Chapter 2

Wavelets

Abstract

The purpose of this chapter is to give a short introduction to the wavelets we use later in this work. We describe them in the more general framework of time-frquency analysis, and remind some of their main properties.

2.1 Local frequency

Wavelets have been introduced about 50 years ago to fill a gap between two extreme ways of representing a signal: its time representation which can be seen as an expansion over the Dirac masses:

$$f(t) = \int_{\mathbb{R}} f(u)\delta(t-u)du$$

and its frequency representation or Fourier transform, which is its expansion over the complex exponentials:

$$f(t) = \int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega t}d\omega$$

The first representation gives an information of maximal time resolution: the value $f(t)$ represents a signal intensity at time t. No frequency information is available. A single pointwise value does not provide any information on the frequency content of f. Conversely, the Fourier representation gives an accurate frequency information, but no time information at all. To give a musical analogy, let us suppose that f represents the time curve of a tune. The time representation of f allows to know when we have notes and when we have rests, but not the pitch. On the other hand, the frequency representation tells what pitches can be heard in the tune, but not when exactly they can be heard.

Each of these representations contains the whole information on the signal, because the Fourier transform is a bijective mapping from one representation to the other. However, in each case, only one kind of information is explicitly available.

We can restate the above observation by saying that the Dirac distributions $\delta(.-t)$ have an infinitely high spatial resolution, but no frequency resolution at all, and the converse for the complex exponentials $t \mapsto e^{i\omega t}$. The question was then to find a way of representing signals between these extremes, in which a mixed information is explicitly given, like "there we hear an A, and there a $B^\flat$". Morlet and Gabor suggested to use basis functions that are half way between Dirac masses and complex exponentials, that are both localized in time and in frequency.

A theoretic limit to this is well known: the Heisenberg inequality. Let f be a square integrable function, of L_2 norm equal to 1:

$$\int |f(t)|^2 dt = 1\,.$$

We define the center $c(f)$ and the width $\Delta(f)$ of such a function as

$$c(f) = \int t|f(t)|^2 dt$$
$$\Delta(f) = \sqrt{\int (t - c(f))^2 |f(t)|^2 dt}$$

The Heisenberg inequality is a fundamental inequality that holds for any function f of norm 1:

$$\Delta(f)\Delta(\hat{f}) \geq \frac{1}{2} \tag{H}$$

If we call the frequency width of a function f the width of its Fourier transform $\hat{f}$, this inequality implies that it is not possible to find a function that has a width and a frequency width both arbitrarily small.

We also know the functions that achieve an optimal trade-off: translated and modulated Gaussian functions:

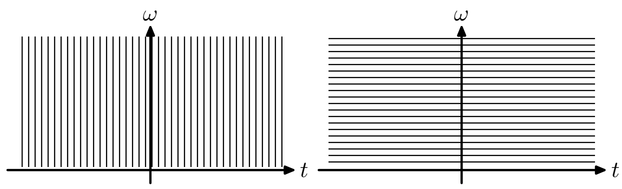
$$Ae^{-(t-t_0)^2/2\Delta t^2} e^{i\omega_0 t}$$

where A is a normalization coefficient such that the above function has an L_2 norm equal to 1. For such functions, and only for them, the Heisenberg inequality is an equality. These functions are now called Gabor wavelets.

2.2 Time-frequency and time-scale representations

To such a function, we associate a time-frequency rectangle on the (t, ω) plane centered around $(c(f), c(\hat{f}))$ and of dimensions $\Delta(f) \times \Delta(\hat{f})$. This rectangle is an intuitive representation of a time-frequency support of a function. To a basis of $L_2(\mathbb{R})$, we also associate a paving of the time-frequency plane, by such rectangles. The rectangles in this paving are centered around the time-frequency center $(c(f), c(\hat{f}))$ of the functions, and their dimensions are proportional to the time frequency dimensions $\Delta(f) \times \Delta(\hat{f})$ in order to draw a partition of the time-frequency plane. This representation is a little arbitrary, as no theoretical result relates a time-frequency paving to the fact that the corresponding function family is a basis.

The time-frequency paving for Dirac functions and complex exponentials are paving with infinitely thin and long rectangle as represented in Fig. 2.1.



