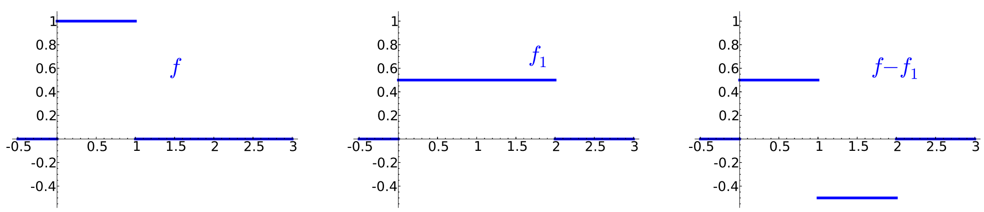
Truly both transforms are not very practical as they are because computing numerically highly oscillatory integrals on $\mathbb{R}^2$ is not so simple to implement. In the next section we will see a Fourier series expansion with wavelets that is more friendly for computations because it allows to consider 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).

Suggested Readings. Many texts employ the windowed Fourier transform or other alike transform to motivate the wavelet transform ant to compare them. In [Dau92, §1.2] there is an interesting discussion about this latter point. Another references are [Mal09, §4.2, §4.3], [Chu92, Ch.3] and [Kai94, Ch.2].

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)

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

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 higher 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:

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

Now there is something mathematically interesting that motivates the early introduction of the Haar wavelet [Haa10]. It turns out that the functions $\psi(2^j x - k)$ are orthogonal with the $L^2(\mathbb{R})$ scalar product. As $\|\psi\|_2 = 1$, the functions

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k) \qquad \text{with} \quad j, k \in \mathbb{Z} \tag{3.26}$$

are orthonormal in $L^2(\mathbb{R})$ and it allows to express the coefficients in (3.25) in a closed form as we did for the Fourier coefficients in (1.34). In a formula, we have for $f \in L^2$

$$f = \sum_{j,k\in\mathbb{Z}} c_{jk}\psi_{jk} \qquad \text{with} \quad c_{jk} = \int_{-\infty}^{\infty} f\overline{\psi}_{jk}. \tag{3.27}$$

The conjugation is superfluous but we keep it to have the result for any *orthonormal basis* of $L^2(\mathbb{R})$. Here we follow the slightly confusing common terminology: In Hilbert spaces, orthonormal basis means complete orthonormal system. It is not a basis in the sense of linear algebra because there only finite linear combinations are considered.

With (3.27) we have a kind of self-similar alternative for Fourier series and we could have obtained (3.24) from it. The self-similarity and the compact support are good having in mind the previous section, because they allow to separate neatly different scales. On the other hand, the lack of regularity does not seem very convenient.

Proposition 3.1.2 suggested a kind of orthogonality of normalized continuum wavelets but (3.27) is, no doubt, a cleaner formula and motivates to introduce a new concept of wavelet, probably the favorite for mathematicians. We define an *orthonormal wavelet* (very often simply a *wavelet*) to be a function $\psi \in L^2(\mathbb{R})$ such that (3.26) form an orthonormal basis for $L^2(\mathbb{R})$. In particular, (3.27) holds in L^2 sense, for orthonormal wavelets.

If we want to compete with Fourier system, we would like to have good convergence properties for smooth functions, say for instance $f \in C_0^\infty$. Let us think for instance about c_{j0} if ψ is bounded we clearly have $c_{j0} = O(2^{j/2}\|f\|_1)$ that goes to 0 when $j \to -\infty$. On the other hand, using the n-th Taylor approximation of f at 0,

$$2^{j/2}c_{j0} = \int_{-\infty}^{\infty} f(2^{-j}x)\overline{\psi}(x)\,dx = \sum_{m=0}^{n-1} 2^{-jm}\frac{f^{(m)}(0)}{m!}M_m + O\big(2^{-jn}\|f^{(n)}\|_\infty\|x^n\psi\|_1\big) \tag{3.28}$$

where $M_m = \int_{-\infty}^{\infty} x^m\psi(x)\,dx$. Then when we have more vanishing moments we have a quicker decay when $j \to \infty$. It can be proved that the very definition of orthonormal wavelet if ψ is smooth and $O(|x|^{-\alpha})$ implies $M_m = 0$ for $m < \alpha - 1$ [HW96, §2.3].

The aim is to construct orthonormal wavelets, mainly theoretically in this section. First of all, let us state a nice result characterizing the orhtonormality under translations.

