Chapter 3

Wavelets and applications

3.1 Multiresolution analysis

3.1.1 The limits of Fourier analysis

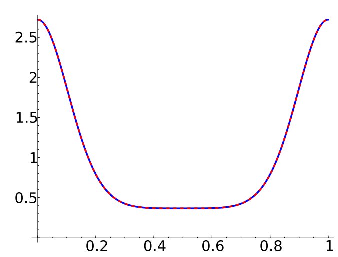
Although along this chapter the underlying Hilbert space will be $L^2(\mathbb{R})$, we start with a completely explicit example with Fourier series to illustrate the situation. We consider the 1-periodic functions

$$(3.1) \qquad f_1(x) = e^{\cos(2\pi x)} \cos\big(\sin(2\pi x)\big) \qquad \text{and} \qquad f_2(x) = f_1(x) + \Big(\frac{\sin(50\pi x)}{50\cos(\pi x)}\Big)^2.$$

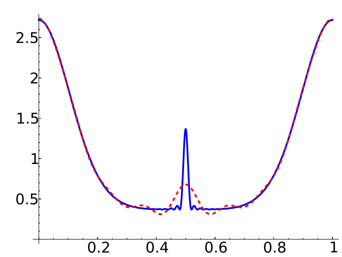
Both functions differ in a scaled and displaced version of Fejér kernel (1.51) and in this way f_2 in $[0,1]$ is like f_1 with a peak of height 1 and width approximately $1/25$ at $x = 1/2$. These $C^\infty(\mathbb{T})$ functions equal their Fourier expansion that admit the closed form

$$(3.2) \qquad f_1(x) = \sum_{n=0}^{\infty} \frac{\cos(2\pi n x)}{n!} \qquad \text{and} \qquad f_2(x) = -\frac{1}{50} + \sum_{n=0}^{\infty} \Big(\frac{1}{n!} + \frac{(-1)^n}{25}\Big(1 - \frac{n}{50}\Big)_+\Big)\cos(2\pi n x),$$

where the index $+$ indicates the positive part. The following figures contain the plots of the functions f_1 and f_2 and the approximation (the dashed line) truncating these Fourier series up to $n = 8$.



f_1 and $S_8 f_1$



f_2 and $S_8 f_2$

The peak occurring in a short amount of time (or space, if you prefer) causes a small change in 50 Fourier coefficients and we have to enlarge a lot the range of frequencies to get a good approximation.

The underlying idea, as we saw, is the uncertainty principle. It is impossible to see fine details with a limited range of frequencies. How to fight against uncertainty principle? You cannot defeat theorems but one can play the usual game in Mathematics: If you do not like the conclusions, circumvent them changing the hypotheses. Imagine that we change the usual complete orthonormal system $\{e(nx)\}_{n\in\mathbb{Z}}$ in $L^2(\mathbb{T})$ by another containing a multiple of $f_2 - f_1$, then to represent the peak we only need to spend a Fourier coefficient.

Let us move to $L^2(\mathbb{R})$ to address the problem in successive steps. Our intuition from Fourier analysis and (1.101) tell us that, due to the lack of compactness of $\mathbb{R}$, we deal here with different animals, Fourier integrals instead of Fourier series, but the epitome of wavelets that we will study next section, will recover the series for $L^2(\mathbb{R})$.

A first natural idea is to introduce *windows*. Have in mind a window as an approximation of a compactly supported function, something that lives mostly in an interval, the one we are looking at. For us a window is a real function $w : \mathbb{R} \longrightarrow \mathbb{R}$ with $\|w\|_2 = 1$ that we assume to be as regular as we wish, for instance rapidly decreasing, because this is a temporary approach. If we are interested in representing details of a certain width we should choose w having the most of its mass in an interval of the same size. If we "localize" the harmonics $e(\xi x)$ of Fourier analysis with $w(x)e(\xi x)$ we will be able to reproduce functions living in the approximate support of the window, whatever it means. To analyze any function we must move the window along $\mathbb{R}$. It suggests to introduce the *windowed Fourier transform*, also called *short-time Fourier transform* when w is compactly supported,

$$G_w f(\xi, b) = \int_{-\infty}^{\infty} f(x)\overline{w}_{\xi,b}(x)\, dx \qquad \text{with} \quad w_{\xi,b}(x) = w(x-b)e(\xi x). \tag{3.3}$$

There exists a fair enough inversion formula and also a Parseval identity. We state them here under overkilling regularity although there is an L^2 version [Bré02, Th.D.1.2].

