



---

# INTRODUCTION TO A TRANSIENT WORLD

**A**fter a few minutes in a restaurant we cease to notice the annoying hub-bub of surrounding conversations, but a sudden silence reminds us of the presence of neighbors. Our attention is clearly attracted by transients and movements as opposed to stationary stimuli, which we soon ignore. Concentrating on transients is probably a strategy for selecting important information from the overwhelming amount of data recorded by our senses. Yet, classical signal processing has devoted most of its efforts to the design of time-invariant and space-invariant operators, that modify stationary signal properties. This has led to the indisputable hegemony of the Fourier transform, but leaves aside many information-processing applications.

The world of transients is considerably larger and more complex than the garden of stationary signals. The search for an ideal Fourier-like basis that would simplify most signal processing is therefore a hopeless quest. Instead, a multitude of different transforms and bases have proliferated, among which wavelets are just one example. This book gives a guided tour in this jungle of new mathematical and algorithmic results, while trying to provide an intuitive sense of orientation. Major ideas are outlined in this first chapter. Section 1.5.2 serves as a travel guide and introduces the *reproducible experiment* approach based on the WAVELAB and LASTWAVE softwares. It also discusses the use of *level numbers*—landmarks that can help the reader keep to the main roads.

### 1.1 FOURIER KINGDOM

The Fourier transform rules over linear time-invariant signal processing because sinusoidal waves  $e^{i\omega t}$  are eigenvectors of linear time-invariant operators. A linear time-invariant operator  $L$  is entirely specified by the eigenvalues  $\hat{h}(\omega)$ :

$$\forall \omega \in \mathbb{R}, \quad L e^{i\omega t} = \hat{h}(\omega) e^{i\omega t}. \quad (1.1)$$

To compute  $Lf$ , a signal  $f$  is decomposed as a sum of sinusoidal eigenvectors  $\{e^{i\omega t}\}_{\omega \in \mathbb{R}}$ :

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} d\omega. \quad (1.2)$$

If  $f$  has finite energy, the theory of Fourier integrals presented in Chapter 2 proves that the amplitude  $\hat{f}(\omega)$  of each sinusoidal wave  $e^{i\omega t}$  is the Fourier transform of  $f$ :

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt. \quad (1.3)$$

Applying the operator  $L$  to  $f$  in (1.2) and inserting the eigenvector expression (1.1) gives

$$Lf(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \hat{h}(\omega) e^{i\omega t} d\omega. \quad (1.4)$$

The operator  $L$  amplifies or attenuates each sinusoidal component  $e^{i\omega t}$  of  $f$  by  $\hat{h}(\omega)$ . It is a frequency *filtering* of  $f$ .

As long as we are satisfied with linear time-invariant operators, the Fourier transform provides simple answers to most questions. Its richness makes it suitable for a wide range of applications such as signal transmissions or stationary signal processing. However, if we are interested in transient phenomena—a word pronounced at a particular time, an apple located in the left corner of an image—the Fourier transform becomes a cumbersome tool.

The Fourier coefficient is obtained in (1.3) by correlating  $f$  with a sinusoidal wave  $e^{i\omega t}$ . Since the support of  $e^{i\omega t}$  covers the whole real line,  $\hat{f}(\omega)$  depends on the values  $f(t)$  for all times  $t \in \mathbb{R}$ . This global “mix” of information makes it difficult to analyze any local property of  $f$  from  $\hat{f}$ . Chapter 4 introduces local time-frequency transforms, which decompose the signal over waveforms that are well localized in time and frequency.

### 1.2 TIME-FREQUENCY WEDDING

The uncertainty principle states that the energy spread of a function and its Fourier transform cannot be simultaneously arbitrarily small. Motivated by quantum mechanics, in 1946 the physicist Gabor [187] defined elementary time-frequency atoms as waveforms that have a minimal spread in a time-frequency plane. To measure time-frequency “information” content, he proposed decomposing signals over these elementary atomic waveforms. By showing that such decompositions

are closely related to our sensitivity to sounds, and that they exhibit important structures in speech and music recordings, Gabor demonstrated the importance of localized time-frequency signal processing.

Chapter 4 studies the properties of windowed Fourier and wavelet transforms, computed by decomposing the signal over different families of time-frequency atoms. Other transforms can also be defined by modifying the family of time-frequency atoms. A unified interpretation of local time-frequency decompositions follows the time-frequency energy density approach of Ville. In parallel to Gabor's contribution, in 1948 Ville [342], who was an electrical engineer, proposed analyzing the time-frequency properties of signals  $f$  with an energy density defined by

$$P_V f(t, \omega) = \int_{-\infty}^{+\infty} f\left(t + \frac{\tau}{2}\right) f^*\left(t - \frac{\tau}{2}\right) e^{-i\tau\omega} d\tau.$$

Once again, theoretical physics was ahead, since this distribution had already been introduced in 1932 by Wigner [351] in the context of quantum mechanics. Chapter 4 explains the path that relates Wigner-Ville distributions to windowed Fourier and wavelet transforms, or any linear time-frequency transform.

