Wavelets

Demetrio Labate, Guido Weiss, Edward Wilson

August 14, 2012

1 Introduction

The subject called "wavelets" is made up of several areas of pure and applied mathematics. It has contributed to the understanding of many problems in various sciences, engineering and other disciplines, and it includes, among its notable successes, the wavelet-based digital fingerprint image compression standard adopted by the FBI in 1993 and JPEG2000, the current standard for image compression.

We will begin by describing what wavelets are in one dimension and, then, pass to more general settings, trying to keep the presentation at a non-technical level as much as possible. We assume that the reader knows a bit of harmonic analysis. In particular, we assume knowledge of the basic properties of Fourier series and Fourier transforms. We start by establishing the basic definitions and notations which will be used in the following.

The space $L^2(\mathbb{R})$ is the Hilbert space of all square (Lebesgue) integrable functions endowed with the inner product $\langle f, g\rangle = \int_{\mathbb{R}} f\,\overline{g}$. The *Fourier transform* $\mathcal{F}$ is the unitary operator that maps $f \in L^2(\mathbb{R})$ into the function $\mathcal{F}f = \hat{f}$ defined by

$$(\mathcal{F}f)(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}} f(x)\, e^{-2\pi i \xi x}\, dx$$

when $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and by the "appropriate" limit for the general $f \in L^2(\mathbb{R})$. We refer to the variable x as the *time* variable and to ξ as the *frequency* variable. Notice that the function $\hat{f}$ is also square integrable. Indeed, $\mathcal{F}$ maps $L^2(\mathbb{R})$ one-to-one onto itself . The inverse $\mathcal{F}^{-1}$ of $\mathcal{F}$ is defined by

$$(\mathcal{F}^{-1}g)(x) = \check{g}(x) = \int_{\widehat{\mathbb{R}}} g(\xi)\, e^{2\pi i x \xi}\, d\xi.$$

The functions $\{e_k(x) = e^{2\pi i k x} : k \in \mathbb{Z}\}$ are 1-periodic and form an orthonormal basis of $L^2(\mathbb{T})$, where $\mathbb{T}$ is the 1-torus and can be identified with any of the sets $(0, 1]$ or $[-\frac{1}{2}, \frac{1}{2})$ or $[-1, -\frac{1}{2}) \cup [\frac{1}{2}, 1)$, . . . , (all having measure one). We denote the *Fourier series* of f, 1-periodic and in $L^p(\mathbb{T})$, by:

$$\sum_{k\in\mathbb{Z}} \langle f, e_k\rangle_{\mathbb{T}}\, e_k \sim f,$$

where $\langle f, e_k\rangle_{\mathbb{T}} = \int_{\mathbb{T}} f\,\overline{e_k}$, and $k \in \mathbb{Z}$.

The paper is organized as follows. In Section 2, we introduce one dimensional wavelets; in Section 3, we discuss wavelets in higher dimensional Euclidean spaces; Section 4 introduces continuous wavelets and some applications; finally, Section 5 discusses other applications and makes some concluding remarks.

2 Wavelets in $L^2(\mathbb{R})$

We consider two sets of unitary operators on $L^2(\mathbb{R})$: the *translations* T_k, $k \in \mathbb{Z}$, defined by $(T_k f)(x) = f(x-k)$ and the (dyadic) *dilations* D_j, $j \in \mathbb{Z}$, defined by $(D_j f)(x) = 2^{j/2} f(2^j x)$. A *wavelet* (more precisely, a *dyadic* *wavelet*) is a function $\psi \in L^2(\mathbb{R})$ having the property that the system $\mathcal{W}_\psi = \{\psi_{j,k} = (D_j\, T_k)\psi : j, k \in \mathbb{Z}\}$ is an orthonormal (ON) basis of $L^2(\mathbb{R})$. Notice that the order of applying first translations and, then, dilations is important: $D_j\, T_k = T_{2^{-j}k}\, D_j$.

In this section, we will explain why there are many wavelets enjoying a large number of useful properties which makes it plausible that various different types of functions (or signals) can be expressed efficiently by appropriate wavelet bases.

It is often stated that Haar in 1910 [19] exhibited a wavelet $\psi = \psi^H$ and it took about 70 years before a large number of different wavelets appeared in the world of Mathematics. This is really not the case. The *Haar wavelet* is defined by $\psi^H = \chi_{[0,\frac{1}{2})} - \chi_{[\frac{1}{2},1)}$ and it is not difficult to show that it is a wavelet (as will be shown below). Another simple example is the *Shannon wavelet*; it appeared in the 1940's and we will explain in what sense it appeared. It is defined as $\psi^S = \check{\chi}_S$, where $S = [-1, -\frac{1}{2}) \cup [\frac{1}{2}, 1)$. A straightforward calculation shows that, if $\psi \in L^2(\mathbb{R})$, then, for $j, k \in \mathbb{Z}$,

$$(\hat{\psi}_{j,k})(\xi) = [2^{-j/2} e^{-2\pi i k 2^{-j}\xi}]\, \hat{\psi}(2^{-j}\xi). \tag{2.1}$$

Let us observe that the sets $2^j S$, $j \in \mathbb{Z}$, form a mutually disjoint covering of $\mathbb{R} \setminus \{0\}$. Moreover, since the system $\{e_{-k}\chi_S : k \in \mathbb{Z}\}$ is an ON basis of $L^2(S)$, the functions within the square bracket in (2.1), restricted to the set $2^j S$, form an ON basis of $L^2(2^j S)$ for each $j \in \mathbb{Z}$. It follows immediately that the set $\{\psi^S_{j,k} : j, k \in \mathbb{Z}\}$ is an ON basis of $L^2(\mathbb{R})$. This shows that ψ^S is a wavelet.

We mentioned that it is not difficult to show that the Haar function ψ^H is a wavelet. We will do this together with the presentation of a general method for constructing wavelets: the *Multiresolution Analysis* (MRA) method introduced by S. Mallat with the help of R. Coifman and Y. Meyer [23, 26].

An MRA is a sequence $\{V_j : j \in \mathbb{Z}\}$ of closed subspaces of $L^2(\mathbb{R})$ satisfying:

(i) $V_j \subset V_{j+1}$ for all $j \in \mathbb{Z}$.

(ii) $V_{j+1} = D_1 V_j$ for all $j \in \mathbb{Z}$; that is, $f \in V_j$ iff $f(2\cdot) \in V_{j+1}$.

(iii) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.

(iv) $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$.

(v) There exists $\phi \in V_0$ such that $\{T_k\, \phi : k \in \mathbb{Z}\}$ is an ON basis of V_0.

The function ϕ described in (v) is called a *scaling* function of this MRA.

If $\{V_j : j \in \mathbb{Z}\}$ is an MRA, let W_j be the orthogonal complement of V_j within V_{j+1}. An immediate consequence of the above properties is that the spaces W_j, $j \in \mathbb{Z}$, are mutually orthogonal and their orthogonal direct sum $\bigoplus_{j\in\mathbb{Z}} W_j$ satisfies

$$\bigoplus_{j\in\mathbb{Z}} W_j = L^2(\mathbb{R}). \tag{2.2}$$

If there exists a function $\psi \in W_0$ such that $\{T_k\, \psi : k \in \mathbb{Z}\}$ is an ON basis of W_0, using the observation that $D_j W_0 = W_j$ for each $j \in \mathbb{Z}$ (an easy consequence of the MRA properties), we see that $\{\psi_{j,k} = D_j T_k\, \psi : k \in \mathbb{Z}\}$ is an ON basis of W_j. It follows from (2.2) that $\{\psi_{j,k} = D_j T_k\, \psi : j, k \in \mathbb{Z}\}$ is an ON basis of $L^2(\mathbb{R})$. Thus, ψ is a wavelet. In the case where $\phi = \chi_{[0,1)}$ and V_0 is the span of the ON system $\{T_k\, \phi : k \in \mathbb{Z}\}$, it is easy to check that $\{V_j = D_j V_0 : j \in \mathbb{Z}\}$ is an MRA. Moreover, it is easy to verify that $\{T_k\, \psi^H : k \in \mathbb{Z}\}$ is an ON basis of the space W_0 defined by $W_0 = V_0^\perp \subset V_1$. It follows that $\{D_j T_k\, \psi^H : j, k \in \mathbb{Z}\}$ is an ON basis of $L^2(\mathbb{R})$. This shows that the Haar function ψ^H is indeed a wavelet.