(a) DIRAC (b) FOURIER

Figure 2.1: Time-frequency paving for Dirac and Fourier representations

To represent a function as a linear combination of such functions, using all possible combinations of the parameters t_0, ω_0 and Δt is overly redundant. Two different approaches have been suggested:

- a time-frequency approach, that consists in choosing the time width of the functions g independently of the modulation frequency. Such functions can be written as

$$g_{t_0,\omega_0}(t) = e^{i\omega_0 t} g_0(t - t_0)$$

  where $g_0(t) = A_0 e^{-t^2/2\Delta t^2}$. Such a representation is called windowed Fourier transform

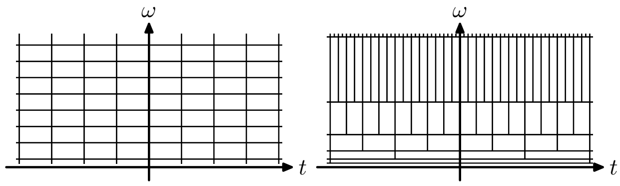
- a time-scale approach, in which the width of the functions is inverse proportional to the frequency (the product $\omega_0 \Delta t$ is a constant equal to $c0$. We then obtain the following parameterized functions:

$$g_{t_0,\Delta t}(t) = \frac{1}{\sqrt{\Delta t}} g_0 \left(\frac{t - t_0}{\Delta t} \right)$$

  where $g_0(t) = A_0 e^{-t^2/2\Delta t^2} e^{ict}$.

The time frequency pavings illustrate the differences between both approaches. In the case of time-frequency analysis, the paving is that of rectangles that are all translates of the same one. In case of time-scale analysis, the rectangles all have the same area, but their *relative* frequency resolution $\Delta\omega/\omega_0$ is constant.

The wavelets we use in this work are time-scale wavelets. The advantages of a time-scale approach are numerous. We dispose of efficient ways to design such family for which the related computation are very fast. The corresponding bases have a very simple form, since all functions have the same shape, and last they have a spatial width which is always proportional to the spatial resolution related to their frequency bandwidth through the Heisenberg inequality.



(a) Time-frequency (b) Time-scale

Figure 2.2: Pavings of the time-frequency plane for the time-frequency representation and the time-scale representation

2.3 Continuous wavelet transform

We choose a basis function ψ called wavelet that fulfills an admissibility condition:

$$C_\psi = \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < +\infty \tag{2.1}$$

We notice that if the Fourier transform of this function is continuous, it has to vanish on 0, which is not the case of the Gabor function above. In practice, we can however consider that the Gabor function satisfies this assumption since for the chosen parameters, the value of its Fourier transform on 0 is very small (but not zero).

We define the continuous wavelet transform $\mathcal{W}f$ of the function f with the following formula:

$$\mathcal{W}f(t,s) = \int_{\mathbb{R}} f(\tau) \frac{1}{\sqrt{s}} \overline{\psi\left(\frac{\tau - t}{s}\right)} d\tau \tag{2.2}$$

The inverse transform is

$$f(t) = \frac{1}{C_\psi} \iint_{\mathbb{R}^2} \mathcal{W}(\tau, s) \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) d\tau \frac{ds}{s^2} \tag{2.3}$$

The same way as for the Fourier transform, this transformation is quasi-isometric:

$$\|f\|^2_{L_2(\mathbb{R})} = \frac{1}{C_\psi} \iint_{\mathbb{R}^2} |\mathcal{W}f(t,s)|^2 dt \frac{ds}{s^2}. \tag{2.4}$$

However, the continuous wavelet transform $\mathcal{W}f$ of a function f is a very redundant representation. $\mathcal{W}f$ fulfills a reproducing kernel equation:

$$\mathcal{W}f(t,s) = \frac{1}{C_\psi} \iint_{\mathbb{R}^2} \mathcal{W}(\tau, \sigma) K(\tau, t, \sigma, s) dt \frac{ds}{s^2}$$

with a kernel K defined as

$$K(\tau, t, \sigma, s) = \int_{\mathbb{R}} \frac{1}{\sqrt{\sigma s}} \psi\left(\frac{t' - t}{s}\right) \overline{\psi\left(\frac{t' - \tau}{\sigma}\right)} dt'$$

2.4 The discrete wavelet transform

Morlet suggested to build bases or frames built according to the following scheme:

$$g_{t_0,\Delta t}(t) = \frac{1}{\sqrt{\Delta t}} g\left(\frac{t - t_0}{\Delta t}\right)$$

where the values of Δt are chosen on a geometric scale, and where the translation steps are proportional to Δt :

$$\Delta t = b^j$$
$$t_0 = k\Delta t$$

A widely used set of scales are dyadic scales 2^j, and we obtain families of functions of the form $g_0(2^j(t - 2^{-j}k)) = g(2^j t - k)$ where j and k are integers. The most common normalization being an L_2 normalization, we obtain families of the form $(\psi_{jk})_{j,k\in\mathbb{Z}}$ where $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$.

