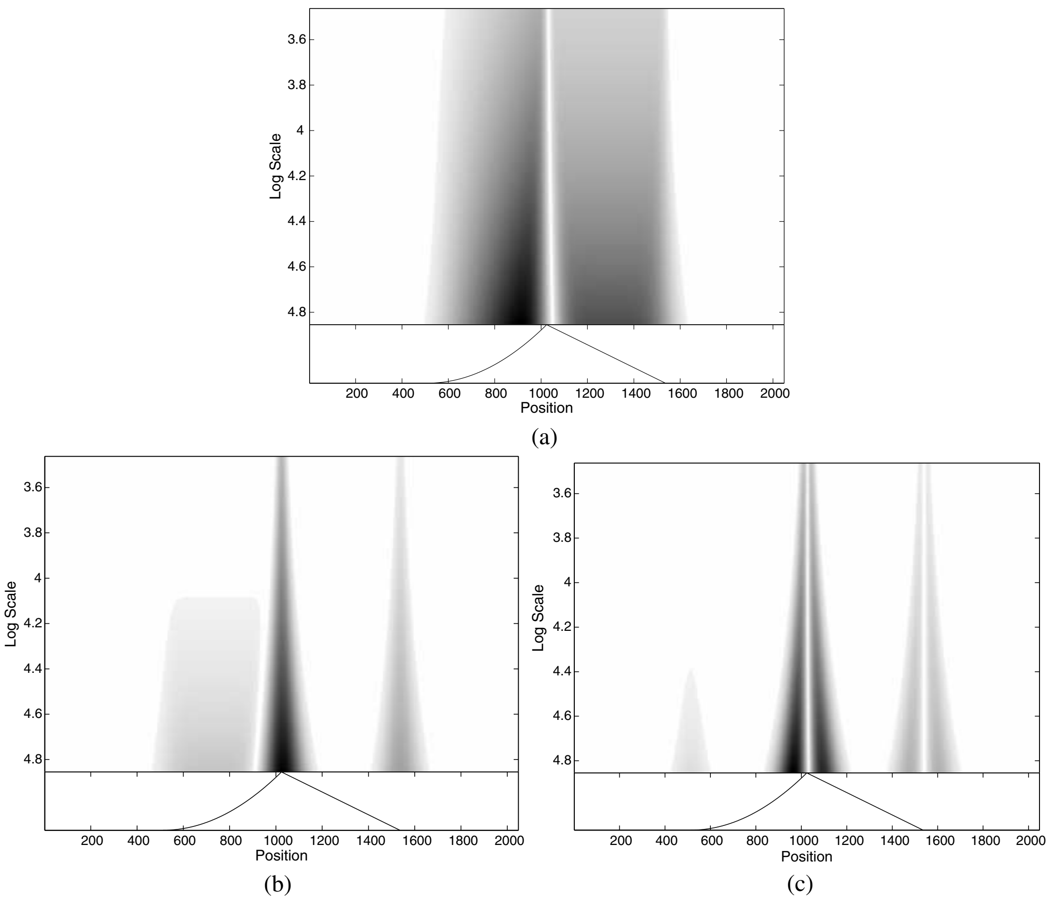

Fig. 1.4. Analysis of a composite signal (bottom panel) with successive derivatives of a Gaussian. (a) First order; (b) second order; (c) third order.

signal, keeping only the three singularities. This example shows the advantage of the *local* filtering effect of the CWT. Notice that a Gabor analysis would be utterly unable to achieve such a discrimination between singularities, let alone to detect them!

As a direct application of this behavior, an interesting technique has been designed by A.Arnéodo *et al.* [49], which consists in analyzing the same signal with several wavelets $\psi_{\mathrm{H}}^{(n)}$, for different n. The features common to all the transforms surely belong to the signal, they are not artifacts of the analysis.

1.3 Discretization of the CWT, frames

All this concerns the continuous WT (CWT). But, in practice, for numerical purposes, the transform must be *discretized*, by restricting the parameters a and b in (1.8) to the points of a discrete lattice $\Gamma = \{a_j, b_k, j, k \in \mathbb{Z}\}$ in the (a, b)-(half)-plane. Then we