### 1.2.1 Windowed Fourier Transform

Gabor atoms are constructed by translating in time and frequency a time window  $g$ :

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

The energy of  $g_{u,\xi}$  is concentrated in the neighborhood of  $u$  over an interval of size  $\sigma_t$ , measured by the standard deviation of  $|g|^2$ . Its Fourier transform is a translation by  $\xi$  of the Fourier transform  $\hat{g}$  of  $g$ :

$$\hat{g}_{u,\xi}(\omega) = \hat{g}(\omega - \xi) e^{-iu(\omega - \xi)}. \quad (1.5)$$

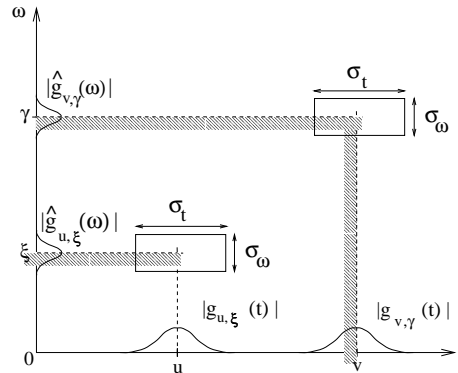
The energy of  $\hat{g}_{u,\xi}$  is therefore localized near the frequency  $\xi$ , over an interval of size  $\sigma_\omega$ , which measures the domain where  $\hat{g}(\omega)$  is non-negligible. In a time-frequency plane  $(t, \omega)$ , the energy spread of the atom  $g_{u,\xi}$  is symbolically represented by the Heisenberg rectangle illustrated by Figure 1.1. This rectangle is centered at  $(u, \xi)$  and has a time width  $\sigma_t$  and a frequency width  $\sigma_\omega$ . The uncertainty principle proves that its area satisfies

$$\sigma_t \sigma_\omega \geq \frac{1}{2}.$$

This area is minimum when  $g$  is a Gaussian, in which case the atoms  $g_{u,\xi}$  are called *Gabor functions*.

The windowed Fourier transform defined by Gabor correlates a signal  $f$  with each atom  $g_{u,\xi}$ :

$$Sf(u, \xi) = \int_{-\infty}^{+\infty} f(t) g_{u,\xi}^*(t) dt = \int_{-\infty}^{+\infty} f(t) g(t-u) e^{-i\xi t} dt. \quad (1.6)$$



**FIGURE 1.1** Time-frequency boxes (“Heisenberg rectangles”) representing the energy spread of two Gabor atoms.

It is a Fourier integral that is localized in the neighborhood of  $u$  by the window  $g(t-u)$ . This time integral can also be written as a frequency integral by applying the Fourier Parseval formula (2.25):

$$Sf(u, \xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \hat{g}_{u,\xi}^*(\omega) d\omega. \quad (1.7)$$

The transform  $Sf(u, \xi)$  thus depends only on the values  $f(t)$  and  $\hat{f}(\omega)$  in the time and frequency neighborhoods where the energies of  $g_{u,\xi}$  and  $\hat{g}_{u,\xi}$  are concentrated. Gabor interprets this as a “quantum of information” over the time-frequency rectangle illustrated in Figure 1.1.

When listening to music, we perceive sounds that have a frequency that varies in time. Measuring time-varying harmonics is an important application of windowed Fourier transforms in both music and speech recognition. A spectral line of  $f$  creates high amplitude windowed Fourier coefficients  $Sf(u, \xi)$  at frequencies  $\xi(u)$  that depend on the time  $u$ . The time evolution of such spectral components is therefore analyzed by following the location of large amplitude coefficients.

### 1.2.2 Wavelet Transform

In reflection seismology, Morlet knew that the modulated pulses sent underground have a duration that is too long at high frequencies to separate the returns of fine, closely-spaced layers. Instead of emitting pulses of equal duration, he thus thought of sending shorter waveforms at high frequencies. Such waveforms are simply obtained by scaling a single function called a *wavelet*. Although Grossmann was working in theoretical physics, he recognized in Morlet’s approach some ideas that were close to his own work on coherent quantum states. Nearly forty years after Gabor, Morlet and Grossmann reactivated a fundamental collaboration between theoretical physics and signal processing, which led to the formalization of the

continuous wavelet transform [200]. Yet, these ideas were not totally new to mathematicians working in harmonic analysis, or to computer vision researchers studying multiscale image processing. It was thus only the beginning of a rapid catalysis that brought together scientists with very different backgrounds, first around coffee tables, then in more luxurious conferences.

A wavelet  $\psi$  is a function of zero average:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0,$$

which is dilated with a scale parameter  $s$ , and translated by  $u$ :

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right). \quad (1.8)$$

The wavelet transform of  $f$  at the scale  $s$  and position  $u$  is computed by correlating  $f$  with a wavelet atom

$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt. \quad (1.9)$$

**Time-Frequency Measurements** Like a windowed Fourier transform, a wavelet transform can measure the time-frequency variations of spectral components, but it has a different time-frequency resolution. A wavelet transform correlates  $f$  with  $\psi_{u,s}$ . By applying the Fourier Parseval formula (2.25), it can also be written as a frequency integration:

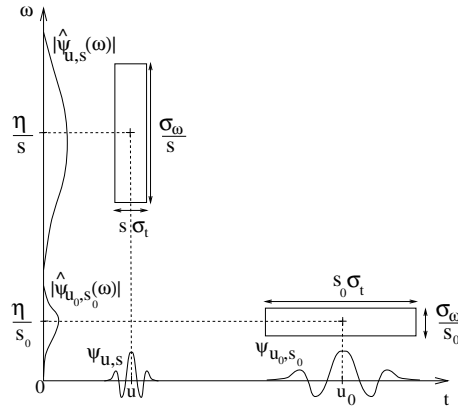
$$Wf(u,s) = \int_{-\infty}^{+\infty} f(t) \psi_{u,s}^*(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \hat{\psi}_{u,s}^*(\omega) d\omega. \quad (1.10)$$

The wavelet coefficient  $Wf(u,s)$  thus depends on the values  $f(t)$  and  $\hat{f}(\omega)$  in the time-frequency region where the energy of  $\psi_{u,s}$  and  $\hat{\psi}_{u,s}$  is concentrated. Time varying harmonics are detected from the position and scale of high amplitude wavelet coefficients.