We leave it to the reader to verify that the Shannon wavelet ψ^S is an MRA wavelet as well. In fact, it corresponds to the scaling function $\phi(x) = \mathrm{sinc}(x) = \frac{\sin x\pi}{x\pi}$ (note: $\mathrm{sinc}(0) = 1$). This is a consequence of the fact that $(\mathrm{sinc})^\wedge(\xi) = \chi_{[-\frac{1}{2},\frac{1}{2})}(\xi) = \hat{\phi}(\xi)$.

We point out that there is an important result involving the function sinc, namely the following elementary theorem.

Theorem 2.1 (Whittaker-Shannon-Kotelnikov Sampling Theorem)**.** *Let* $f \in L^2(\mathbb{R})$ *and* $\operatorname{supp}\hat{f} \subset [-\frac{1}{2}, \frac{1}{2}]$*. Then*

$$f(x) = \sum_{k\in\mathbb{Z}} f(k)\,\operatorname{sinc}(x-k),$$

where the symmetric partial sums of this series converge in the L^2*-norm, as well as absolutely and uniformly.*

We will explain how this result is related to wavelets even though, when it was obtained, the notion of wavelets had not yet appeared. The word *sampling* reflects the fact that the functions involved are completely determined if we know their values on the countable set $\mathbb{Z}$. The name *Shannon* is singled out because he is associated with many important aspects and applications of sampling.

Let $\phi \in L^2(\mathbb{R})$, ϕ not the zero function, and $\mathcal{T}_\phi = \{\phi_k = T_k\,\phi : k \in \mathbb{Z}\}$. Then $\mathcal{T}_\phi$ generates the closed space $V_\phi := \langle\phi\rangle = \overline{\operatorname{span}\{\phi_k : k \in \mathbb{Z}\}}$, that is, the closure of all finite linear combinations of the functions ϕ_k. This space is shift-invariant and is called the *Principal Shift-Invariant Space* (PSIS) generated by ϕ. In case $\phi = \operatorname{sinc}$, then $\mathcal{T}_\phi = \langle\operatorname{sinc}\rangle$ is an orthonormal system (recall that $(\operatorname{sinc})^\wedge(\xi) = \chi_{[-\frac{1}{2},\frac{1}{2})}(\xi))$ and we have that $\sum_{k\in\mathbb{Z}}|f(k)|^2 < \infty$, where f is the function in Theorem 2.1. If $V_0 = \langle\operatorname{sinc}\rangle$, then the set $\{V_j = D_j V_0 : j \in \mathbb{Z}\}$ is an MRA and $\phi = \operatorname{sinc}$ is a scaling function for this MRA. For a general MRA with a scaling function ϕ, there is a bounded 1-periodic function m_0 known as a *low-pass filter* and an associated *high-pass filter* $m_1(\xi) = e^{2\pi i\xi}\,\overline{m_0(\xi + \frac{1}{2})}$ that produce the so-called *two-scale equations*:

$$\hat{\phi}(2\xi) = m_0(\xi)\hat{\phi}(\xi), \quad \hat{\psi}(2\xi) = m_1(\xi)\hat{\phi}(\xi). \tag{2.3}$$

In fact, these equations produce the desired wavelet ψ generated by the scaling function ϕ. In the special case we are considering, where $\phi = \operatorname{sinc}$, the low-pass filter, when restricted to the interval $[-\frac{1}{2}, \frac{1}{2}]$, is the function $\chi_{[-\frac{1}{4},\frac{1}{4}]}$. One can verify indeed that, in this case, $\hat{\psi}(\xi) = e^{-i\pi\xi}\,\hat{\psi}^S(\xi) = e^{-i\pi\xi}\,\chi_S(\xi)$. Thus, the wavelet we obtain is essentially the Shannon wavelet, since the factor $e^{-i\pi\xi}$ is irrelevant to the orthonormality of the system.

Let us point out that the spaces V_0 and W_0, in general, are PSIS's, but they have an important difference with respect to the dilation operators we are considering: $V_j = D_j\,V_0$ is an increasing sequence of closed spaces as $j \to \infty$, while the spaces $W_j = D_j\,W_0$ are disjoint and satisfy (2.2).

The properties of shift-invariant spaces have many consequences in the theory of wavelets. If ϕ is not the zero function in $L^2(\mathbb{R})$, let $p_\phi(\xi) = \sum_{j\in\mathbb{Z}}|\hat{\phi}(\xi + j)|^2$ and consider the space $\mathcal{M}_\phi = L^2([0,1], p_\phi)$ of all 1-periodic functions m satisfying

$$\int_0^1 |m(\xi)|^2\,p_\phi(\xi)\,d\xi := \|m\|^2_{\mathcal{M}_\phi} < \infty.$$

It is easy to check that the mapping $J_\phi : \mathcal{M}_\phi \mapsto \langle\phi\rangle = V_\phi$ defined by $J_\phi m = (m\,\hat{\phi})^\vee$ is an isometry onto $V_\phi \subset L^2(\mathbb{R})$. That is, the two spaces $\mathcal{M}_\phi$ and V_ϕ are "essentially equivalent" via the map J_ϕ. It is natural, therefore, to ask how the properties of the weight p_ϕ correspond to the properties of the generating system $\mathcal{T}_\phi$. For example, the functions $e_{-k}(\xi) = e^{-2\pi i k\xi}$, $k \in \mathbb{Z}$, are mapped by the mapping J_ϕ onto the functions $\phi(\cdot - k)$, $k \in \mathbb{Z}$. Since the set $\{e_{-k}(\xi) : k \in \mathbb{Z}\}$ is algebraically linearly independent, so is the system $\mathcal{T}_\phi$. It follows immediately that

(i) $\mathcal{T}_\phi$ *is an ON system iff* $p_\phi(\xi) = 1$ *a.e.*

(ii) $p_\phi(\xi) > 0$ *a.e. iff there exists an ON basis of* V_ϕ *of the form* $\{T_k\psi : k \in \mathbb{Z}\}$ *for some* $\psi \in V_\phi$*.*

In [20, 7], many properties of p_ϕ are shown to be equivalent to properties of V_ϕ or $\mathcal{T}_\phi$. For example one can show the following.

(iii) The system $\mathcal{T}_\phi$ *is a frame for* V_ϕ *in the sense that we have constants* $0 < A \le B < \infty$ *for which*

$$A\sum_{k\in\mathbb{Z}}|\langle f, T_k\phi\rangle|^2 \le \langle f, f\rangle \le B\sum_{k\in\mathbb{Z}}|\langle f, T_k\phi\rangle|^2$$

for each $f \in V_\phi$ *iff*

$$A\, \chi_{\Omega_\phi}(\xi) \le p_\phi(\xi) \le B\, \chi_{\Omega_\phi}(\xi), \quad a.e.,$$

where $\Omega_\phi = \{\xi \in [0,1] : p_\phi(\xi) > 0\}$.