Proposition 3.1.3. *Let $f \in L^2(\mathbb{R})$. Then $\{f(\cdot - k)\}_{k\in\mathbb{Z}}$ is an orthonormal system if and only if*

$$\sum_{k\in\mathbb{Z}} |\widehat{f}(\xi+k)|^2 = 1 \qquad \textit{almost everywhere.} \tag{3.29}$$

Proof. By Parseval identity,

$$\int_{-\infty}^{\infty} \overline{f}(x) f(x-k)\, dx = \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2 e(-k\xi)\, d\xi = \int_0^1 \sum_{l\in\mathbb{Z}} |\widehat{f}(\xi+l)|^2 e(-k\xi)\, d\xi \tag{3.30}$$

where in the last equality the reasoning is as in (2.10) and the interchange of the sum and the integral is justified by Lebesgue's dominated convergence theorem. If $\{f(\cdot - k)\}_{k\in\mathbb{Z}}$ is orthonormal then the sum in (3.29) minus 1 defines an integrable function on $\mathbb{T}$ and the integral equality shows that all of its Fourier coefficients are zero, then it is zero almost everywhere[2] [Zyg88, §I.6]. The converse is similar. □

It is possible to generalize this result in the following way: The sum in (3.29) is bounded by positive constants c_1 and c_2 from below and above if and only if $\{f(\cdot - k)\}_{k\in\mathbb{Z}}$ is a *Riesz system* with the same constants. This means

$$c_1 \sum |\lambda_k|^2 \le \big\| \sum \lambda_k f(\cdot - k)\big\|_2^2 \le c_2 \sum |\lambda_k|^2 \qquad \text{for any } \{\lambda_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z}). \tag{3.31}$$

As a curiosity, for the Haar wavelet the orthonormality of the translates is obvious but (3.29) is not and the same can be said for the characteristic function of $[0,1)$. Recalling (3.14) we have, respectively

$$\sum_{k\in\mathbb{Z}} \frac{\sin^4\big(\pi(\xi+l)/2\big)}{(\xi+l)^2} = \frac{\pi^2}{4} \qquad \text{and} \qquad \sum_{k\in\mathbb{Z}} \frac{\sin^2(\pi\xi)}{(\xi+l)^2} = \pi^2. \tag{3.32}$$

For the Shannon wavelet (3.15) and any of its translations the condition (3.29) is obvious by (3.16) and Proposition 3.1.3 gives the orthogonality with zero calculations.

There exists a full characterization of orthonormal wavelets using Proposition 3.1.3 and a corresponding result involving the scaling [HW96, §7.1] but it does not any clue about how to fulfill our aim of constructing wavelets.

The common theoretical approach is to define firstly the spaces in which we are going to work at each scale and the "brick" we are going to use. To get (3.25) for the Haar wavelet we employed the approximation of L^2 functions in spaces of step functions with step-width a power of 2. Each step is a scaled version of the characteristic function of $[0,1)$ that becomes our brick.

[2] I do not resist the temptation of citing [New74] with a surprising and short complex variable proof of the analogue of this fact for Fourier integrals.

In general, a *multiresolution analysis*, abbreviated *MRA*, is a sequence of nested subspaces $V_j \subset V_{j+1} \subset L^2(\mathbb{R})$, $j \in \mathbb{Z}$, and a function $\phi \in L^2(\mathbb{R})$ such that

$$\text{(3.33)} \qquad \begin{cases} \text{a)} \quad \{\phi(\cdot - k)\}_{k\in\mathbb{Z}} \text{ is an orthonormal basis of } V_0, \\ \text{b)} \quad V_j = \{f \,:\, f(2^{-j}\cdot) \in V_0\}, \\ \text{c)} \quad \bigcup V_j \text{ is dense in } L^2(\mathbb{R}) \text{ and } \bigcap V_j = \{0\}. \end{cases}$$

The function is called the *scaling function* of the MRA. It is the "brick" that dictates the subspaces when we combine the first and the second properties. Note that $2^{1/2}\phi(2x-k)$ is an orthonormal basis of V_1 and in general $2^{j/2}\phi(2^jx-k)$ is an orthonormal basis of V_j. It does not mean that ϕ is an orthonormal wavelet. At the contrary, the inclusion $V_{-1} \subset V_0$ implies that the functions $2^{-1/2}\phi(x/2-k)$ can be expanded in terms of $\phi(x-k)$, in particular they are not orthogonal to them. Let c_n be the coefficients of $2^{-1/2}\phi(x/2)$ and define $\sqrt{2}m_0(-\xi)$ having these Fourier coefficients:

$$\text{(3.34)} \qquad 2^{-1/2}\phi(x/2) = \sum_{k\in\mathbb{Z}} c_k\phi(x-k) \qquad \text{and} \qquad m_0(\xi) = 2^{-1/2}\sum_{k\in\mathbb{Z}} c_k e(-k\xi).$$