In time,  $\psi_{u,s}$  is centered at  $u$  with a spread proportional to  $s$ . Its Fourier transform is calculated from (1.8):

$$\hat{\psi}_{u,s}(\omega) = e^{-iu\omega} \sqrt{s} \hat{\psi}(s\omega),$$

where  $\hat{\psi}$  is the Fourier transform of  $\psi$ . To analyze the phase information of signals, a complex analytic wavelet is used. This means that  $\hat{\psi}(\omega) = 0$  for  $\omega < 0$ . Its energy is concentrated in a positive frequency interval centered at  $\eta$ . The energy of  $\hat{\psi}_{u,s}(\omega)$  is therefore concentrated over a positive frequency interval centered at  $\eta/s$ , whose size is scaled by  $1/s$ . In the time-frequency plane, a wavelet atom  $\psi_{u,s}$  is symbolically represented by a rectangle centered at  $(u, \eta/s)$ . The time and frequency spread are respectively proportional to  $s$  and  $1/s$ . When  $s$  varies, the height and width of the rectangle change but its area remains constant, as illustrated by Figure 1.2.



**FIGURE 1.2** Time-frequency boxes of two wavelets  $\psi_{u,s}$  and  $\psi_{u_0,s_0}$ . When the scale  $s$  decreases, the time support is reduced but the frequency spread increases and covers an interval that is shifted towards high frequencies.

**Multiscale Zooming** The wavelet transform can also detect and characterize transients with a zooming procedure across scales. Suppose that  $\psi$  is real. Since it has a zero average, a wavelet coefficient  $Wf(u, s)$  measures the variation of  $f$  in a neighborhood of  $u$  whose size is proportional to  $s$ . Sharp signal transitions create large amplitude wavelet coefficients. Chapter 6 relates the pointwise regularity of  $f$  to the asymptotic decay of the wavelet transform  $Wf(u, s)$ , when  $s$  goes to zero. Singularities are detected by following across scales the local maxima of the wavelet transform. In images, high amplitude wavelet coefficients indicate the position of edges, which are sharp variations of the image intensity. Different scales provide the contours of image structures of varying sizes. Such multiscale edge detection is particularly effective for pattern recognition in computer vision [113].

The zooming capability of the wavelet transform not only locates isolated singular events, but can also characterize more complex multifractal signals having non-isolated singularities. Mandelbrot [43] was the first to recognize the existence of multifractals in most corners of nature. Scaling one part of a multifractal produces a signal that is statistically similar to the whole. This self-similarity appears in the wavelet transform, which modifies the analyzing scale. From the global wavelet transform decay, one can measure the singularity distribution of multifractals. This is particularly important in analyzing their properties and testing models that explain the formation of multifractals in physics.

### 1.3 BASES OF TIME-FREQUENCY ATOMS

The continuous windowed Fourier transform  $Sf(u, \xi)$  and the wavelet transform  $Wf(u, s)$  are two-dimensional representations of a one-dimensional signal  $f$ . This

indicates the existence of some redundancy that can be reduced and even removed by subsampling the parameters of these transforms.

**Frames** Windowed Fourier transforms and wavelet transforms can be written as inner products in  $\mathbf{L}^2(\mathbb{R})$ , with their respective time-frequency atoms

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

and

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

Subsampling both transforms defines a complete signal representation if any signal can be reconstructed from linear combinations of discrete families of windowed Fourier atoms  $\{g_{u_n, \xi_k}\}_{(n, k) \in \mathbb{Z}^2}$  and wavelet atoms  $\{\psi_{u_n, s_j}\}_{(j, n) \in \mathbb{Z}^2}$ . The frame theory of Chapter 5 discusses what conditions these families of waveforms must meet if they are to provide stable and complete representations.

Completely eliminating the redundancy is equivalent to building a basis of the signal space. Although wavelet bases were the first to arrive on the research market, they have quickly been followed by other families of orthogonal bases, such as wavelet packet and local cosine bases.

### 1.3.1 Wavelet Bases and Filter Banks

In 1910, Haar [202] realized that one can construct a simple piecewise constant function

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1/2 \\ -1 & \text{if } 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

whose dilations and translations generate an orthonormal basis of  $\mathbf{L}^2(\mathbb{R})$ :

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

Any finite energy signal  $f$  can be decomposed over this wavelet orthogonal basis  $\{\psi_{j, n}\}_{(j, n) \in \mathbb{Z}^2}$

$$f = \sum_{j=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \langle f, \psi_{j, n} \rangle \psi_{j, n}. \quad (1.11)$$

Since  $\psi(t)$  has a zero average, each partial sum

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

can be interpreted as detail variations at the scale  $2^j$ . These layers of details are added at all scales to progressively improve the approximation of  $f$ , and ultimately recover  $f$ .