In a paper written in 1987, Stéphane MALLAT has drawn a parallel between the time-scale representations of Morlet and the quadrature mirror filters Burt, Adelson and Simoncelli used in image compression.

He described a family of wavelet decompositions that can be implemented very efficiently with a "fast wavelet transform", for which the wavelet ψ is an infinite convolution of discrete filters. More precisely, we can find discrete filters m_0 and m_1:

$$k \mapsto m_0[k] \qquad k \in \mathbb{Z}, \tag{2.5}$$
$$k \mapsto m_1[k] \qquad k \in \mathbb{Z}; \tag{2.6}$$

whose Fourier transforms $\omega \mapsto m_0(\omega)$ and $\omega \mapsto m_1(\omega)$ are 2π–periodic functions. We assume that there exists a scaling function ϕ and a wavelet ψ in $L_2(\mathbb{R})$ such that

$$\hat{\phi}(\omega) = \prod_{k=1}^{+\infty} m_0\left(\frac{\omega}{2^k}\right) \tag{2.7}$$
$$\hat{\psi}(\omega) = m_1\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \tag{2.8}$$

Under some conditions on m_0 and m_1, the family (ψ_{jk}) is an orthonormal basis, and the wavelet decomposition of a sampled signal can be done in a short time with a sequence of filtering and subsampling steps.

This approach reduces significantly the complexity of the object to design. Instead of choosing a function, we choose the discrete (and most usually finite) set of coefficients of the filters.

2.5 Multiresolution analyses

The theory and the tools for the wavelets that can be designed from discrete filters as been widely developed in this last decade, and there exist a number of theorems relating the properties of the discrete filters with those of the resulting wavelets. Moreover, a number of wavelet families have been constructed that carry either the name of some property they have, or the name of their inventor.

2.5.1 Theoretic framework

The theoretic framework developed by Stéphane Mallat is based on the definition of *multiresolution analyses*. A multiresolution analysis is a sequence of closed subspaces of $L_2(\mathbb{R})$ denoted by $(V_j)_{j\in\mathbb{Z}}$, which has the following properties:

$$V_j = \left\{ \sum_{k\in\mathbb{Z}} a_k \phi_{jk} : a_k \in \mathbb{R} \right\} \qquad \text{(Riesz space)} \tag{2.9a}$$

$$V_j \subset V_{j+1} \tag{2.9b}$$

$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\} \tag{2.9c}$$

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L_2(\mathbb{R}) \tag{2.9d}$$

We can make the following remarks:

- The assumption (2.9a) means that V_j is a Riesz space spanned by the family $(\phi_{jk})_{k\in\mathbb{Z}}$. The definition depends on the underlying topology for the functional space, because it can be defined as the closure of the space of finite linear combinations of ϕ_{jk}. This also imposes a constraint on the function ϕ. For an L_2 metric, the mapping

$$\begin{aligned} \ell_2(\mathbb{Z}) &\to L_2(\mathbb{R}) \\ (a_k)_{k\in\mathbb{Z}} &\mapsto \sum_{k\in\mathbb{Z}} a_k \phi_{0k} \end{aligned}$$

  has to be continuous. A function ϕ with a too slow decay at infinity cannot be used.

- Intuitively, we can consider that the set of functions V_{j+1} is a "richer" or "denser" set as V_j, which does not means that the latter is a subset of the former. However, equation (2.9b) assumes this.

  Because of scale and translation invariance of the above definitions, we can check that this last assumption is equivalent to assuming that $\phi \in V_1$, which means that there exists a sequence of coefficients $(m_0[k])_{k\in\mathbb{Z}}$ such that

$$\phi(t) = 2 \sum_{k\in\mathbb{Z}} m_0[k] \phi(2t-k) \tag{2.10}$$

  This is how the discrete filter m_0 appears.

- The assumption (2.9c) is stated more or less for clarity, because it always fulfilled whenever ϕ is not 0. The assumption (2.9d) is also fulfilled whenever $\hat{\phi}$ does vanish on 0.

Wavelets then appear as a natural way to express the difference between two successive spaces V_j and V_{j+1}. We build for this a Riesz space W_0 such that:

$$V_0 \oplus W_0 = V_1 \tag{2.11}$$

The subspace W_0 is spanned by translates of a function ψ:

$$W_0 = \left\{ t \mapsto \sum_{k\in\mathbb{R}} d_k \psi(t-k) : a_k \in \mathbb{R} \right\}$$

This imposes that the function ψ be in the space V_1, thus again:

$$\psi(t) = \sum_{k\in\mathbb{Z}} m_1[k]\phi(2t-k) \tag{2.12}$$

Hence the second discrete filter m_1.