Proposition 3.1.1. *Let w be as before. For any rapidly decreasing f we have*

$$f(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_w f(\xi, b) w_{\xi,b}(x)\, d\xi db \tag{3.4}$$

and

$$\|f\|_2^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big|G_w f(\xi, b)\big|^2\, d\xi db. \tag{3.5}$$

Proof. By Parseval identity (1.74) and $\widehat{w}_{\xi,b}(t) = e\big((\xi - t)b\big)\widehat{w}(t-\xi)$, we have

$$G_w f(\xi, b) = \int_{-\infty}^{\infty} \widehat{f}(t)e(tb)\widehat{w}(\xi - t)e(-\xi b)\, dt. \tag{3.6}$$

Note that the conjugate of $\widehat{w}(u)$ is $\widehat{w}(-u)$ because w is real. Changing the order of integration and using the inversion formula for the Fourier transform,

$$\int_{-\infty}^{\infty} G_w f(\xi, b)e(\xi x)\, d\xi = \int_{-\infty}^{\infty} \widehat{f}(t)e(tb)e\big(t(x-b)\big)\widehat{w}(x-b)\, dt = w(x-b)f(x). \tag{3.7}$$

Then the right hand side of the first formula of the statement is $f(x)\|w\|_2^2 = f(x)$.

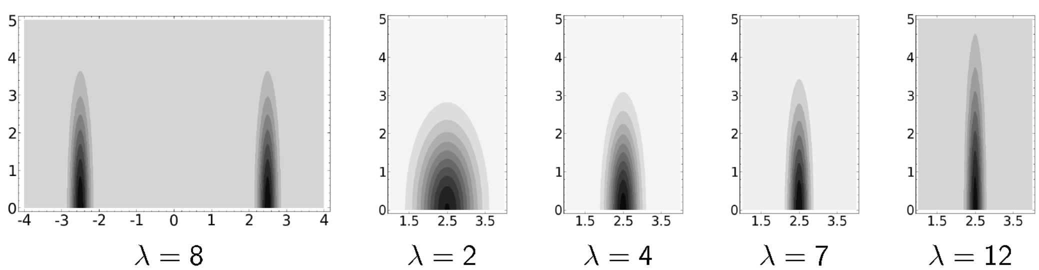
On the other hand, applying Parseval identity (1.74) to (3.7),

$$\int_{-\infty}^{\infty} \left|G_w f(\xi,b)\right|^2 d\xi = \int_{-\infty}^{\infty} \left|w(x-b)f(x)\right|^2 dx \tag{3.8}$$

and integrating on b, this is $\|w\|_2^2\|f\|_2^2 = \|f\|_2^2$. □

The definition (3.3) and the equality (3.6) used in the proof indicate that w cuts the function and $\widehat{w}$ cuts its Fourier transform. According to Theorem 1.2.9, in a certain vague sense the "best" choice is $w(x) = \pi^{-1/4}\lambda^{1/2}e^{-\lambda^2x^2/2}$ with λ^{-1} somewhat the size of the details we want to observe. In this case $G_w f(\xi,b)$ is called the *Gabor transform*.

As an illustration of how the windowed Fourier transform works, look the following contour plots (darker means larger values) of the Gabor transform $|G_w f(\xi,b)|$ where f is the function $e^{-32(x+5/2)^2} + e^{-32(x-5/2)^2}$ representing two peaks. The horizontal axis is b and the vertical axis is ξ. The complete plot is symmetric with respect to both axes.



As expected, the most of the mass is concentrated around $b = \pm 5/2$. The width of the peaks and of w coincide for $\lambda = 8$. For higher λ the latter is smaller and we have to pay with extra higher frequencies (if you use a too short measuring-tape you will have to use it many times). On the other hand, for λ much smaller the situation gets closer to usual Fourier analysis and for λ very close to zero we would obtain something approaching to a horizontal band i.e., not depending on b. Roughly speaking, the blobs have less area for λ close to 8, meaning that we have to consider "less values" of b and ξ to get a good approximation of the function.

The obvious shortcoming of the windowed Fourier transform is that it imposes a fixed size for the details we can analyze efficiently. The success of wavelets in practice relies on the capability to manage infinitely many windows at the same time that operate at different scales. A *wavelet* is essentially a fixed profile to be translated and scaled to make the windows.

There are three important avatars of wavelets. The first one is related to the windowed Fourier transform and we will define it right now. The second, mathematically more challenging, is related to the construction of orthonormal systems spanning $L^2(\mathbb{R})$. The third one is a discrete version, very simple from the mathematical point of view but the most useful in practice.