If  $f$  has smooth variations, we should obtain a precise approximation when removing fine scale details, which is done by truncating the sum (1.11). The resulting approximation at a scale  $2^J$  is

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

For a Haar basis,  $f_J$  is piecewise constant. Piecewise constant approximations of smooth functions are far from optimal. For example, a piecewise linear approximation produces a smaller approximation error. The story continues in 1980, when Strömberg [322] found a piecewise linear function  $\psi$  that also generates an orthonormal basis and gives better approximations of smooth functions. Meyer was not aware of this result, and motivated by the work of Morlet and Grossmann he tried to prove that there exists no regular wavelet  $\psi$  that generates an orthonormal basis. This attempt was a failure since he ended up constructing a whole family of orthonormal wavelet bases, with functions  $\psi$  that are infinitely continuously differentiable [270]. This was the fundamental impulse that led to a widespread search for new orthonormal wavelet bases, which culminated in the celebrated Daubechies wavelets of compact support [144].

The systematic theory for constructing orthonormal wavelet bases was established by Meyer and Mallat through the elaboration of multiresolution signal approximations [254], presented in Chapter 7. It was inspired by original ideas developed in computer vision by Burt and Adelson [108] to analyze images at several resolutions. Digging more into the properties of orthogonal wavelets and multiresolution approximations brought to light a surprising relation with filter banks constructed with conjugate mirror filters.

**Filter Banks** Motivated by speech compression, in 1976 Croisier, Esteban and Galand [141] introduced an invertible filter bank, which decomposes a discrete signal  $f[n]$  in two signals of half its size, using a filtering and subsampling procedure. They showed that  $f[n]$  can be recovered from these subsampled signals by canceling the aliasing terms with a particular class of filters called *conjugate mirror filters*. This breakthrough led to a 10-year research effort to build a complete filter bank theory. Necessary and sufficient conditions for decomposing a signal in subsampled components with a filtering scheme, and recovering the same signal with an inverse transform, were established by Smith and Barnwell [316], Vaidyanathan [336] and Vetterli [339].

The multiresolution theory of orthogonal wavelets proves that any conjugate mirror filter characterizes a wavelet  $\psi$  that generates an orthonormal basis of  $\mathbf{L}^2(\mathbb{R})$ . Moreover, a fast discrete wavelet transform is implemented by cascading these conjugate mirror filters. The equivalence between this continuous time wavelet

theory and discrete filter banks led to a new fruitful interface between digital signal processing and harmonic analysis, but also created a culture shock that is not totally resolved.

**Continuous Versus Discrete and Finite** Many signal processors have been and still are wondering what is the point of these continuous time wavelets, since all computations are performed over discrete signals, with conjugate mirror filters. Why bother with the convergence of infinite convolution cascades if in practice we only compute a finite number of convolutions? Answering these important questions is necessary in order to understand why throughout this book we alternate between theorems on continuous time functions and discrete algorithms applied to finite sequences.

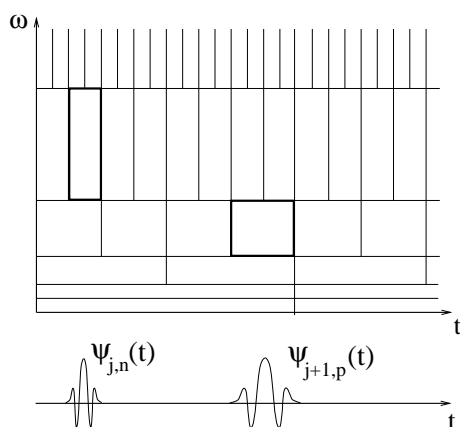
A short answer would be “simplicity”. In  $\mathbf{L}^2(\mathbb{R})$ , a wavelet basis is constructed by dilating and translating a single function  $\psi$ . Several important theorems relate the amplitude of wavelet coefficients to the local regularity of the signal  $f$ . Dilations are not defined over discrete sequences, and discrete wavelet bases have therefore a more complicated structure. The regularity of a discrete sequence is not well defined either, which makes it more difficult to interpret the amplitude of wavelet coefficients. A theory of continuous time functions gives asymptotic results for discrete sequences with sampling intervals decreasing to zero. This theory is useful because these asymptotic results are precise enough to understand the behavior of discrete algorithms.

Continuous time models are not sufficient for elaborating discrete signal processing algorithms. Uniformly sampling the continuous time wavelets  $\{\psi_{j,n}(t)\}_{(j,n) \in \mathbb{Z}^2}$  does not produce a discrete orthonormal basis. The transition between continuous and discrete signals must be done with great care. Restricting the constructions to finite discrete signals adds another layer of complexity because of border problems. How these border issues affect numerical implementations is carefully addressed once the properties of the bases are well understood. To simplify the mathematical analysis, throughout the book continuous time transforms are introduced first. Their discretization is explained afterwards, with fast numerical algorithms over finite signals.

### 1.3.2 Tilings of Wavelet Packet and Local Cosine Bases

Orthonormal wavelet bases are just an appetizer. Their construction showed that it is not only possible but relatively simple to build orthonormal bases of  $\mathbf{L}^2(\mathbb{R})$  composed of local time-frequency atoms. The completeness and orthogonality of a wavelet basis is represented by a tiling that covers the time-frequency plane with the wavelets’ time-frequency boxes. Figure 1.3 shows the time-frequency box of each  $\psi_{j,n}$ , which is translated by  $2^j n$ , with a time and a frequency width scaled respectively by  $2^j$  and  $2^{-j}$ .