The role of $m_0(\xi)$ is clarified if we take Fourier transforms in the first equation to get

$$\text{(3.35)} \qquad \widehat{\phi}(2\xi) = m_0(\xi)\widehat{\phi}(\xi).$$

Then $m_0 \in L^2(\mathbb{T})$ indicates how to filter the frequencies of the signal ϕ when the scale is changed. In the literature m_0 is called the *low-pass filter* of the MRA. In some way, the functions in V_{-1} oscillate twice slower than in V_0 then m_0 suppresses high frequencies, it is truly a low-pass filter.

Comparing with the example with the Haar wavelets, the V_j would be the spaces of finer and finer step functions and the wavelets of different scales live in the "differences" between consecutive spaces. More precisely, let us define W_j to be the orthogonal complement of V_j in V_{j+1}

$$\text{(3.36)} \qquad W_j = V_{j+1} \cap V_j^{\perp}.$$

If we find a ψ such that $\{\psi(\cdot - k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0, then $\{\psi_{jk}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_j for each j and this is enough to prove that ψ is a wavelet because the third property of (3.1.4) implies

$$\text{(3.37)} \qquad V_j = \bigoplus_{k<j} W_k \qquad \text{and} \qquad L^2(\mathbb{R}) = \overline{\bigoplus_{j\in\mathbb{Z}} W_j}.$$

In some sense, W_k is the "difference" between V_{k+1} and V_j and these sums telescope resembling the argument in (3.23) and (3.24).

The main result in this section says that if we have a MRA we can construct an orthonormal wavelet $\psi \in W_0$ from ϕ and m_0.

Theorem 3.1.4. *Let ϕ be the scaling function of a MRA and m_0 the associated low-pass filter defined by* (3.34) *or* (3.35). *Then a function $\psi \in W_0$ is an orthonormal wavelet if and only if*

$$\widehat{\psi}(2\xi) = e(\xi)\nu(2\xi)\overline{m}_0(\xi + \frac{1}{2})\widehat{\phi}(\xi) \tag{3.38}$$

almost everywhere for some $\nu \in L^2(\mathbb{T})$ with $|\nu(\xi)| = 1$.

The proof of this result takes some effort. Before entering into it let us see two examples producing the simplest orthonormal wavelets.

If ϕ is the characteristic function of $[0, 1)$ the V_j's are spaces of step functions which become finer when j grows and the conditions of MRA are fulfilled. It is clear

$$2^{-1/2}\phi(x/2) = 2^{-1/2}\phi(x) + 2^{-1/2}\phi(x-1). \tag{3.39}$$

Then by (3.34), $m_0(\xi) = \big(1+e(-\xi)\big)/2$. By Theorem 3.1.4 we have infinitely many wavelets to our disposal choosing ν. Let us take $\nu(\xi) = -e(-\xi)$ to get

$$2\widehat{\psi}(2\xi) = \widehat{\phi}(\xi) - e(-\xi)\widehat{\phi}(\xi). \tag{3.40}$$

Taking Fourier inverses, $\psi(x/2) = \phi(x) - \phi(x - 1/2)$ that is just the definition of the Haar wavelet (3.13).

Now we take $\phi(x) = \operatorname{sinc} x$. Note that $\widehat{\phi} = \chi^*_{[-1/2,1/2]}$ with χ^* as in (3.16) that satisfies (3.29) trivially, then the first condition in (3.33) is fulfilled for V_0 generated by the orthonormal basis, the second can be taken as a definition and the third could be checked noting that the Fourier transforms of the finite linear combinations of $\phi(2^j x - k)$ are step functions as before, but we skip this point now because later we will see a general result that gives a simple condition to get the last property in (3.33). The relation (3.35) proves $m_0 = \chi^*_{[-1/4,1/4]}$ for $|\xi| < 1/2$ and by the periodicity

$$m_0(\xi) = \sum_{k\in\mathbb{Z}} \chi^*_{[-1/4,1/4]}(\xi + k) \qquad \text{almost everywhere.} \tag{3.41}$$

Choosing $\nu(\xi) = e(-\xi)$ in (3.38), we have

$$\widehat{\psi}(2\xi) = e(-\xi)\big(\chi^*_{[-1/2,-1/4]}(\xi) + \chi^*_{[1/4,1/2]}(\xi)\big) = e(-\xi)\widehat{\psi}_S(2\xi) \tag{3.42}$$

where ψ_S is the Shannon wavelet as in (3.15) and (3.16). Then we conclude that $\psi_S(x-1/2)$ is an orthonormal wavelet.