Note that, in general, when $\tilde{\phi} = (\frac{\chi_{\Omega_\phi}}{p_\phi}\hat{\phi})^\vee$, then $\tilde{\phi} \in V_\phi = V_{\tilde{\phi}}$ and $p_{\tilde{\phi}} = \chi_\Omega$ a.e. Moreover, $\mathcal{T}_{\tilde{\phi}}$ is a Parseval frame (PF) for V_ϕ; that is, it is a frame with $A = B = 1$. Slightly more general than ON MRA wavelets are PF MRA wavelets ψ where $\mathcal{T}_\psi$ is a PF for W_0 and there is a scaling function ϕ generating a PF for V_0. Examples include the function ψ given by $\hat{\psi} = \chi_{U \setminus \frac{1}{2}U}$, for $U \subset [-\frac{1}{2}, \frac{1}{2})$, where U has positive measure and $\frac{1}{2}U \subset U$.

One of the most celebrated contributions to the construction of MRA wavelets was made by I. Daubechies [1, 2], who used an ingenious construction to produce MRA wavelets which are compactly supported, and can have high regularity and many vanishing moments, where the *kth moment* of ψ is defined as the integral $\int_{\mathbb{R}} x^k\, \psi(x)\, dx$. These wavelets are very useful for applications in numerical analysis and engineering since the wavelet expansions of a piecewise smooth function converge very rapidly to the function. Specifically, suppose that $f \in C^R(\mathbb{R})$, the space of R times differentiable functions such that $\|f\|_{C^R} = \max\{\|f^{(s)}\|_\infty : s = 0, \dots, R\} < \infty$, and ψ is a compactly supported wavelet having at least R vanishing moments. Choose a bijection $\pi : \mathbb{N} \mapsto \mathbb{Z} \times \mathbb{Z}$ such that $|\langle f, \psi_{\pi(k)}\rangle| \ge |\langle f, \psi_{\pi(k+1)}\rangle|$ for all $k \ge 1$; that is, the wavelet coefficients of f are ordered in non-increasing order of magnitude. Then one can show [21, Thm. 7.16] that

$$|\langle f, \psi_{\pi(m)}\rangle| \le C\, \|f\|_{C^R}\, m^{-(R+\frac{1}{2})}, \tag{2.4}$$

where C is a constant independent of f and m. The implication of this is that relatively few coefficients are needed to get a good approximation of f. In fact, letting f_N be the best N-term nonlinear approximation of f, the *nonlinear approximation error* decays as

$$\|f - f_N\|_{L^2}^2 \le C \sum_{m>N} |\langle f, \psi_{\pi(m)}\rangle|^2 \le C\, \|f\|_{C^R}\, N^{-2R}. \tag{2.5}$$

Remarkably, this result holds also if f is R times continuously differentiable up to finitely many jump discontinuities. That is, the wavelet approximation behaves as if the functions had no discontinuities. This behaviour is very different from Fourier approximations, in which case the error rate is of the order $O(N^{-2})$. These results have extensions to higher dimensions (see further discussion in Sec. 3).

Another important property of an MRA is that it enables an efficient algorithmic implementation of the wavelet decomposition. To explain this, suppose that ψ is a compactly supported MRA wavelet associated with a compactly supported scaling function ϕ. Because ϕ and ψ are in V_1, it follows that $D_{-1}\phi$ and $D_{-1}\psi$ are in V_0. Let us examine equations (2.3) which, in the time domain, can be written as

$$(D_{-1}\phi)(x) = \frac{1}{\sqrt{2}}\, \phi(\frac{x}{2}) = \sum_{k\in\mathbb{Z}} a_k\, \phi(x-k),\ (D_{-1}\psi)(x) = \frac{1}{\sqrt{2}}\, \psi(\frac{x}{2}) = \sum_{k\in\mathbb{Z}} b_k\, \phi(x-k), \tag{2.6}$$

where only finitely many coefficients a_k and b_k are not zero. Since D_{-1} is unitary, from (2.6) we obtain

$$a_k = \langle D_{-1}\phi, T_k\phi\rangle = \langle D_{j-1}\phi, D_j T_k \phi\rangle,\ b_k = \langle D_{-1}\psi, T_k\phi\rangle = \langle D_{j-1}\psi, D_j T_k \phi\rangle. \tag{2.7}$$

We see, therefore, that there are two ON bases for $V_j = W_{j-1} \oplus V_{j-1}$; they are $D_j\mathcal{T}_\phi$ and $D_{j-1}\mathcal{T}_\psi \cup D_{j-1}\mathcal{T}_\phi$. From equalities (2.7), we can calculate the matrices of the change of bases derived from these two ON bases (keeping in mind that the order of dilations and translations is important: $T_k D_1 = D_1 T_{2k}$). For a given compactly supported $f_j \in V_j$, we apply the appropriate change of basis matrix to compute the orthogonal projection f_{j-1} of f_0 into V_{j-1} and obtain the "correction term" $e_{j-1} = f_j - f_{j-1} \in W_{j-1}$. Iterating this procedure, we see that arithmetic manipulations with finite sets of coefficients are all that is involved to compute $f_0 \in V_0$ and $e_i \in W_i$, $0 \le i \le j-1$, so that $f_j = f_0 + e_1 + \cdots + e_{j-1}$. Together with the excellent approximation properties of wavelets, this remarkably simple technique is one of the main reasons why engineers adopted wavelets (specifically, the so-called compactly supported biorthogonal wavelets [8]) in the design of JPEG2000, the industrial standard for image compression replacing the older Fourier-based JPEG standard.

Another property of MRA wavelets is that, considered as members of a subset of the unit sphere in $L^2(\mathbb{R})$, they form an arcwise connected set. In particular, it is not hard to show that there are continuous paths of wavelets. Suppose that ψ is an MRA wavelet and $\tilde{\psi}$ is a wavelet in the same MRA. Then one can easily show that $(\tilde{\psi})^\wedge = s\,\hat{\psi}$, where s is a 1-periodic unimodular function. In particular, we can choose a 1-periodic function θ for which $(\tilde{\psi})^\wedge = e^{i\theta}\,\hat{\psi}$ and set $s^t = e^{it\theta}$, for $t \in [0,1]$, to establish a continuous path $t \mapsto (s^t\hat{\psi})^\vee$ in $L^2(\mathbb{R})$ connecting ψ to $\tilde{\psi}$. A more complicated argument shows how ψ is continuously connected to the Haar wavelet [13]. Other related questions arise naturally. For example: are all ON wavelets connected? are any two frame wavelets connected? The answer to this last question is “yes”, whereas the previous question is still an open problem.

Before moving to the topic of wavelets in higher dimensions, let us state that there are many other facts about one-dimensional wavelets we have not discussed. In particular, there are wavelets not arising from an MRA. Also, wavelets can be defined by replacing dyadic dilations with dilations by $r > 1$, where r need not be an integer. In the situation of non-dyadic dilations, the construction of the orthonormal bases associated with the wavelet may require more than one generator; namely, if $r = \frac{p}{q} > 1$, and p, q are relatively prime, then $p - q$ generators are needed.

3 Wavelets in higher dimensions

Many of the concepts in Section 2 extend naturally to n dimensions ($n \in \mathbb{N}$, $n > 1$) with $\mathbb{Z}$-translations replaced by $\mathbb{Z}^n$-translations and the dilation set $\{2^j : j \in \mathbb{Z}\}$ replaced by $\{u^j : j \in \mathbb{Z}\}$, where u is an $n \times n$ real matrix each of whose eigenvalues has magnitude larger than one. Both for theoretical and practical purposes, however, it is convenient to focus our attention on PF wavelets rather than ON wavelets. Thus, given the matrix u, we seek functions $\psi \in L^2(\mathbb{R}^n)$ for which the wavelet system

$$\{\psi_{j,k}(x) = (D_u^j\, T_k\, \psi)(x) = |\det u|^{j/2}\psi(u^j x - k),\ j \in \mathbb{Z}, k \in \mathbb{Z}^n\}, \tag{3.8}$$

is a PF for $L^2(\mathbb{R}^n)$. That is,

$$\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}^n} |\langle f, (D_u^j\, T_k\, \psi\rangle|^2 = \|f\|^2_{L^2(\mathbb{R}^2)},$$

for all $f \in L^2(\mathbb{R}^n)$. For example, let $n = 2$ and $u = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. For $U \subset [-\frac{1}{2}, \frac{1}{2})^2$ where U has positive measure and $\frac{1}{2}U \subset U$, the function ψ defined by $\hat{\psi} = \chi_{U\setminus\frac{1}{2}U}$ is a PF MRA wavelet. On the other hand, to obtain an ON MRA wavelet system, we need to use 3 wavelet generators.