Following [Pin02] we call *continuum wavelet* to any $\psi \in L^2(\mathbb{R}) - \{0\}$ satisfying

$$\int_{-\infty}^{\infty} |\xi|^{-1} |\widehat{\psi}(\xi)|^2 \, d\xi < \infty. \tag{3.9}$$

We say that ψ is *normalized* if this integral equals 1. We can always normalize a continuum wavelet multiplying it by a positive real constant.

Given a certain wavelet ψ we define the *wavelet transform* associated to it as the operator that applies $f \in L^2(\mathbb{R})$ into

$$W_\psi f(a,b) = \int_{-\infty}^{\infty} f(x)\overline{\psi}_{a,b}(x) \, dx \qquad \text{where} \quad \psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi\Big(\frac{x-b}{a}\Big), \ a,b \in \mathbb{R}, \quad a \neq 0. \tag{3.10}$$

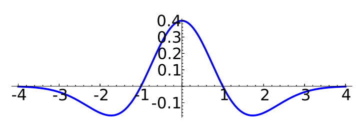
Here $\psi_{a,b}$ plays the role of a variable window where b gives the position and a the scale. Note that $\|\psi_{a,b}\|_2 = \|\psi\|_2$ then Cauchy-Schwarz inequality assures that for each $f \in L^2(\mathbb{R})$ its wavelet transform is bounded.

We expect some oscillation in ψ because we are mimicking $w_{\xi,b}$ rather than w. The condition (3.9) requires, for $\widehat{\psi}$ continuous, $\widehat{\psi}(0) = \int_{-\infty}^{\infty} \psi = 0$ and then involves a minimal oscillation. In general, the vanishing of moments $\int_{-\infty}^{\infty} x^j \psi(x) \, dx$ until certain n implies a better behavior of the wavelet transform. We do not expand this idea here. It will reappear later.

Of course, something with zero average and a minimal regularity and decay qualifies to be a continuum wavelet. Let us review three celebrated examples.

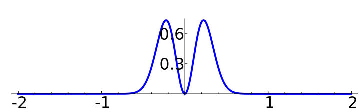
A normalized continuum wavelet is the so-called *Mexican hat wavelet*, with a self-explanatory name,

$$\psi(x) = \frac{1}{\sqrt{2\pi}}(1-x^2)e^{-x^2/2} \tag{3.11}$$



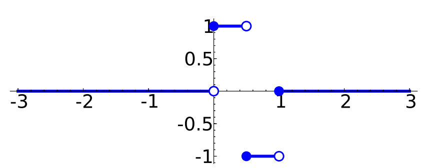
Its Fourier transform vanishes with order 2 at the origin and shows a quicker exponential decay:

$$\widehat{\psi}(\xi) = (2\pi\xi)^2 e^{-2\pi^2\xi^2} \tag{3.12}$$



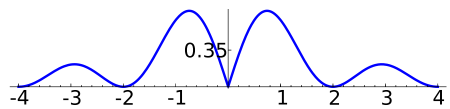
The next two wavelets receive proper names and appear in any academic textbook because they provide examples with rather explicit calculations in the next section. Actually, they are not very practical. The first one is the *Haar wavelet* and it was introduced at the beginning of the 20th century [Haa10] to address a problem that would be central in the development of wavelets. It is a humble horizontal broken line:

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2, \\ -1 & \text{if } 1/2 \le x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.13}$$



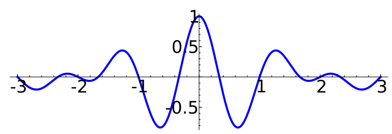
Its Fourier transform vanishes with order 1 at $\xi = 0$ and decays as $|\xi|^{-1}$, then (3.9) is assured. We plot $|\widehat{\psi}|$ because $\widehat{\psi}$ is complex.

$$\widehat{\psi}(\xi) = \frac{\big(1 - e(-\xi/2)\big)^2}{2\pi i \xi}, \qquad |\widehat{\psi}| \tag{3.14}$$



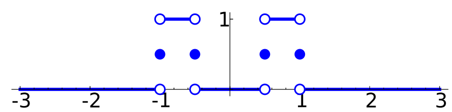
The *Shannon wavelet* involves the function sinc introduced in (1.60):

$$\psi(x) = \operatorname{sinc}\big(\frac{x}{2}\big)\cos\big(\frac{3\pi x}{2}\big) \tag{3.15}$$



Its Fourier transform is almost as simple as the Haar wavelet:

$$\widehat{\psi}(\xi) = \chi^*_{[-1,-1/2]}(\xi) + \chi^*_{[1/2,1]}(\xi) \tag{3.16}$$



where $\chi^*_{[a,b]}$ means the characteristic function of the interval $[a,b]$ putting $1/2$ as the value at the extremes.