We show that the functions ϕ and ψ are then defined by the sole choice of these two filters: m_0 and m_1. The formulae are

$$\hat{\phi}(\omega) = \prod_{k=1}^{+\infty} m_0\left(\frac{\omega}{2^k}\right) \qquad \text{par itération de (2.10),} \tag{2.13a}$$

$$\hat{\psi}(\omega) = m_1\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \qquad \text{par (2.12).} \tag{2.13b}$$

2.5.2 Wavelet bases

The equation (2.11) can be transposed at any scale j:

$$V_j \oplus W_j = V_{j+1} \tag{2.14}$$

and by iterating, we get:

$$V_j \oplus W_j \oplus \cdots \oplus W_{j'-1} = W_{j'} \qquad \text{if } j < j' \tag{2.15}$$

If j' goes to $+\infty$ (and resp. also j to $-\infty$), we obtain two decompositions:

$$L_2(\mathbb{R}) = \overline{V_j \oplus \bigoplus_{j'=j}^{+\infty} W_{j'}} \qquad \text{for all } j \in \mathbb{Z} \tag{2.16}$$

$$L_2(\mathbb{R}) = \overline{\bigoplus_{j'=-\infty}^{+\infty} W_{j'}} \tag{2.17}$$

The union of all Riesz bases in each of these direct sums thus provides several wavelet bases:

$$\mathcal{B}_j = \{\phi_{jk} : k \in \mathbb{Z}\} \cup \{\psi_{j'k} : j' \geq j,\, k \in \mathbb{Z}\} \tag{2.18}$$

$$\mathcal{B} = \{\psi_{jk} : j \in \mathbb{Z},\, k \in \mathbb{Z}\} \tag{2.19}$$

2.5.3 Wavelet transform

The wavelet transform is used for sampled signals. Most of the time, this sampling is written as

$$f = \sum_{k\in\mathbb{Z}} 2^{j/2} f[k/2^j]\phi_{jk}$$

where the sample $f[k/2^j]$ is estimated with:

$$f[k/2^j] \simeq f(k/2^j)$$

The signal to be transformed is thus supposed to be expanded over the Riesz basis of V_j. The transformation consists in finding its expansion over the basis corresponding to the direct sum:

$$V_L \oplus W_L \oplus W_{L+1} \oplus \cdots \oplus W_{j-1}$$

The transformation is recursive and consists in replacing the representation of a component expanded over some $V_{j'}$ with a representation over $V_{j'-1} \oplus W_{j'-1}$. We successively obtain decompositions adapted to the following direct sums:

$$\begin{gathered} V_{j-1} \oplus W_{j-1} \\ V_{j-2} \oplus W_{j-2} \oplus W_{j-1} \\ \vdots \\ V_L \oplus W_L \oplus W_{L+1} \oplus \cdots \oplus W_{j-1} \end{gathered}$$

2.5.4 Dual filters, dual wavelets

The fundamental step in a wavelet transform is therefore the following basis change:

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

It can be written as a mapping

$$\begin{aligned} \ell_2(\mathbb{Z}) &\to \ell_2(\mathbb{Z}) \times \ell_2(\mathbb{Z}) \\ (a_{j+1,k})_{k\in\mathbb{Z}} &\mapsto [(a_{jk})_{k\in\mathbb{Z}}, (d_{jk})_{k\in\mathbb{Z}}] \end{aligned}$$

If we denote by A_j and D_j the 2π–periodic functions whose coefficients are the discrete sequences $k \mapsto a_{jk}$ and $k \mapsto d_{jk}$:

$$\begin{aligned} A_j(\omega) &= \sum_{k\in\mathbb{Z}} a_{jk} e^{-ik\omega} \\ D_j(\omega) &= \sum_{k\in\mathbb{Z}} d_{jk} e^{-ik\omega} \end{aligned}$$

this fundamental step can be written as a multiplication with a transfer matrix:

$$\begin{bmatrix} A_j(2\omega) \\ D_j(2\omega) \end{bmatrix} = \begin{bmatrix} m_0(\omega) & m_0(\omega+\pi) \\ m_1(\omega) & m_1(\omega+\pi) \end{bmatrix} \begin{bmatrix} A_{j+1}(\omega) \\ A_{j+1}(\omega+\pi) \end{bmatrix}$$