As in the 1-dimensional case, we avoid multiple wavelet generators by restricting our attention to $n \times n$ integer matrices u with $|\det u| = 2$. For example, let u be chosen to be the *quincunx matrix* $q = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$, representing a counterclockwise rotation by $\pi/4$ multiplied by $\sqrt{2}$. We do encounter an “unexpected” fact if we try to find a Haar-type wavelet. There exists a set $D \subset \mathbb{R}^2$ such that χ_D is a scaling function for an MRA (defined as the obvious two-dimensional analogue of the of the one-dimensional MRA) that produces a Haar-type wavelet as the difference of two disjoint sets. Yet, these sets are rather complicated fractal sets known as the *twin dragons* (see Figure 1), as was observed in [14].

These observations indicate that the general construction of two-dimensional wavelets is significantly more complicated than the one-dimensional case. In particular, it is not known whether there exist continuous compactly supported ON wavelets analogues of the one-dimensional Daubechies wavelets associated with the dilation matrix q. However, it turns out that a rather simple change in the definition of the dilations in (3.8) produces much simpler constructions of Haar-type wavelets in dimension two.

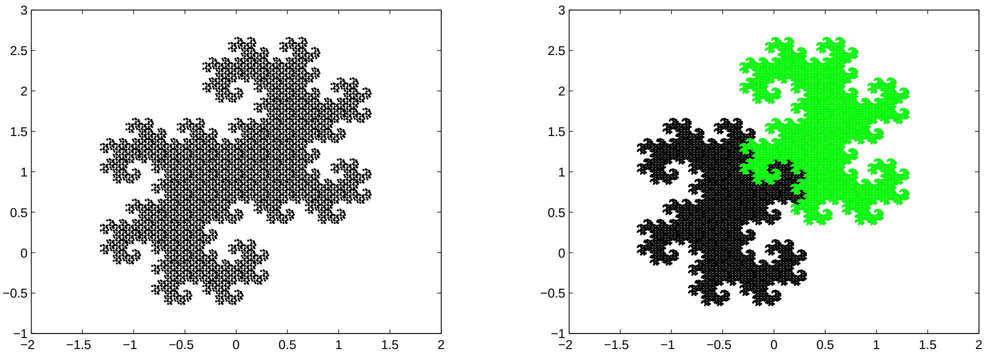

Figure 1: On the left is the fractal set known as the "Twin Dragon", whose characteristic function is the scaling function for the 2-dimensional Haar-type wavelet associated with the dilation matrix q. On the right we see the support of the resulting wavelet ψ, whose values are 1 on the darker set, -1 on the lighter set and 0 elsewhere.

Essentially, the idea consists in adding an additional set of dilations to the ones produced by the integer powers of the quincunx matrix. Specifically, let B be the group of the eight symmetries of the square, given by $B = \{b_j : j = 0, 1, \ldots, 7\}$, where $b_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $b_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $b_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $b_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, and $b_j = -b_{j-4}$, for $j = 4, \ldots, 7$. Let R_0 be the triangle with vertices $(0,0)$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$ and $R_i = b_i\, R_0$ for $i = 0, \ldots, 7$ (see Figure 2). Let $\phi = 2^{3/4}\chi_{R_0}$ and V_0 the closed linear span of $\{D_b\, T_k\, \phi : b \in B, k \in \mathbb{Z}^2\}$. Note that $|\det b| = 1$ for all $b \in B$ and V_0 is the subspace of $L^2(\mathbb{R}^2)$ of square integrable functions which are constant on each $\mathbb{Z}^2$-translate of the triangles R_i, $i = 0, \ldots, 7$. It is not difficult to show that there is a structure very similar to the classical MRA consisting of the spaces $V_j = D_q^j\, V_0$, $j \in \mathbb{Z}$. In fact, let us define the vector-valued function

$$\Phi = \begin{pmatrix} D_{b_0}\phi \\ \vdots \\ D_{b_7}\phi \end{pmatrix} = \begin{pmatrix} \phi^0 \\ \vdots \\ \phi^7 \end{pmatrix}.$$

Then $\{V_j : j \in \mathbb{Z}\}$ is an MRA with a vector-valued scaling function Φ. To derive a Haar-like wavelet, we observe that $R_0 = q^{-1}R_1 \cup q^{-1}(R_6 + \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right))$ (see Figure 2). This equality implies that

$$\phi(x) = \phi^0(x) = \phi^1(x) + \phi^6(qx - \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)).$$

Applying D_{b_j}, $j = 1, \ldots, 7$ to the expression above, we obtain similar equalities for ϕ^j, $j = 1, \ldots, 7$. Now, let $\psi(x) = \phi^1(qx) - \phi^6(qx - \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right))$. This function is the difference of appropriately normalized characteristic functions of disjoint triangles and leads indeed to the desired Haar-like wavelet. In fact, we can define the vector-valued function

$$\Psi = \begin{pmatrix} D_{b_0}\psi \\ \vdots \\ D_{b_7}\psi \end{pmatrix} = \begin{pmatrix} \psi^0 \\ \vdots \\ \psi^7 \end{pmatrix},$$

and observe that the system $\{D_q^j\, T_k\, \Psi : j \in \mathbb{Z}, k \in \mathbb{Z}^2\}$ is an ON basis of $L^2(\mathbb{R}^2)$.

The construction above is representative of a much more general situation. For example, let $u = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}$, the matrix of counterclockwise rotation by $\pi/3$, normalized to produce $\det u = 2$. Also in this case, there is a fractal Haar wavelet associated with this dilation matrix, but the introduction of

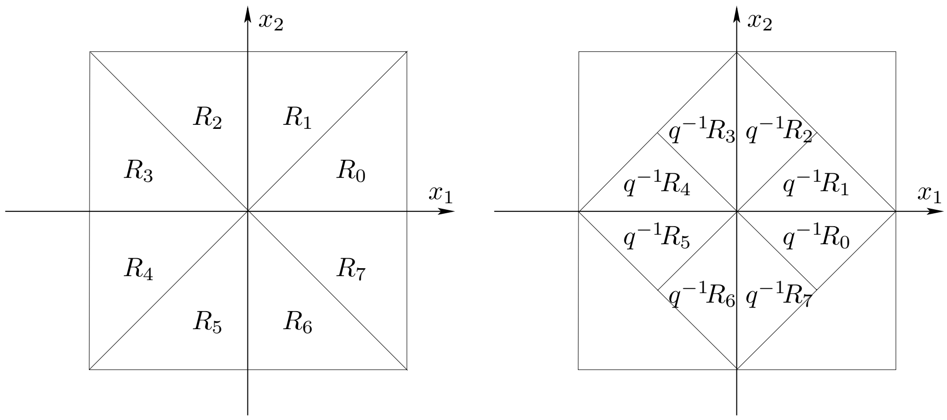

Figure 2: Example of construction of a wavelet system with composite dilations. On the left is illustrated the triangle R_0 and the triangles R_i, $i = 1, \dots 7$, obtained as $b_i R_0$ (the matrices b_i are described in the text). On the right is illustrated the action of the inverse of the quincunx matrix q on the triangles R_i.

an appropriate additional finite group of dilations allows one to derive simpler Haar-like wavelets similar to what we did above. See [18] for details.