For general normalized continuum wavelets, one can deduce an inversion formula and a Parseval identity formally similar to Proposition 3.1.1. Due to the low regularity assumed here the proof and even the statement requires finer considerations (as it happens with the Fourier transform in L^2 [DM72]). We state the result in this general context but we assume some regularity for the proof (see [Pin02, §6.2] or [Bré02, D2·2] for a proof without this assumption).

Proposition 3.1.2. *Let ψ be a normalized continuum wavelet. Then for every $f \in L^2(\mathbb{R})$*

$$f(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_\psi f(a,b)\psi_{a,b}(x)\,\frac{da\,db}{a^2}. \tag{3.17}$$

where the integrals are understood as principal values[1] *at $a = 0$ and $a, b = \infty$. Moreover*

$$\int_{-\infty}^{\infty} |f|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big|W_\psi f(a,b)\big|^2\,\frac{da\,db}{a^2}. \tag{3.18}$$

Proof. Let us assume $\psi \in L^1$ (in this way $\widehat{\psi} \in L^\infty \cap C$) and $f, \widehat{f} \in L^1 \cap C^1$ to have the Fourier inversion formula at every point thanks to Theorem 1.2.4.

The Fourier transform of $\psi_{a,b}$ is $|a|^{1/2} e(-b\xi)\widehat{\psi}(a\xi)$ by (1.71) and (1.72). Then Parseval identity for the Fourier transform (1.74) implies

$$W_\psi f(a,b) = |a|^{1/2}\int_{-\infty}^{\infty} \widehat{f}(\xi)\overline{\widehat{\psi}}(a\xi)e(b\xi)\,d\xi. \tag{3.19}$$

[1] This means $\int_{|b|<M_1}\int_{\epsilon<|a|<M_2}$ with $\epsilon \to 0$ and $M_1, M_2 \to \infty$ and it converges in L^2 to the function f.

This means that $|a|^{1/2}\widehat{f}(\xi)\overline{\widehat{\psi}}(a\xi)$ is the Fourier transform of $W_\psi f(a,\cdot)$. On the other hand the Fourier transform of $\overline{\psi}_{a,\cdot}(x)$ is $|a|^{1/2}e(-x\xi)\overline{\widehat{\psi}}(a\xi)$. Then again by Parseval identity,

(3.20)
$$\int_{-\infty}^{\infty} W_\psi f(a,b)\psi_{a,b}(x)\,db = \int_{-\infty}^{\infty} W_\psi f(a,b)\overline{\overline{\psi}_{a,b}(x)}\,db = |a|\int_{-\infty}^{\infty} \widehat{f}(\xi)|\widehat{\psi}(a\xi)|^2 e(x\xi)\,d\xi.$$

Integrating against a^{-2} and making the change $a \mapsto a/\xi$ this is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widehat{f}(\xi)|a|^{-1}|\widehat{\psi}(a\xi)|^2 e(x\xi)\,d\xi = \int_{-\infty}^{\infty} \widehat{f}(\xi)e(x\xi)\,d\xi \int_{-\infty}^{\infty} |a|^{-1}|\widehat{\psi}(a)|^2\,da. \tag{3.21}$$

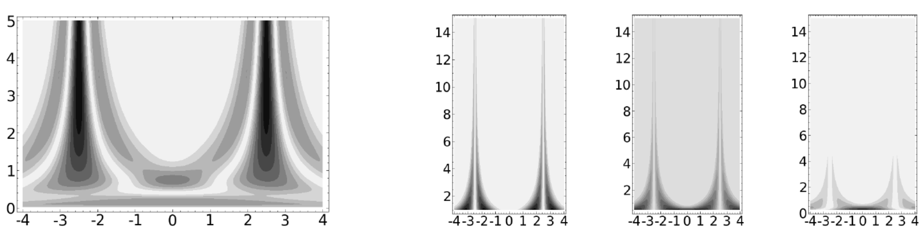
The first integral factor is $f(x)$ by the inversion formula and the second is 1 by the normalization condition. Then we have proved (3.17).