A necessary condition for such a transformation to be invertible in $L_2([0,2\pi])^2$ is that the transfer matrix

$$T(\omega) = \begin{bmatrix} m_0(\omega) & m_0(\omega+\pi) \\ m_1(\omega) & m_1(\omega+\pi) \end{bmatrix}$$

be bounded and of bounded inverse uniformly on $[0,2\pi]$. In this case, the dual transfer matrix is $\tilde{T}(\omega) = T(\omega)^{-T}$. There exist two other 2π–periodic functions $\tilde{m}_0$ and $\tilde{m}_1$ such that $\tilde{T}(\omega)$ is written:

$$\tilde{T}(\omega) = \begin{bmatrix} \tilde{m}_0(\omega) & \tilde{m}_0(\omega+\pi) \\ \tilde{m}_1(\omega) & \tilde{m}_1(\omega+\pi) \end{bmatrix}$$

These filters define dual wavelets $\tilde{\phi}$ and $\tilde{\psi}$ with equations similar to (2.13a) and (2.13b):

$$\hat{\tilde{\phi}}(\omega) = \prod_{k=1}^{+\infty} \tilde{m}_0\left(\frac{\omega}{2^k}\right) \tag{2.20a}$$

$$\hat{\tilde{\psi}}(\omega) = \tilde{m}_1\left(\frac{\omega}{2}\right)\hat{\tilde{\phi}}\left(\frac{\omega}{2}\right) \tag{2.20b}$$

The functions $\tilde{\phi}$ and $\tilde{\psi}$ are dual wavelets in the sense that for all j, we have the expansions on $L_2(\mathbb{R})$:

$$f = \sum_{k\in\mathbb{Z}} \left\langle f, \tilde{\phi}_{jk} \right\rangle \phi_{jk} + \sum_{j'\geq j, k\in\mathbb{Z}} \left\langle f, \tilde{\psi}_{j'k} \right\rangle \psi_{j'k} \tag{2.21}$$

for all $j \in \mathbb{Z}$ and $f \in L_2(\mathbb{Z})$, and with j going to $-\infty$, we obtain:

$$f = \sum_{j,k\in\mathbb{Z}} \left\langle f, \tilde{\psi}_{jk} \right\rangle \psi_{jk} \tag{2.22}$$

2.5.5 The fast wavelet transform

The coefficients of the filters m_0 and m_1, and of the dual filters $\tilde{m}_0$ and $\tilde{m}_1$ are used in the basis change

$$\{\phi_{jk} : k \in \mathbb{Z}\} \cup \{\psi_{jk} : k \in \mathbb{Z}\} \leftrightarrow \{\phi_{j+1,k} : k \in \mathbb{Z}\}$$

with the following formulae:

- for the forward wavelet transform, we have

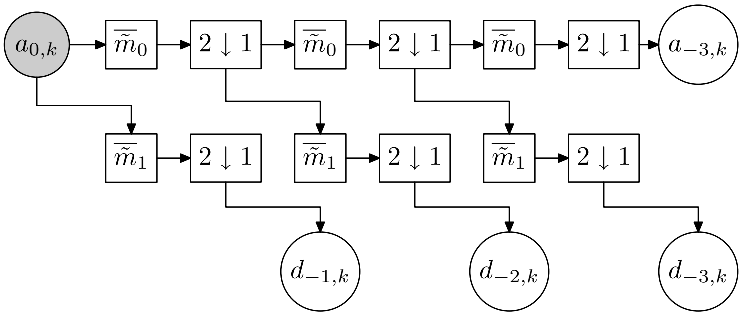
$$a_{jk} = 2\sum_{l\in\mathbb{Z}} \tilde{m}_0[k] a_{j+1,2l-k}$$

$$d_{jk} = 2\sum_{l\in\mathbb{Z}} \tilde{m}_1[k] a_{j+1,2l-k}$$

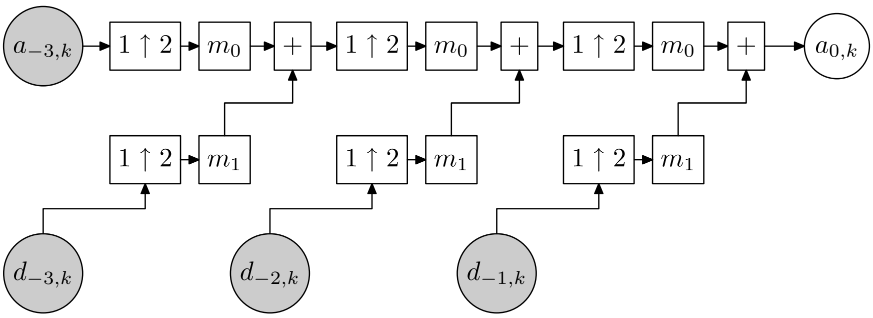
- and for the inverse wavelet transform, we get

$$a_{j+1,k} = \frac{1}{2} \sum_{l \in \mathbb{Z}} m_0[2l-k] a_{jl} + m_1[2l-k] d_{jl}$$

The cascade filtering schemes are displayed for the direct and the inverse transforms in Fig. 2.3 between resolutions $j = 0$ and $j = -3$.