In [18], we have shown that all this can be formalized by introducing the notion of *wavelet systems with composite dilations*, which are the systems of the form

$$\{D_a\, D_b\, T_k\, \psi : \, j \in \mathbb{Z}, a \in A, b \in B, k \in \mathbb{Z}^2\}, \tag{3.9}$$

where B is a group of matrices with determinant of absolute value 1 (as in the examples above) and A is a group of expanding matrices, in the sense that all eigenvalue have magnitude larger than one (as in the example above where A is the group of the integer powers of the quincunx matrix). Many different groups of B-dilations have been considered in the literature, such as the crystallographic and shear groups, which are associated to very different properties. As we will see below, one special benefit of this framework is the ability to produce wavelet-like systems with geometric properties going far beyond traditional wavelets. For example, one can construct waveforms whose supports are highly anisotropic and that are ranging not only over various scales and locations, but also over various orientations, making these functions particularly useful in image processing applications.

As discussed in Sec. 2, one of the most important properties of wavelets in $L^2(\mathbb{R})$ is their ability to provide rapidly convergent approximations for piecewise smooth functions. This property implies that wavelet expansion are useful to *compress* functions efficiently since - as described by the nonlinear approximation error estimate (2.5) - most of the L^2-norm of the function can be recovered from a relatively small number of expansion coefficients and not much information is lost by discarding the remaining ones. This idea is the basis for the construction of several wavelet-based data compression algorithm, such as JPEG2000 [27].

There is another important perhaps less obvious implication of the wavelet approximation properties and which has to do with the classical problem of data denoising. Suppose that we want to recover a function f which is corrupted by zero-mean white Gaussian noise with variance σ^2. Let f_n denote the noisy function. In this case, Donoho and Johnstone [12] have shown that there is a very simple and very effective procedure for estimating f. This consists, essentially, in (i) computing the wavelet expansion of f_n, (ii) setting to zero the wavelet coefficients of f_n whose magnitudes are below a fixed value (which depends on σ), and (iii) computing an estimator of f as a reconstruction from the wavelet coefficients of f_n which have not been set to zero. It turns out that the performance of this procedure depends directly on the decay rate of the nonlinear approximation error estimate (2.5).

The above observations underline the fundamental importance of the approximation properties of one-dimensional wavelets for applications. Unfortunately, as will be discussed in more details below, the situation is different in higher dimensions, where the standard multidimensional generalization of dyadic wavelets does not lead to the same type of approximation properties as in the one-dimensional case.

Let us restrict our attention to dimension $n = 2$ (the cases $n > 2$ are similar). Recall that, in dimension $n = 1$, to achieve the desired approximation properties of the wavelet expansions, we required wavelets having compact support and sufficiently many vanishing moments. In dimension $n = 2$, we can easily construct a two-dimensional dyadic ON wavelet system starting from a one-dimensional MRA with scaling function ϕ_1 and wavelet ψ_1, as follows. We define three wavelets $\psi^{(1)}(x_1, x_2) = \phi_1(x_1)\psi_1(x_2)$, $\psi^{(2)}(x_1, x_2) = \psi_1(x_1)\phi_1(x_2)$, $\psi^{(3)}(x_1, x_2) = \psi_1(x_1)\psi_1(x_2)$. Then the system

$$\{\psi^{(\ell)}_{j,k_1,k_2}(x_1, x_2) = 2^j \psi^{(\ell)}(2^j(x_1 - k_1), 2^j(x_2 - k_2)) : j, k_1, k_2 \in \mathbb{Z}, \ell = 1, 2, 3\}$$

is an ON basis for $L^2(\mathbb{R}^2)$. This is called a *separable* MRA wavelet basis. Clearly, we can choose ψ_1 and ϕ_1 having compact support and R vanishing moments. In this case, similar to the one-dimensional result, if $f \in C^R(\mathbb{R}^2)$ and $\|f\|_{C^R} < \infty$, then the m-th largest wavelet coefficient in magnitude satisfies the estimate:

$$|\langle f, \psi_{\pi(m)}\rangle| \le C\, \|f\|_{C^R}\, m^{-(R+1)/2}, \tag{3.10}$$

where C is a constant independent of f and m. This implies that, letting f_N be the best N-term nonlinear approximation of f, the *nonlinear approximation error* decays as

$$\|f - f_N\|^2_{L^2} \le C \sum_{m>N} |\langle f, \psi_{\pi(m)}\rangle|^2 \le C\, \|f\|^2_{C^R}\, N^{-R}.$$

However, while in the one-dimensional case the approximation properties of the wavelet expansion are not affected if one allows the function f to have a finite number of isolated singularities, the situation now is very different. Let us examine, for example, the wavelet approximation of the function $f = \chi_D$, where $D \subset \mathbb{R}^2$ is a compact set whose boundary has finite length L. Let us consider in particular the wavelet coefficients associated with the boundary of the region D. Since f is bounded, these wavelet coefficients have size $|\langle f, \psi^{(\ell)}_{j,k_1,k_2}\rangle| \sim 2^{-j}$. In addition, since the boundary of D has finite length and the wavelets have compact support, at each scale j, there are about $L\, 2^j$ wavelets with support overlapping the boundary of D. It follows from these observations that the N-th largest wavelet coefficient in this class is bounded by $C(L)N^{-1}$ and this implies that the nonlinear approximation error rate is only of the order $O(N^{-1})$. Hence, there is a large number of significant wavelet coefficients associated with the edge discontinuity of f and this is limiting the wavelet approximation rate.

The reason for this limitation of two-dimensional separable MRA wavelet bases is the fact that their supports are isotropic (they are supported on a box of size $\sim 2^{-j} \times 2^{-j}$) so that there are 'many' wavelets overlapping the edge singularity and producing significant expansion coefficients. To address this issue and produce better approximations of piecewise smooth multidimensional functions, one has to consider alternative multiscale systems which are more flexible at representing anisotropic features. Several constructions have been introduced starting with the wedgelets [3] and ridgelets [5]. Among the most successful constructions proposed in the literature, the curvelets [6] and shearlets [15] achieve this additional flexibility by defining a collection of analyzing functions ranging not only over various scales and locations, like traditional wavelets, but also over various orientations and with highly anisotropic supports. As a result, these systems are able to produce (nonlinear) approximations of two-dimensional piecewise C^2 functions for which the nonlinear approximation error decays essentially like N^{-2}, that is, as if the functions had no discontinuities. To give a better insight into this approach and show how the wavelet machinery can be modified to obtain these types of systems, we will briefly describe the shearlet construction in dimension $n = 2$, which is closely related to the framework of wavelets with composite dilations, which we mentioned above. To keep the presentation self-contained, we will only sketch the main ideas and refer the reader to [15, 17] for more details.