By (3.19) and the standard Parseval identity,

$$\int_{-\infty}^{\infty} |W_\psi f(a,b)|^2\,db = |a|\int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2|\widehat{\psi}(a\xi)|^2\,d\xi. \tag{3.22}$$

Integrating against a^{-2} as before, we obtain (3.18). □

Let us illustrate the wavelet transform with the example we studied before. In some sense in a wavelet the inverse of the scale $A = 1/a$ corresponds to the frequency. The first figure below is the plot of $|W_\psi f(1/A,b)|$ for the Mexican hat wavelet and the function $e^{-32(x+5/2)^2} + e^{-32(x-5/2)^2}$ that we analyzed with the Gabor transform. To keep the analogy, A is in the vertical axis and in the same range as before.



$|W_\psi f(1/A,b)|$ $\quad A > 1 \quad A > 0.5 \quad A > 0.25$

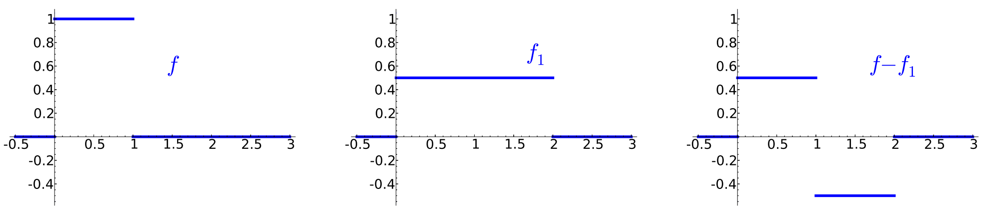
The other plots corresponds to $|W_\psi f(1/A,b)|$ for the characteristic function of the interval $[-1,1]$ in a larger range. They show that the large values corresponding to A close to zero mask the rest of the values when distributing the contour levels. This gives an idea about the variation of these levels.

The conclusion to extract from the asymptotes is that $\psi_{a,b}$ with small scale a (large "frequency" $1/A$) only contribute when there are fine details to study. One may argue that apparently we got much better results with the Gabor transform because now the area occupied by the relevant values could be infinite. This conclusion is unclear because if we only consider absolute values of the transforms for ξ and A large, in Proposition 3.1.1 we lose the cancellation induced by the oscillation of $w_{\xi,b}$ while in Proposition 3.1.2 when $A = 1/a$ grows the mass of $\psi_{a,b}$ is constrained to a set of size comparable to a.

Truly both transforms are not very practical as they are because computing numerically highly oscillatory integrals on $\mathbb{R}^2$ is not so simple to implement. In the next section we will see a Fourier series expansion with wavelets that is more friendly for computations because it allows to consider approximations by partial sums. The most common approach in applications employs a fully discrete wavelet transform involving only a finite number of values that it is only vaguely related to (3.10).

3.1.2 The theoretical framework

To get some intuition about the idea behind multiresolution analysis, let us consider the characteristic function f of the interval $[0,1)$ and let us try to analyze it in terms of the Haar wavelet (3.13) following an iterative procedure. We define f_1 to be the function that gives the average of f in the doubled interval $[0,2)$ along this interval and that is 0 otherwise. The difference is easily related to the Haar wavelet ψ.



In a formula, $f(x) - f_1(x) = 2^{-1}\psi(x/2)$. Now f_1 is the same as f changing the scale in the X and Y axes by factors 2 and 2^{-1} and we repeat the procedure defining in general f_n as the average of f_{n-1} in its doubled support to get

(3.23)
$$f(x) - f_1(x) = \frac{1}{2}\psi(\frac{x}{2}), \quad f_1(x) - f_2(x) = \frac{1}{4}\psi(\frac{x}{4}), \ \ldots \quad f_{n-1}(x) - f_n(x) = \frac{1}{2^n}\psi(\frac{x}{2^n}), \ldots$$

Weierstrass M-test allows to sum all of these formulas to get

$$f(x) = \sum_{n=1}^{\infty} 2^{-n}\psi(x/2^n) \tag{3.24}$$

with uniform convergence. We have also convergence in L^2 but not in L^1 because the integral of ψ vanishes and the integral of f does not.

Imagine that instead this silly f we have something more involved, say a function in L^2 (as smooth as you wish, if you prefer so). It can be approximated by step functions. If the steps are of width 2^{-m} we can perform the same analysis with each step as we did with the characteristic function of $[0,1)$ but now applying the scaling and translation $x \mapsto 2^m x - k$. When the approximation by step functions is finer we will obtain more positive exponents $m - n$ in the powers of 2. In the limit, assuming the convergence, we would obtain an expansion of any $f \in L^2$ in terms of the Haar wavelet:

$$f(x) = \sum_{j,k\in\mathbb{Z}} a_{jk}\psi(2^j x - k). \tag{3.25}$$