(a) Direct transform



(b) Inverse transform

Figure 2.3: Fast wavelet transform. The shaded circle represent the input, while the white circle represent the output.

For a finite number N of samples, a wavelet transform (until any resolution possible with such a number of samples) costs less that $A \times N$ flops, where the constant A depends on the chosen filters. This in theory better than a fast Fourier transform which costs $N \log N$ flops.

2.5.6 Orthogonal wavelets

Orthogonal wavelets are wavelets ψ such that the family $(t \mapsto 2^{j/2}\psi(2^j t - k))_{j,k\in\mathbb{Z}}$ is an orthonormal basis of $L_2(\mathbb{R})$. This means that $\phi = \tilde{\phi}$ and $\psi = \tilde{\psi}$, and is equivalent to saying that the transfer matrix and the dual transfer matrix are equal, or that the transfer matrix is unitary for all ω. This can be written in terms of the filters m_0 and m_1 as:

$$\begin{array}{ccccc} |m_0(\omega)|^2 & + & |m_0(\omega+\pi)|^2 & = & 1 & \forall\omega \\ m_0(\omega)\overline{m_1(\omega)} & + & m_0(\omega+\pi)\overline{m_1(\omega+\pi)} & = & 0 & \forall\omega \\ |m_1(\omega)|^2 & + & |m_1(\omega+\pi)|^2 & = & 1 & \forall\omega \end{array}$$

In this case, m_0 and m_1 are called quadrature mirror filters, as Esteban and Galand called them, and after them Adelson and Simoncelli. Moreover, all the direct sums above are then orthogonal direct sums. In practice, once the filter m_0 is chosen, almost a single choice for m_1 is possible, which is

$$m_1(\omega) = e^{i\omega}\overline{m_0(\omega+\pi)}$$

Historically, the first wavelets to be defined in the multiresolution analysis framework were orthogonal wavelets (Meyer, Mallat), so that other wavelets were afterwards called biorthogonal wavelets. The prefix “bi” meaning that we need two wavelet bases instead of one. The first for the decomposition, and the second for the signal reconstruction. A systematic analysis of biorthogonal wavelets has been done by Cohen, Daubechies *et al.*

There exist a number of orthogonal wavelets that are frequently used. The most known such wavelets are Daubechies’ wavelets. They achieve an optimal trade-off between support size and number of vanishing moments (two opposite constraints, in some way comparable to the Heisenberg inequality). There exist also many other wavelets, like the *coiflets* named after Ronald Coifman, or the *symmlets*, that are almost symmetric.

Orthogonal wavelet bases have a considerable theoretic advantage in compression or denoising applications : first the error metric is in general an L_2 metric, and this can be easily expressed in terms of orthogonal wavelet coefficients. In the case of denoising, a Gaussian white noise has expansion coefficients over an orthonormal basis that are i.i.d Gaussian variables.

In practice however, orthogonal wavelets do not have the same flexibility as that offered by the biorthogonal wavelets. One can for example show that they can never (up to some trivial cases) be symmetric. The dual relationship between dual wavelets for computing expansion coefficients $c_{jk} = \langle f, \tilde{\psi}_{jk}\rangle$ and wavelets ψ_{jk} for reconstruction $f = \sum c_{jk}\psi_{jk}$ can be reversed, to get:

$$f = \sum_{j,k\in\mathbb{Z}} \langle \psi_{jk}, f\rangle \tilde{\psi}_{jk}$$

that can be compared to (2.22).

2.6 Coefficient decay, smoothness and approximation rates

If we want subexpansions of a function to converge rapidly to the original function, the expansion coefficients have to decay rapidly as $j \to +\infty$. This decay is connected to the number of *vanishing moments* of the dual wavelet and to the smoothness of the function.

We say that $\tilde{\psi}$ has p vanishing moments if

$$\int_{\mathbb{R}} \tilde{\phi}(t) t^k dt = 0$$

for all k in $\{0, \dots, p-1\}$. This is the same as assuming that the Fourier transform of $\tilde{\psi}$ has a zero of order p in $\omega = 0$, or as assuming that ψ is orthogonal to any polynomial of degree less that p.