For an appropriate function $\psi \in L^2(\mathbb{R}^2)$, a *system of shearlets* is a collection of functions of the form (3.9), where

$$a = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{3.11}$$

Note that a is a dilation matrix whose integer powers produce anisotropic dilations and, more precisely, *parabolic scaling dilations* since the dilation factors grow quadratically in one coordinate with respect to the other one; the *shear matrix* b is non-expanding and, as we will see below, its integer powers control the orientations of the elements of the shearlet system. The generator ψ of the system is defined in the frequency domain as

$$\hat{\psi}(\xi) = \hat{\psi}(\xi_1, \xi_2) = w(\xi_1)\, v(\frac{\xi_2}{\xi_1}),$$

where $w,\ v \in C^\infty(\mathbb{R})$, $\operatorname{supp} w \subset [-\frac{1}{2}, \frac{1}{2}] \setminus [-\frac{1}{16}, \frac{1}{16}]$ and $\operatorname{supp} v \subset [-1, 1]$. Furthermore, it is possible to choose the functions w, v so that the corresponding system (3.9) is a PF (Parseval frame) for $L^2(\mathbb{R}^2)$. The geometrical properties of the shearlet system are more evident in the Fourier domain. In fact, a direct calculation gives that

$$\hat{\psi}_{j,\ell,k}(\xi) := (D_a^j\, D_b^\ell\, T_k\, \psi)^\wedge(\xi) = 2^{-3j/2}\, w(2^{-2j}\xi_1)\, v(2^j \tfrac{\xi_2}{\xi_1} - \ell)\, e^{2\pi i \xi a^{-j} b^{-\ell} k}, \tag{3.12}$$

implying that the functions $\hat{\psi}_{j,\ell,k}$ are supported in the trapezoidal regions

$$\{(\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_1 \in [-2^{2j-1}, 2^{2j-1}] \setminus [-2^{2j-4}, 2^{2j-4}], |\frac{\xi_2}{\xi_1} - \ell 2^{-j}| \le 2^{-j}\}.$$

The last expression shows that the frequency supports of the elements of the shearlet system are increasingly more elongated at fine scales (as $j \to \infty$) and orientable, with orientation controlled by the index ℓ (this is illustrated in Fig. 3). These properties show that shearlets are much more flexible than "isotropic" wavelets and explain why shearlets can achieve better approximation properties for functions which are piecewise smooth. Similar properties hold for the curvelets and can be extended to higher dimensions.

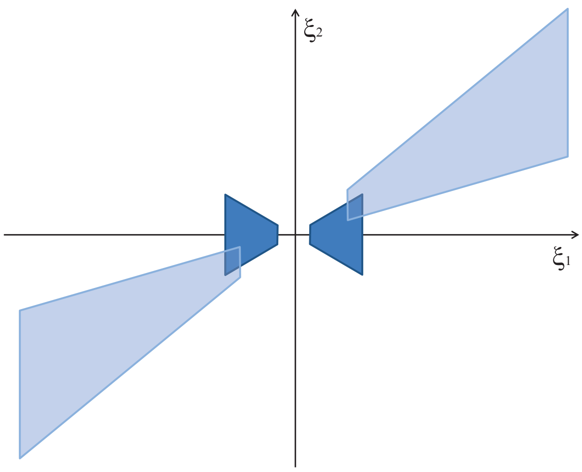

Figure 3: The frequency supports of the elements of the shearlet system are pairs of trapezoidal regions defined at various scales and orientations, dependent on j and ℓ respectively. The figure shows the frequency supports of 2 representative shearlet elements: the darker region corresponds to $j = 0$, $\ell = 0$ and the lighter region to $j = 1$, $\ell = 1$.

As indicated above, shearlets and curvelets are only some of the methods introduce during the last decade to extend or generalize the wavelet approach. We also recall the construction of *bandelets* [25]; these systems achieve improved approximations for functions which are piecewise C^α (α may be larger than 2) by using an adaptive construction. We refer to the volume [24] for the description of several such systems.

4 Continuous wavelets

In this section, we examine *continuous wavelets* on $\mathbb{R}^n$. The general linear group $GL(n,\mathbb{R})$ of $n\times n$ invertible real matrices acts on $\mathbb{R}$ by linear transformations. The semi-direct product G of $GL(n,\mathbb{R})$ and $\mathbb{R}^n$ is called the *general affine group* on $\mathbb{R}^n$ since each invertible affine map on $\mathbb{R}^n$ has the form $(a,t)\cdot x = a(x+t) = ax+at$ for a unique $(a,t)\in G$. Thus, the group law $(a,t)\cdot(b,s) = (ab, b^{-1}t+s)$ for G corresponds to the composition of the associated affine maps and $(a,t)^{-1} = (a^{-1}, -a^{-1}t)$. We have a unitary representation τ of G acting on $L^2(\mathbb{R}^n)$ defined by

$$(\tau_{(a,t)}\psi)(x) = |\det a|^{-\frac{1}{2}}\psi((a,t)^{-1}\cdot x) = |\det a|^{-\frac{1}{2}}\psi(a^{-1}x - t) := \psi_{a,t}(x).$$

We then have that

$$(\tau_{(a,t)}\psi)^\wedge(\xi) = |\det a|^{1/2}\hat{\psi}(a^*\xi)\, e^{-2\pi i\xi\cdot at}.$$

For $\psi\in L^2(\mathbb{R}^n)$, the *continuous wavelet transform* W_ψ associated with ψ and G is defined by

$$(W_\psi f)(a,t) = \langle f, \psi_{(a,t)}\rangle = |\det a|^{-\frac{1}{2}}\int_{\mathbb{R}^n} f(x)\,\overline{\psi(a^{-1}x - t)}\,dx, \tag{4.13}$$

and it maps $f\in L^2(\mathbb{R}^n)$ into a space of functions on G. For D a closed subgroup of $GL(n,\mathbb{R})$, $H = \{(a,t) : a\in D, t\in\mathbb{R}^n\}$ is a closed subgroup of G and the left Haar measures on H are the product measures $d\lambda(a,t) = d\mu(a)\,dt$, where μ is a left Haar measure on D. In the special case where D is a discrete subgroup of $GL(n,\mathbb{R})$ such as $\{u^j : j\in\mathbb{Z}\}$, for some $u\in GL(n,\mathbb{R})$, then we can take μ to be the counting measure on D.

We then seek conditions on D and μ for which restricting $W_\psi f$ to H gives an isometry from $L^2(\mathbb{R}^n)$ into $L^2(H, d\lambda)$ or, equivalently, we have a continuous reproducing formula

$$f = \int_H \langle f, \tau_{(a,t)}\psi\rangle\, \tau_{(a,t)}\psi\, d\lambda(a,t),$$

for each $f\in L^2(\mathbb{R}^n)$. As shown in [4], this holds if and only if

$$\int_D |\hat{\psi}(a^*\xi)|^2\, d\mu(a) = 1, \quad \text{for a.e. } \xi\in\mathbb{R}^n, \tag{4.14}$$

in which case ψ is a *continuous wavelet* (with respect to D). In the special case $D = \{aI_n : a>0\}$, where I_n is the $n\times n$ identity matrix, and $d\mu(aI_n) = \frac{da}{a}$, the expression (4.14) reduces to the classical *Calderón condition*:

$$\int_0^\infty |\hat{\psi}(a\xi)|^2\,\frac{da}{a} = 1, \quad \text{for a.e. } \xi\in\mathbb{R}^n. \tag{4.15}$$

Clearly, the set

$$\{f\in L^2(\mathbb{R}^n) : \ c\,f \text{ satisfies (4.15) for some } c>0\}$$

is dense in $L^2(\mathbb{R}^n)$. When $D = \{u^j : j\in\mathbb{Z}\}$, for $u>1$, one can show that if $\{\psi_{(u^j,k)} : j\in\mathbb{Z}, k\in\mathbb{Z}\}$ is a Parseval frame, then ψ ia also a continuous wavelet with respect to D. Conversely, (4.15) is only necessary but not sufficient for $\{\psi_{(u^j,k)} : j\in\mathbb{Z}, k\in\mathbb{Z}\}$ to be a Parseval frame.