The choice $\nu(\xi) = -e(-\xi)$ employed before leads to a neat expansion of the wavelet in terms of the scaling function

Corolary 3.1.5. *If ϕ is the scaling function of a MRA then*

$$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} (-1)^k \overline{c}_{1-k}\phi(2x - k), \tag{3.43}$$

with c_k as in (3.34), *is an orthonormal wavelet.*

Proof. By (3.1.4) with $\nu(\xi) = -e(-\xi)$ and the formula for m_0 in (3.34),

$$\widehat{\psi}(\xi) = -2^{-1/2} \sum_{k\in\mathbb{Z}} (-1)^k \overline{c}_k e((k-1)\xi/2)\widehat{\phi}(\xi/2). \tag{3.44}$$

By Fourier inversion

$$\psi(x) = -2^{1/2} \sum_{k\in\mathbb{Z}} (-1)^k \overline{c}_k \phi(2x+k-1), \tag{3.45}$$

that is the result when $k \mapsto 1-k$. □

To prove Theorem 3.1.4, we proceed in two steps characterizing the Fourier transform of the functions belonging to the space V_{-1} and to its orthogonal complement in V_0. We follow [HW96, §2.2] through [LW18] to keep the usual normalization of the Fourier transform.

Lemma 3.1.6. *In a MRA with scaling function* ϕ*,*

$$V_{-1} = \Big\{ f \in L^2(\mathbb{R}) \ : \ \widehat{f}(\xi) = p(2\xi) m_0(\xi)\widehat{\phi}(\xi) \quad \textit{for some} \quad p \in L^2(\mathbb{T}) \Big\}. \tag{3.46}$$

Lemma 3.1.7. *In a MRA with scaling function* ϕ*,*

$$W_{-1} = \Big\{ f \in L^2(\mathbb{R}) \ : \ \widehat{f}(\xi) = e(\xi) p(2\xi) \overline{m}_0(\xi + \frac{1}{2})\widehat{\phi}(\xi) \quad \textit{for some} \quad p \in L^2(\mathbb{T}) \Big\}. \tag{3.47}$$

The first result follows easily from the definition of V_{-1} through the second property of (3.33).

Proof of Lemma 3.1.6. We know that $\{\phi(\cdot - k)\}_{k\in\mathbb{Z}}$ spans V_0 then any $f \in V_{-1}$ can be expanded as

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \phi(x-k) \qquad \text{with} \quad \{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z}). \tag{3.48}$$

Taking Fourier transforms,

$$\widehat{f}(\xi) = 2 \sum_{k\in\mathbb{Z}} a_k e(-2k\xi)\widehat{\phi}(2\xi) = 2 m_0(\xi)\widehat{\phi}(\xi) \sum_{k\in\mathbb{Z}} a_k e(-2k\xi) \tag{3.49}$$

by (3.35). As $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ is arbitrary, every function of $L^2(\mathbb{T})$ evaluated at 2ξ can be written as the previous sum. □

In the proof of the second result, a noticeable identity plays a role which we separate for further reference.

Lemma 3.1.8. *Let* m_0 *the low-pass filter of a MRA, defined by* (3.34)*. Then*

$$|m_0(\xi)|^2 + |m_0(\xi + 1/2)|^2 = 1 \quad \textit{almost everywhere.} \tag{3.50}$$

Proof. By Proposition 3.1.3, (3.35) and the periodicity of m_0,

$$1 = \sum_{k\in\mathbb{Z}} |\widehat{\phi}(2\xi+k)|^2 = |m_0(\xi)|^2 \sum_{2|k} |\widehat{\phi}(\xi+k/2)|^2 + |m_0(\xi+1/2)|^2 \sum_{2\nmid k} |\widehat{\phi}(\xi+k/2)|^2. \tag{3.51}$$

In the right hand side both sums equal 1 by Proposition 3.1.3, the first one by a direct application and the second one replacing ξ by $\xi - 1/2$. ☐

Proof of Lemma 3.1.7. If $f \in V_0$ then $f(x) = \sum a_k \phi(x-k)$ with $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ or, equivalently, the functions of V_0 are characterized by

$$\widehat{f}(\xi) = q(\xi)\widehat{\phi}(\xi) \qquad \text{with} \quad q(\xi) = \sum_{k\in\mathbb{Z}} a_k e(-k\xi) \in L^2(\mathbb{T}). \tag{3.52}$$