We can show that if f is p times differentiable with a bounded derivative of order p over an interval I, its wavelet coefficients decay in $2^{-j(p+1/2)}$ on I, i.e. there exists a bound M such that

$$\left| \langle \tilde{\psi}_{jk}, f \rangle \right| \leq M 2^{-(p+1/2)j} \qquad \text{if } \operatorname{supp} \tilde{\psi}_{jk} \subset I$$

To show this bound, we write a Taylor expansion of f around the center of the wavelet $\tilde{\psi}_{jk}$. If $u = k/2^j$, we have:

$$f(t) = \sum_{k=0}^{p-1} \frac{(t-u)^p}{p!} f^{(p)}(u) + (t-u)^p r(t)$$

where the function $r(t)$ is bounded by the p^{th} derivative of f. By inner product of f with $\tilde{\psi}_{jk}$, the sum of polynomial terms vanishes, and we have the following term left

$$\begin{aligned} \langle \tilde{\psi}_{jk}, f \rangle &= \int (t-u)^p r(t) \overline{\psi_{jk}(t)} dt \\ &= \int t^p r(t+u) \overline{\psi_{j0}(t)} dt \end{aligned}$$

which can be bounded with a simple variable change by $M 2^{-(p+1/2)j}$. We thus see that the local regularity of the function is strongly connected to the wavelet coefficient decay. A quasi converse statement is also true: if the coefficient decay is $\langle f, \tilde{\psi}_j \rangle \leq M 2^{-(p+1/2)j}$ and if ψ is p–Lipschitz, then f is r–Lipschitz for any $r < p$.

To illustrate the relationship between smoothness and coefficient decay, an example of wavelet coefficient expansion is displayed in Fig. 2.4.

Remark

There is a strong connection between two dual parameters of a wavelet basis: (1) the number of vanishing moments of the dual wavelet, and (2) the smoothness of the wavelet. These parameters are involved respectively in the theorem estimating the wavelet coefficient decay from the function smoothness, and the converse theorem.

Yves Meyer has shown that if a wavelet is p times derivable, then the dual wavelet has at least $p+1$ vanishing moments. The converse is not true: the smoothness index of a wavelet is usually much lower than the number of vanishing moments of the dual wavelet.

For Daubechies wavelets, Cohen and Conze showed that the Lipschitz smoothness index α_N of a wavelet with N vanishing moments is asymptotically equal to:

$$\alpha \sim \left(1 - \frac{\log 3}{2 \log 2}\right) N \simeq 0,2075\, N$$

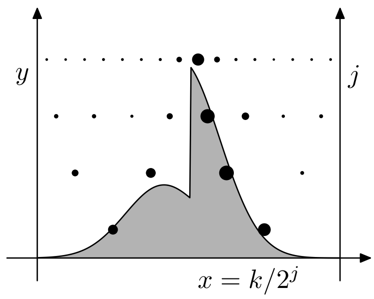


Figure 2.4: Wavelet coefficients of compactly supported function. Note that large coefficients at large j concentrate around singularities.

2.7 Design of smooth wavelets and of the corresponding filters

Designing a wavelet family thus consists in choosing a pair of discrete filters (m_0, m_1). The corresponding transfer matrix

$$T(\omega) = \begin{bmatrix} m_0(\omega) & m_0(\omega+\pi) \\ m_1(\omega) & m_1(\omega+\pi) \end{bmatrix}$$

has to fulfill a perfect reconstruction condition, i.e. it must be bounded and of bounded inverse, which is relatively easy to check for compactly supported filters.

This is however not sufficient. Indeed, for some filters m_0 the associated function ϕ may not be in $L_2(\mathbb{R})$. If we simply take $m_0(\omega) = 1$, the corresponding scale function is a Dirac mass: $\delta(\omega)$. The generated functions also have to be smooth enough. In addition we must also require that the partial products

$$1_{[-2^j\pi, 2^j\pi]}(\omega) \prod_{k=1}^{j} m_0\left(\frac{\omega}{2^k}\right)$$

converge to $\hat{\phi}$ fast enough (at least in L_2 norm), so that the perfect reconstruction property of the filters can be effectively transposed into the biorthogonality of the wavelets:

$$\langle \phi_{jk}, \tilde{\phi}_{jk'} \rangle = \delta_{kk'}$$

This assumption is really necessary, as can be shown *a contrario* with the following example. The filter $m_0(\omega) = (1 + e^{2i\omega})/2$ is orthogonal, and the resulting scaling function is in $L_2(\mathbb{R})$

$$\phi(t) = \frac{1}{2} 1_{[0,2]}(t)$$

which is obviously not an orthogonal scaling function, since

$$\langle \phi_{jk}, \tilde{\phi}_{jk'} \rangle \neq \delta_{kk'} \qquad \text{for } |k - k'| = 1$$