One can draw many other tilings of the time-frequency plane, with boxes of minimal surface as imposed by the uncertainty principle. Chapter 8 presents



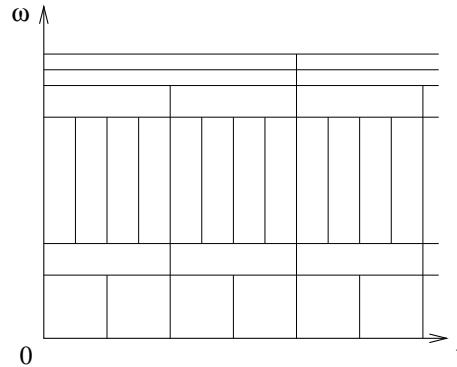
**FIGURE 1.3** The time-frequency boxes of a wavelet basis define a tiling of the time-frequency plane.

several constructions that associate large families of orthonormal bases of  $L^2(\mathbb{R})$  to such new tilings.

**Wavelet Packet Bases** A wavelet orthonormal basis decomposes the frequency axis in dyadic intervals whose sizes have an exponential growth, as shown by Figure 1.3. Coifman, Meyer and Wickerhauser [139] have generalized this fixed dyadic construction by decomposing the frequency in intervals whose bandwidths may vary. Each frequency interval is covered by the time-frequency boxes of wavelet packet functions that are uniformly translated in time in order to cover the whole plane, as shown by Figure 1.4.

Wavelet packet functions are designed by generalizing the filter bank tree that relates wavelets and conjugate mirror filters. The frequency axis division of wavelet packets is implemented with an appropriate sequence of iterated convolutions with conjugate mirror filters. Fast numerical wavelet packet decompositions are thus implemented with discrete filter banks.

**Local Cosine Bases** Orthonormal bases of  $L^2(\mathbb{R})$  can also be constructed by dividing the time axis instead of the frequency axis. The time axis is segmented in successive finite intervals  $[a_p, a_{p+1}]$ . The local cosine bases of Malvar [262] are obtained by designing smooth windows  $g_p(t)$  that cover each interval  $[a_p, a_{p+1}]$ , and multiplying them by cosine functions  $\cos(\xi t + \phi)$  of different frequencies. This is yet another idea that was independently studied in physics, signal processing and mathematics. Malvar's original construction was done for discrete signals. At the same time, the physicist Wilson [353] was designing a local cosine basis with smooth windows of infinite support, to analyze the properties of quantum coherent states. Malvar bases were also rediscovered and generalized by the



**FIGURE 1.4** A wavelet packet basis divides the frequency axis in separate intervals of varying sizes. A tiling is obtained by translating in time the wavelet packets covering each frequency interval.

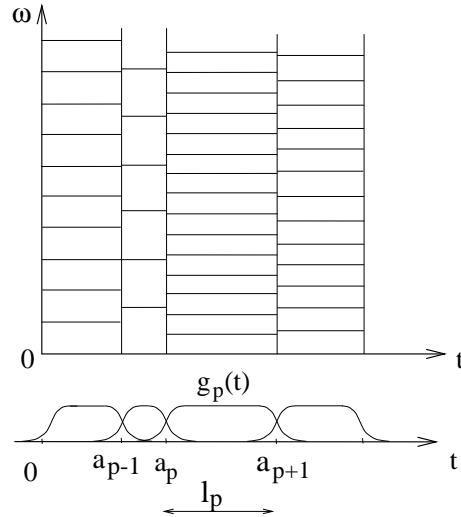
harmonic analysts Coifman and Meyer [138]. These different views of the same bases brought to light mathematical and algorithmic properties that opened new applications.

A multiplication by  $\cos(\xi t + \phi)$  translates the Fourier transform  $\hat{g}_p(\omega)$  of  $g_p(t)$  by  $\pm\xi$ . Over positive frequencies, the time-frequency box of the modulated window  $g_p(t) \cos(\xi t + \phi)$  is therefore equal to the time-frequency box of  $g_p$  translated by  $\xi$  along frequencies. The time-frequency boxes of local cosine basis vectors define a tiling of the time-frequency plane illustrated by Figure 1.5.

#### 1.4 BASES FOR WHAT?

The tiling game is clearly unlimited. Local cosine and wavelet packet bases are important examples, but many other kinds of bases can be constructed. It is thus time to wonder how to select an appropriate basis for processing a particular class of signals. The decomposition coefficients of a signal in a basis define a representation that highlights some particular signal properties. For example, wavelet coefficients provide explicit information on the location and type of signal singularities. The problem is to find a criterion for selecting a basis that is intrinsically well adapted to represent a class of signals.

Mathematical approximation theory suggests choosing a basis that can construct precise signal approximations with a linear combination of a small number of vectors selected inside the basis. These selected vectors can be interpreted as intrinsic signal structures. Compact coding and signal estimation in noise are applications where this criterion is a good measure of the efficiency of a basis. Linear and non-linear procedures are studied and compared. This will be the occasion to show that non-linear does not always mean complicated.