For $\psi\in L^2(\mathbb{R}^n)$, the *modified continuous wavelet transform* $\widetilde{W}_\psi$ associated with ψ and G is defined by

$$(\widetilde{W}_\psi f)(a,t) = (W_\psi f)(a, a^{-1}t) = |\det a|^{-\frac{1}{2}}\int_{\mathbb{R}^n} f(x)\,\overline{\psi(a^{-1}(x-t))}\,dx, \tag{4.16}$$

and it also maps $f\in L^2(\mathbb{R}^n)$ into functions on $G = D\times\mathbb{R}^n$. The modification arises from using the analyzing function $|\det a|^{-\frac{1}{2}}\psi(a^{-1}(x-t))$ rather than $|\det a|^{-\frac{1}{2}}\psi(a^{-1}x-t)$ and it simplifies the problem of estimating the asymptotic decay properties of the continuous wavelet transform.

Indeed for $G = D \times \mathbb{R}^n$, where $D = \{aI_n : a > 0\}$, let us consider the modified continuous wavelet transform

$$(\widetilde{W}_\psi f)(a,t) := (\widetilde{W}_\psi f)(aI_n,t) = a^{-n/2} \int_{\mathbb{R}^n} f(x)\, \overline{\psi(a^{-1}(x-t))}\, dy. \tag{4.17}$$

A fundamental property of this transform is its ability to characterize the local regularity of functions. For example, let f be a bounded function on $\mathbb{R}$ which is Hölder continuous at x_0, with exponent $\alpha \in (0,1]$, that is, there is $C > 0$ for which

$$|f(x_0+h) - f(x_0)| \le C|h|^\alpha.$$

Suppose that $\int_{\mathbb{R}} (1+|x|)|\psi(x)|\, dx < \infty$ and that $\hat{\psi}(0) = 0$. Since the last condition implies that $\int_{\mathbb{R}} \psi(x)\, dx = 0$, then

$$(\widetilde{W}_\psi f)(a,t) = a^{-1/2} \int_{\mathbb{R}} (f(x) - f(x_0))\, \overline{\psi(a^{-1}(x-t))}\, dx.$$

Thus, using the Hölder continuity and a change of variables, we have:

$$\begin{aligned} |(\widetilde{W}_\psi f)(a,t)| &\le a^{-1/2} \int_{\mathbb{R}} |f(x) - f(x_0)|\, |\psi(a^{-1}(x-t))|\, dx && (4.18)\\ &\le C\, a^{\alpha+1/2} \int_{\mathbb{R}} |y + a^{-1}(t-x_0)|^\alpha\, |\psi(y)|\, dy. && (4.19) \end{aligned}$$

This shows that, at $t = x_0$, the continuous wavelet transform of f decays (at least) like $a^{\alpha+1/2}$, as $a \to 0$. Under slightly stronger condition on ψ, one can show that the converse also holds, hence providing a characterization result. It is also possible to extend this analysis to discontinuous functions and even distributions. For example, if f has a jump discontinuity at x_0, then one can show that the continuous wavelet transform of f decays like $a^{1/2}$, as $a \to 0$, and similar properties hold in higher dimensions (cf. [22]).

While the continuous wavelet transform (4.17) is able to describe the local regularity of functions and distribution and detect the location of singularity points through its decay at fine scales, it does not provide additional information about the *geometry* of the set of singularities. In order to achieve this additional capability, one has to consider wavelet transforms associated with more general dilation groups.

For example, in dimension $n = 2$, let M be the subgroup of $GL(2,\mathbb{R})$ of the matrices

$$\left\{ m_{a,s} = \begin{pmatrix} a & -a^{1/2}\, s \\ 0 & a^{1/2} \end{pmatrix} : a > 0, s \in \mathbb{R} \right\},$$

and let us consider the corresponding generalized continuous wavelet transform

$$(\widetilde{W}_\psi f)(a,s,t) := (\widetilde{W}_\psi f)(m_{a,s},t) = a^{-3/4} \int_{\mathbb{R}^2} f(x)\, \overline{\psi(m_{a,s}^{-1}(x-t))}\, dx, \tag{4.20}$$

where $a > 0$, $s \in \mathbb{R}$ and $t \in \mathbb{R}^2$. It is easy to verify that we have the factorization $m_{a,s} = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{1/2} \end{pmatrix}$, that is, $m_{a,s}$ is the product of an anisotropic dilation matrix and a shear matrix. As a result, the analyzing function $\psi_{a,s,t} = a^{-3/4}\psi(m_{a,s}^{-1}(x-t))$ associated with this transform range over various scales, orientations and locations, controlled by the variables a, s, t, respectively. This is similar to the discrete shearlets in Section 3. The transform $(\widetilde{W}_\psi f)(a,s,t)$ is called the *continuous shearlet transform* of f.

Thanks to the properties associated with dilation group M, the continuous shearlet transform is able to detect not only the location of singularity points through its decay at fine scales, but also the geometric information of the singularity set. In particular, there is a general characterization of step discontinuities along 2D piecewise smooth curves, which can summarized as follows [16]. Let $B = \chi_S$, where $S \subset \mathbb{R}^2$ and its boundary ∂S is a piecewise smooth curve.

- If $t \notin \partial S$, then $\widetilde{W}_\psi B(a,s,t)$ has rapid asymptotic decay, as $a \to 0$, for each $s \in \mathbb{R}$. That is,

$$\lim_{a \to 0} a^{-N} \widetilde{W}_\psi B(a,s,t) = 0, \qquad \text{for all } N > 0.$$

- If $t \in \partial S$ and ∂S is smooth near t, then $\widetilde{W}_\psi B(a,s,t)$ has rapid asymptotic decay, as $a \to 0$, for each $s \in \mathbb{R}$ unless $s = \tan\theta_0$ and $(\cos\theta_0, \sin\theta_0)$ is the normal orientation to ∂S at t. In this last case, $\widetilde{W}_\psi B(a,s_0,t) \sim a^{\frac{3}{4}}$, as $a \to 0$. That is,

$$\lim_{a\to 0} a^{-\frac{3}{4}} \widetilde{W}_\psi B(a,s,t) = C \neq 0.$$

- If t is a corner point of ∂S and $s = \tan\theta_0$ where $(\cos\theta_0, \sin\theta_0)$ is one of the normal orientations to ∂S at t, then $\widetilde{W}_\psi B(a,s_0,t) \sim a^{\frac{3}{4}}$, as $a \to 0$. For all other orientations, the asymptotic decay of $\widetilde{W}_\psi B(a,s,t)$ is faster and depends in a complicated way on the curvature of the boundary ∂S near t [16].

Similar results hold in higher dimensions and for other types of singularity sets.

Note that, in the definition of $\widetilde{W}_\psi B(a,s,t)$, we are taking the inner product of f with the *continuous* shearlets $T_t D^{-1}_{m_{as}}$. On the other hand, the discrete shearlets in Section 3 involve the reverse order of operators. Despite this fact, the decay properties of the continuous shearlet transform are related to the approximation properties of discrete shearlets.

5 Various other "wavelet topics", applications and conclusions

The number of researchers who have worked and are working on wavelets is very large. This field and its applications are enormous. We could not cover all the topics that are most important and interesting in such a short article. We make no claim that we have chosen to cover all "the most important" topics on wavelets. In this article, we have defined wavelets to be elements of the Hilbert space $L^2(\mathbb{R}^n)$. In fact, we can apply the wavelet techniques we described to the Banach spaces $L^p(\mathbb{R}^n)$, $p \geq 1$. As usual, the flavor for $1 \leq p < 2$ is very different from $p > 1$. The roles played by $\mathbb{R}$ and the 1-torus $\mathbb{T}$ were shown to extend to higher dimensions. It is well known that the harmonic analysis involving $(\mathbb{Z}, \mathbb{T})$ extends to the setting $(G, \widehat{G})$ where G is a locally compact Abelian group and $\widehat{G}$ its dual. Wavelet theory extends to $(G, \widehat{G})$ and other abstract settings.