To sum up, the three basic ingredients of an unconditional wavelet basis of L_2 built from filters are:

1. a pair of filters (m_0, m_1) fulfilling the perfect reconstruction assumption, and of dual filters $(\tilde{m}_0, \tilde{m}_1)$;
2. ϕ defined as
$$\hat{\phi}(\omega) = \prod_{k=1}^{+\infty} m_0\left(\frac{\omega}{2^k}\right)$$
should be in $L_2(\mathbb{R})$, as well as $\tilde{\phi}$ defined in a similar way with $\tilde{m}_0$;
3. the partial convolution products truncated on the frequency band $[-2^j\pi, 2^j\pi]$
$$1_{[-2^j\pi, 2^j\pi]}(\omega) \prod_{k=1}^{j} m_0\left(\frac{\omega}{2^k}\right)$$
should converge to ϕ (and the same for $\tilde{\phi}$) in $L_2(\mathbb{R})$ norm, so that the filter duality is transposed into wavelet duality.

2.7.1 Sufficient conditions in the orthonormal case

A sufficient condition has been given by Mallat in his paper of 1987, for compactly supported orthogonal wavelets. He showed that an orthogonal filter m_0 (with $m_0(0) = 1$ and $|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1 \forall\omega$) generates an orthonormal basis of $L_2(\mathbb{R})$ if

$$m_0(\omega) \neq 0 \qquad \text{for } \omega \in [-\pi/2, \pi/2]$$

This assumption has been relaxed afterwards, by imposing only that ϕ does not vanish on $[-\pi/3, \pi/3]$.

2.7.2 Sufficient condition for regularity

A theorem of Daubechies and Tchamitchian relates the smoothness of the scaling function ϕ to a factorization of the filter m_0.

Theorem 2.1 (Daubechies, Tchamitchian)

Let m_0 be a discrete filter such that the function

$$\omega \mapsto m_0(\omega)$$

is bounded, differentiable on 0 and such that $m_0(0) = 1$, and let ϕ be the function (tempered distribution) defined by

$$\hat{\phi}(\omega) = \prod_{k=1}^{+\infty} m_0\left(\frac{\omega}{2^k}\right)$$

We factor m_0 *as*

$$m_0(\omega) = \left(\frac{e^{i\omega}+1}{2}\right)^N r(\omega)$$

where N *is as large as possible, and* r *is a bounded. We set*

$$B_j = \sup_\omega \left|\prod_{k=1}^{j} r(2^k\omega)\right|$$

and

$$b_j = \frac{\log B_j}{j \log 2}$$

$$b = \inf_{j>0} b_j \qquad \text{(critical exponent)}$$

Then the Fourier transform of ϕ *is bounded by the following formula:*

$$|\hat{\phi}(\omega)| \leq \frac{M}{1+|\omega|^{N-b_j}} \qquad \forall j > 0$$

If we denote by G^α *the Hölder space of exponent* α*, i.e. the set of functions* f *such that*

$$\int |\hat{f}(\omega)|(1+|\omega|)^\alpha d\omega < +\infty$$

ϕ *is in* H^α *for any* $\alpha < N - b_j - 1$*.*

Moreover, for any $\alpha < N - b_j - 1/2$*, the function* ϕ *is in the Sobolev space of exponent* α*.*

This an ingredient of the following theorem by Cohen

Theorem 2.2 (Cohen)

Let (m_0, m_1) *and* $(\tilde{m}_0, \tilde{m}_1)$ *be two dual pairs of filters (thus perfect reconstruction filters). We denote by* N*,* $\tilde{N}$*,* b *and* $\tilde{b}$ *the factorization exponents and the critical exponents of* m_0 *and* $\tilde{m}_0$*. If* $N - b > 1/2$ *and* $\tilde{N} - \tilde{b} > 1/2$*, then the wavelets generated by these filters make up unconditional biorthogonal bases of* $L_2(\mathbb{R})$*.*

Bibliography

[BCD+91] G. Beylkin, R. Coifman, I. Daubechies, S. G. Mallat, Y. Meyer, L. Raphael and M. B. Ruskai. *Wavelets*. Jones and Bartlett, Boston, 1991.

[Dau92] I. Daubechies. *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[Mal89a] S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315(1):69–87, 1989.

[Mal89b] S. G. Mallat. A theory for multiscale signal decomposition: The wavelet representation. *IEEE Trans. on Pattern and Machine Intelligence*, 11(7):674–693, 1989.

[Mal97] S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1997.