**FIGURE 1.5** A local cosine basis divides the time axis with smooth windows  $g_p(t)$ . Multiplications with cosine functions translate these windows in frequency and yield a complete cover of the time-frequency plane.

#### 1.4.1 Approximation

The development of orthonormal wavelet bases has opened a new bridge between approximation theory and signal processing. This exchange is not quite new since the fundamental sampling theorem comes from an interpolation theory result proved in 1935 by Whittaker [349]. However, the state of the art of approximation theory has changed since 1935. In particular, the properties of non-linear approximation schemes are much better understood, and give a firm foundation for analyzing the performance of many non-linear signal processing algorithms. Chapter 9 introduces important approximation theory results that are used in signal estimation and data compression.

**Linear Approximation** A linear approximation projects the signal  $f$  over  $M$  vectors that are chosen *a priori* in an orthonormal basis  $\mathcal{B} = \{g_m\}_{m \in \mathbb{N}}$ , say the first  $M$ :

$$f_M = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m. \quad (1.12)$$

Since the basis is orthonormal, the approximation error is the sum of the remaining squared inner products

$$\epsilon[M] = \|f - f_M\|^2 = \sum_{m=M}^{+\infty} |\langle f, g_m \rangle|^2.$$

The accuracy of this approximation clearly depends on the properties of  $f$  relative to the basis  $\mathcal{B}$ .

A Fourier basis yields efficient linear approximations of uniformly smooth signals, which are projected over the  $M$  lower frequency sinusoidal waves. When  $M$  increases, the decay of the error  $\epsilon[M]$  can be related to the global regularity of  $f$ . Chapter 9 characterizes spaces of smooth functions from the asymptotic decay of  $\epsilon[M]$  in a Fourier basis.

In a wavelet basis, the signal is projected over the  $M$  larger scale wavelets, which is equivalent to approximating the signal at a fixed resolution. Linear approximations of uniformly smooth signals in wavelet and Fourier bases have similar properties and characterize nearly the same function spaces.

Suppose that we want to approximate a class of discrete signals of size  $N$ , modeled by a random vector  $F[n]$ . The average approximation error when projecting  $F$  over the first  $M$  basis vectors of an orthonormal basis  $\mathcal{B} = \{g_m\}_{0 \leq m < N}$  is

$$\epsilon[M] = E\{\|F - F_M\|^2\} = \sum_{m=M}^{N-1} E\{|\langle F, g_m \rangle|^2\}.$$

Chapter 9 proves that the basis that minimizes this error is the Karhunen-Loève basis, which diagonalizes the covariance matrix of  $F$ . This remarkable property explains the fundamental importance of the Karhunen-Loève basis in optimal linear signal processing schemes. This is however only a beginning.

**Non-linear Approximation** The linear approximation (1.12) is improved if we choose *a posteriori* the  $M$  vectors  $g_m$ , depending on  $f$ . The approximation of  $f$  with  $M$  vectors whose indexes are in  $I_M$  is

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle g_m. \quad (1.13)$$

The approximation error is the sum of the squared inner products with vectors not in  $I_M$ :

$$\epsilon[M] = \|f - f_M\|^2 = \sum_{n \notin I_M} |\langle f, g_n \rangle|^2.$$

To minimize this error, we choose  $I_M$  to be the set of  $M$  vectors that have the largest inner product amplitude  $|\langle f, g_m \rangle|$ . This approximation scheme is non-linear because the approximation vectors change with  $f$ .

The amplitude of inner products in a wavelet basis is related to the local regularity of the signal. A non-linear approximation that keeps the largest wavelet inner products is equivalent to constructing an adaptive approximation grid, whose resolution is locally increased where the signal is irregular. If the signal has isolated singularities, this non-linear approximation is much more precise than a linear scheme that maintains the same resolution over the whole signal support. The spaces of functions that are well approximated by non-linear wavelet schemes are

thus much larger than for linear schemes, and include functions with isolated singularities. Bounded variation signals are important examples that provide useful models for images.

In this non-linear setting, Karhunen-Loève bases are not optimal for approximating the realizations of a process  $F$ . It is often easy to find a basis that produces a smaller non-linear error than a Karhunen-Loève basis, but there is yet no procedure for computing the optimal basis that minimizes the average non-linear error.

**Adaptive Basis Choice** Approximations of non-linear signals can be improved by choosing the approximation vectors in families that are much larger than a basis. Music recordings, which include harmonic and transient structures of very different types, are examples of complex signals that are not well approximated by a few vectors chosen from a single basis.

A new degree of freedom is introduced if instead of choosing *a priori* the basis  $\mathcal{B}$ , we adaptively select a “best” basis, depending on the signal  $f$ . This best basis minimizes a cost function related to the non-linear approximation error of  $f$ . A fast dynamical programming algorithm can find the best basis in families of wavelet packet basis or local cosine bases [140]. The selected basis corresponds to a time-frequency tiling that “best” concentrates the signal energy over a few time-frequency atoms.

Orthogonality is often not crucial in the post-processing of signal coefficients. One may thus further enlarge the freedom of choice by approximating the signal  $f$  with  $M$  non-orthogonal vectors  $\{g_{\gamma_m}\}_{0 \leq m < M}$ , chosen from a large and redundant dictionary  $\mathcal{D} = \{g_{\gamma}\}_{\gamma \in \Gamma}$ :

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

Globally optimizing the choice of these  $M$  vectors in  $\mathcal{D}$  can lead to a combinatorial explosion. Chapter 9 introduces sub-optimal pursuit algorithms that reduce the numerical complexity, while constructing efficient approximations [119, 259].