The necessary and sufficient condition to have $f \in V_{-1}^{\perp}$ is, after Parseval identity and Lemma 3.1.6,

$$0 = \int_{-\infty}^{\infty} \overline{p}(2\xi)\overline{m}_0(\xi)q(\xi)|\widehat{\phi}(\xi)|^2\, d\xi = \sum_{k\in\mathbb{Z}} \int_0^1 \overline{p}(2\xi)\overline{m}_0(\xi)q(\xi)|\widehat{\phi}(\xi+k)|^2\, d\xi \tag{3.53}$$

for every $p \in L^2(\mathbb{T})$, where we have used the periodicity for the last equality. Using Proposition 3.1.3 and defining $F(\xi) = e(-\xi)\overline{m}_0(\xi)q(\xi)$, the condition is

$$0 = \int_0^1 e(\xi)\overline{p}(2\xi)F(\xi)\, d\xi = \int_0^{1/2} e(\xi)\overline{p}(2\xi)\big(F(\xi) - F(\xi+1/2)\big)\, d\xi. \tag{3.54}$$

As p is arbitrary in $[0,1/2]$, this is the same as saying that F is 1/2-periodic. Lemma 3.1.8 implies that $m_0(\xi)$ and $m_0(\xi+1/2)$ are bounded and do not vanish simultaneously. Then we can write $F(\xi) = F(\xi+1/2) = \overline{m}_0(\xi)\overline{m}_0(\xi+1/2)p(2\xi)$ with $p \in L^2(\mathbb{T})$ arbitrary. Then $q(\xi) = e(\xi)\overline{m}_0(\xi+1/2)p(2\xi)$ and substituting in (3.52) the proof is finished. ☐

Proof of Theorem 3.1.4. Let us check firstly that any orthonormal wavelet $\psi \in W_0$ is of the form (3.38). It is easy to see by the second property of (3.33) and the definition (3.36),

$$f \in W_0 \qquad \text{if and only if} \qquad f(2^j\cdot) \in W_j. \tag{3.55}$$

Considering $j = -1$ and taking Fourier transforms, we have by Lemma 3.1.7

$$\widehat{\psi}(2\xi) = e(\xi)p(2\xi)\overline{m}_0(\xi+\frac{1}{2})\widehat{\phi}(\xi). \tag{3.56}$$

We apply Proposition 3.1.3 separating even and odd integers as in the proof of Lemma 3.1.8,

$$\sum_{k\in\mathbb{Z}} |\widehat{\psi}(2\xi+k)|^2 = |p(2\xi)|^2|m_0(\xi+1/2)|^2 \sum_{2|k} |\widehat{\phi}(\xi+k/2)|^2 + |p(2\xi)|^2|m_0(\xi)|^2 \sum_{2\nmid k} |\widehat{\phi}(\xi+k/2)|^2. \tag{3.57}$$

The sums are 1 by Proposition 3.1.3 and Lemma 3.1.8 gives $|p(2\xi)| = 1$ as required in (3.38).

We have to prove now that (3.38) defines an orthonormal wavelet. It belongs to W_0 by Lemma 3.1.7 and (3.55) with $j=-1$. We can write any $f \in W_{-1}$ as

$$\widehat{f}(\xi) = p(2\xi)\overline{\nu}(2\xi)\widehat{\psi}(2\xi) \tag{3.58}$$

If $\sum a_k e(-k\xi)$ is the Fourier expansion of the periodic function $p(\xi)\overline{\nu}(\xi)$, taking inverse Fourier transforms

$$f(x) = \frac{1}{2}\sum_{k\in\mathbb{Z}} c_k \psi(x/2-k). \tag{3.59}$$

Then $\{\psi_{-1k}\}_{k\in\mathbb{Z}}$ spans W_{-1} and it is orthonormal proceeding as in (3.57). By (3.55), $\{\psi_{0k}\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0 and as explained in (3.37) it is enough to assure that ψ is an orthonormal wavelet. □

Suggested Readings. The rigorous development of the theory can be read in books addressed to a mathematical audience. My favorites are [Bré02], [Pin02] and [HW96]. The latter, one of the pioneering textbooks, takes quite effort to discuss some properties related to analysis, like convergence or atomic decomposition of functions. By reasons that probably rely on tradition, my feeling is that very rarely the literature about wavelets uses the most standard normalization of the Fourier transform (1.36), [Pin02] is an exception.