## 1.4.2 Estimation

The estimation of a signal embedded in noise requires taking advantage of any prior information about the signal and the noise. Chapter 10 studies and contrasts several approaches: Bayes versus minimax, linear versus non-linear. Until recently, signal processing estimation was mostly Bayesian and linear. Non-linear smoothing algorithms existed in statistics, but these procedures were often ad-hoc and complex. Two statisticians, Donoho and Johnstone [167], changed the game by proving that a simple thresholding algorithm in an appropriate basis can be a nearly optimal non-linear estimator.

**Linear versus Non-Linear** A signal  $f[n]$  of size  $N$  is contaminated by the addition of a noise. This noise is modeled as the realization of a random process  $W[n]$ , whose

probability distribution is known. The measured data are

$$X[n] = f[n] + W[n] .$$

The signal  $f$  is estimated by transforming the noisy data  $X$  with an operator  $D$ :

$$\tilde{F} = DX .$$

The risk of the estimator  $\tilde{F}$  of  $f$  is the average error, calculated with respect to the probability distribution of the noise  $W$ :

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

It is tempting to restrict oneself to linear operators  $D$ , because of their simplicity. Yet, non-linear operators may yield a much lower risk. To keep the simplicity, we concentrate on diagonal operators in a basis  $\mathcal{B}$ . If the basis  $\mathcal{B}$  gives a sparse signal representation, Donoho and Johnstone [167] prove that a nearly optimal non-linear estimator is obtained with a simple thresholding:

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

The thresholding function  $\rho_T(x)$  sets to zero all coefficients below  $T$ :

$$\rho_T(x) = \begin{cases} 0 & \text{if } |x| < T \\ x & \text{if } |x| \geq T \end{cases} .$$

In a wavelet basis, such a thresholding implements an adaptive smoothing, which averages the data  $X$  with a kernel that depends on the regularity of the underlying signal  $f$ .

**Bayes Versus Minimax** To optimize the estimation operator  $D$ , one must take advantage of any prior information available about the signal  $f$ . In a Bayes framework,  $f$  is considered as a realization of a random vector  $F$ , whose probability distribution  $\pi$  is known a priori. Thomas Bayes was a XVII century philosopher, who first suggested and investigated methods sometimes referred as “inverse probability methods,” which are basic to the study of Bayes estimators. The Bayes risk is the expected risk calculated with respect to the prior probability distribution  $\pi$  of the signal:

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

Optimizing  $D$  among all possible operators yields the *minimum Bayes risk*:

$$r_n(\pi) = \inf_{all D} r(D, \pi) .$$

Complex signals such as images are clearly non-Gaussian, and there is yet no reliable probabilistic model that incorporates the diversity of structures such as edges and textures.

In the 1940's, Wald brought a new perspective on statistics, through a decision theory partly imported from the theory of games. This point of view offers a simpler way to incorporate prior information on complex signals. Signals are modeled as elements of a particular set  $\Theta$ , without specifying their probability distribution in this set. For example, large classes of images belong to the set of signals whose total variation is bounded by a constant. To control the risk for any  $f \in \Theta$ , we compute the maximum risk

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

The *minimax risk* is the lower bound computed over all operators  $D$ :

$$r_n(\Theta) = \inf_{all D} r(D, \Theta).$$

In practice, the goal is to find an operator  $D$  that is simple to implement and which yields a risk close the minimax lower bound.

Unless  $\Theta$  has particular convexity properties, non-linear estimators have a much lower risk than linear estimators. If  $W$  is a white noise and signals in  $\Theta$  have a sparse representation in  $\mathcal{B}$ , then Chapter 10 shows that thresholding estimators are nearly minimax optimal. In particular, the risk of wavelet thresholding estimators is close to the minimax risk for wide classes of piecewise smooth signals, including bounded variation images. Thresholding estimators are extended to more complex problems such as signal restorations and deconvolutions. The performance of a thresholding may also be improved with a best basis search or a pursuit algorithm that adapts the basis  $\mathcal{B}$  to the noisy data. However, more adaptivity does not necessarily means less risk.

### 1.4.3 Compression

Limited storage space and transmission through narrow band-width channels create a need for compressing signals while minimizing their degradation. Transform codes compress signals by decomposing them in an orthonormal basis. Chapter 11 introduces the basic information theory needed to understand these codes and optimize their performance. Bayes and minimax approaches are studied.

A transform code decomposes a signal  $f$  in an orthonormal basis  $\mathcal{B} = \{g_m\}_{0 \leq m < N}$ :

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

The coefficients  $\langle f, g_m \rangle$  are approximated by quantized values  $Q(\langle f, g_m \rangle)$ . A signal  $\tilde{f}$  is restored from these quantized coefficients:

$$\tilde{f} = \sum_{m=0}^{N-1} Q(\langle f, g_m \rangle) g_m.$$

A binary code is used to record the quantized coefficients  $Q(\langle f, g_m \rangle)$  with  $R$  bits. The resulting distortion is

$$d(R, f) = \|f - \tilde{f}\|^2.$$

At the compression rates currently used for images,  $d(R, f)$  has a highly non-linear behavior, which depends on the precision of non-linear approximations of  $f$  from a few vectors in the basis  $\mathcal{B}$ .

To compute the distortion rate over a whole signal class, the Bayes framework models signals as realizations of a random vector  $F$  whose probability distribution  $\pi$  is known. The goal is then to optimize the quantization and the basis  $\mathcal{B}$  in order to minimize the average distortion rate  $d(R, \pi) = \mathbb{E}_\pi\{d(R, F)\}$ . This approach applies particularly well to audio signals, which are relatively well modeled by Gaussian processes.

In the absence of stochastic models for complex signals such as images, the minimax approach computes the maximum distortion by assuming only that the signal belongs to a prior set  $\Theta$ . Chapter 11 describes the implementation of image transform codes in wavelet bases and block cosine bases. The minimax distortion rate is calculated for bounded variation images, and wavelet transform codes are proved to be nearly minimax optimal.

For video compression, one must also take advantage of the similarity of images across time. The most effective algorithms predict each image from a previous one by compensating for the motion, and the error is recorded with a transform code. MPEG video compression standards are described.

## 1.5 TRAVEL GUIDE

### 1.5.1 Reproducible Computational Science

The book covers the whole spectrum from theorems on functions of continuous variables to fast discrete algorithms and their applications. Section 1.3.1 argues that models based on continuous time functions give useful asymptotic results for understanding the behavior of discrete algorithms. Yet, a mathematical analysis alone is often unable to predict fully the behavior and suitability of algorithms for specific signals. Experiments are necessary and such experiments ought in principle be reproducible, just like experiments in other fields of sciences.

In recent years, the idea of reproducible algorithmic results has been championed by Claerbout [127] in exploration geophysics. The goal of exploration seismology is to produce the highest possible quality image of the subsurface. Part of the scientific know-how involved includes appropriate parameter settings that lead to good results on real datasets. The reproducibility of experiments thus requires having the complete software and full source code for inspection, modification and application under varied parameter settings.

Donoho has advocated the reproducibility of algorithms in wavelet signal processing, through the development of a `WAVELAB` toolbox, which is a large library of `MATLAB` routines. He summarizes Claerbout's insight in a slogan: [105]

*An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software environment and the complete set of instructions which generated the figures.*

Following this perspective, all wavelet and time-frequency tools presented in this book are available in `WAVELAB`. The figures can be reproduced as demos and the source code is available. The `LASTWAVE` package offers a similar library of wavelet related algorithms that are programmed in C, with a user-friendly shell interface and graphics. Appendix B explains how to retrieve these toolboxes, and relates their subroutines to the algorithms described in the book.

### 1.5.2 Road Map

Sections are kept as independent as possible, and some redundancy is introduced to avoid imposing a linear progression through the book. The preface describes several possible paths for a graduate signal processing or an applied mathematics course. A partial hierarchy between sections is provided by a level number. If a section has a level number then all sub-sections without number inherit this level, but a higher level number indicates that a subsection is more advanced.

Sections of level <sup>1</sup> introduce central ideas and techniques for wavelet and time-frequency signal processing. These would typically be taught in an introductory course. The first sections of Chapter 7 on wavelet orthonormal bases are examples. Sections of level <sup>2</sup> concern results that are important but which are either more advanced or dedicated to an application. Wavelet packets and local cosine bases in Chapter 8 are of that sort. Applications to estimation and data compression belong to this level, including fundamental results such as Wiener filtering. Sections of level <sup>3</sup> describe advanced results that are at the frontier of research or mathematically more difficult. These sections open the book to research problems.

All theorems are explained in the text and reading the proofs is not necessary to understand the results. Proofs also have a level index specifying their difficulty, as well as their conceptual or technical importance. These levels have been set by trying to answer the question: “Should this proof be taught in an introductory course?” Level <sup>1</sup> means probably, level <sup>2</sup> probably not, level <sup>3</sup> certainly not. Problems at the end of each chapter follow this hierarchy of levels. Direct applications of the course are at the level <sup>1</sup>. Problems at level <sup>2</sup> require more thinking. Problems of level <sup>3</sup> are often at the interface of research and can provide topics for deeper projects.

The book begins with Chapters 2 and 3, which review the Fourier transform properties and elementary discrete signal processing. They provide the necessary background for readers with no signal processing experience. Fundamental properties of local time-frequency transforms are presented in Chapter 4. The wavelet and windowed Fourier transforms are introduced and compared. The measurement of instantaneous frequencies is used to illustrate the limitations of their time-frequency resolution. Wigner-Ville time-frequency distributions give a

global perspective which relates all quadratic time-frequency distributions. Frame theory is explained in Chapter 5. It offers a flexible framework for analyzing the properties of redundant or non-linear adaptive decompositions. Chapter 6 explains the relations between the decay of the wavelet transform amplitude across scales and local signal properties. It studies applications involving the detection of singularities and analysis of multifractals.

The construction of wavelet bases and their relations with filter banks are fundamental results presented in Chapter 7. An overdose of orthonormal bases can strike the reader while studying the construction and properties of wavelet packets and local cosine bases in Chapter 8. It is thus important to read in parallel Chapter 9, which studies the approximation performance of orthogonal bases. The estimation and data compression applications of Chapters 10 and 11 give life to most theoretical and algorithmic results of the book. These chapters offer a practical perspective on the relevance of these linear and non-linear signal processing algorithms.