Wavelets continue to stimulate and inspire active research going beyond the area of harmonic analysis where they were originally introduced. While during the 1980s and 1990s most of wavelet theory was devoted to the construction of "nice" wavelet bases and their applications to denoising and compression, during the last decade wavelet research was focused more on the subject of approximations and the so-called *sparse* approximations. As we mentioned above, several generalizations and extensions of wavelets were introduced with the goal to provide improved approximations properties for special classes of functions where the more traditional wavelet approach is not as effective. This research has stimulated the investigation of redundant function systems (that is, frames which are not necessarily tight) and their applications using techniques coming not only from harmonic analysis but also from approximation theory and probability. The emerging area of compressed sensing, for example, can be seen as a method for achieving the same nonlinear approximation properties of wavelets and their generalization by using linear measurements defined according to a certain clever strategy.

Some fundamental ideas from wavelet theory, most notably the multiresolution analysis, have appeared in other forms in very different contexts. For example, the theory of diffusion wavelets [11] provide a method for the multiscale analysis of manifolds, graphs and point clouds in Euclidean space. Rather than using dilations as in the classical wavelet theory, this approach uses "diffusion operators" acting on functions on the space. For example, let T be a diffusion operator (e.g. the heat operator) acting on a graph (the graph can be a discretization of a manifold). The study of the eigenfunctions and eigenvalues of T is known as Spectral Graph Theory and can be viewed a generalization of the theory of Fourier series on the torus. The main idea of diffusion wavelets is to compute dyadic powers of the operator T to establish a scale for performing multiresolution analysis on the graph. This approach has many useful applications, since it allows one to apply the advantages of multiresolution analysis to objects that can be modeled as graphs, such as chemical structures, social networks, etc.

In order to describe the more recent applications inspired by wavelets, we quote Coifman from [9, p. 159] "Over the last twenty years we have seen the introduction of adaptive computational analytic tools that enable flexible transcriptions of the physical world. These tools enable orchestration of signals into constituents (mathematical musical scores) and opened doors to a variety of digital implementations/applications in engineering and science. Of course I am referring to wavelet and various versions of computational Harmonic Analysis. The main concepts underlying these ideas involved the adaptation of Fourier analysis to changing geometries as well as multiscale structures of natural data. As such, these methodologies seem to be limited to analyze and process physical data alone. Remarkably, the last few years have seen an explosion of activity in machine learning, data analysis and search, implying that similar ideas and concepts, inspired by signal processing might carry as much power in the context of the orchestration of massive high dimensional data sets. This digital data, e.g., text documents, medical records, music, sensor data, financial data etc., can be structured into geometries that result in new organizations of language and knowledge building. In these structures, the conventional hierarchical ontology building paradigm, merges with a blend of Harmonic Analysis and combinatorial geometry. Conceptually these tools enable the integration of local association models into global structures in much the same way that calculus enables the recovery of a global function from a local linear model for its variation. As it turns out, such extensions of differential calculus into the digital data fields are now possible and open the door to the usage of mathematics similar in scope to the Newtonian revolution in the physical sciences. Specifically we see these tools as engendering the field of mathematical learning in which raw data viewed as clouds of points in high dimensional parameter space is organized geometrically much the same way as in our memory, simultaneously organizing and linking associated events, as well as building a descriptive language."

Let us illustrate three examples that will give the reader a more concrete idea of what Coifman asserts (see also [10]).

(a) In the oil exploration and mining industry, one needs to decide where to drill or mine to greatest advantage for finding oil, gas, copper or other minerals. This involves an analysis of the composition and structure of the soil in a certain region. From such analysis one would find properties that optimize where these resources are most likely to be found.

(b) Suppose that we would like to decompose a large collection of books into subclasses of "similar" books, e.g., novels, histories, physics books, mathematical books, etc. Possibly we also want to assign "distances" between these subclasses. It is not unreasonable that the distributions of many particular words contained in each book can identify various kind of books so that they can be assigned in a specific subclass.

(c) In medical diagnostics, important information can be gleaned from the analysis of data obtained from radiological, histological, chemical tests and this is important for arriving at an early detection of potentially dangerous tumors and other pathologies.

What is surprising is that the analysis of these very different types of data can be performed very efficiently using the type of "mathematics" based on the ideas presented in this paper. For example, several hospitals have adopted medical diagnostics methods that were developed by Coifman and his group using these ideas.

In conclusion, we want to stress that "wavelets" is a huge field. Many have helped to create it. We want to state, however, that Yves Meyer has contributed and introduced many ideas that were most important in its creation. The material in the first chapter of [21] describes many constructions which are due to him and the ideas that paved the way for many of the topics (e.g., the MRA) we presented in this paper.

References

[1] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **41**(7) (1988), 909–996.

[2] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[3] D. L. Donoho, *Wedgelets: Nearly-minimax estimation of edges*, Ann. Statist. **27** (1999), 859–897.

[4] A. P. Calderón, *Intermediate spaces and interpolation, the complex method*, Studia Math. **24** (1964), 113–190.

[5] E. J. Candès and D. L. Donoho, *Ridgelets: the key to high dimensional intermittency?*, Philos. Trans. R. Soc. Lond. Ser. A **357** (1999), 2495–2509.

[6] E. J. Candès and D. L. Donoho, *New tight frames of curvelets and optimal representations of objects with C^2 singularities*, Comm. Pure Appl. Math. **57** (2004), 219–266.

[7] P. G. Casazza, O. Christensen and N. J. Kalton, *Frames of translates*, Collect. Math. **52** (2001), 35–54.

[8] A. Cohen, I. Daubechies and J.C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **45** (5) (1992), 485–560.

[9] J. Cohen and A. I. Zayed, (Editors), *Wavelets and Multiscale Analysis: Theory and Applications*, Birkäuser, 2011.

[10] R. R. Coifman and M. Gavish, *Harmonic analysis of digital data bases* in [9], 161–197.

[11] R. R. Coifman and M. Maggioni, *Diffusion wavelets*, Appl. Comp. Harm. Anal. **21**(1) (2006), 53-94.

[12] D. L. Donoho and I. M. Johnstone. *Ideal spatial adaptation by wavelet shrinkage*, Biometrika **81**(3) (1994), 425-455.

[13] G. Garrigos, E. Hernández, H. Sikic, F. Soria, G. Weiss and E.N. Wilson, *Connectivity in the Set of Tight Frame Wavelets*, Glas. Mat. **38**(1) (2003), 75–98.

[14] K. Gröchenig and W. R. Madych, *Multiresolution analysis, Haar bases, and self-similar tilings of $\mathbb{R}^n$*, IEEE Trans. Info. Theory **38**(2) (1992), 556–568.

[15] K. Guo and D. Labate, *Optimally Sparse Multidimensional Representation using Shearlets*, SIAM J. Math. Anal. **9** (2007), 298–318

[16] K. Guo and D. Labate, *Characterization and analysis of edges using the continuous shearlet transform*, SIAM J. Imag. Sci. **2** (2009), 959–986.

[17] K. Guo and D. Labate, *Optimally sparse representations of 3D Data with C^2 surface singularities using Parseval frames of shearlets*, SIAM J Math. Anal. **44** (2012), 851–886.

[18] K. Guo, W-Q. Lim, D. Labate, G. Weiss and E. Wilson, *Wavelets with composite dilations and their MRA properties*, Appl. Comput. Harmon. Anal. **20** (2006), 231–249.

[19] A. Haar, *Zur Theorie der orthogonalen Funktionensysteme*, Math. Ann. **69**(3) (1910), 331–371.

[20] E. Hernández, H. Sikic, G. Weiss and E. Wilson, *On the properties of the integer translates of a square integrable function*, Contemp. Math. **505** (2010), 233–249.

[21] E. Hernández and G. Weiss, *A First Course on Wavelets*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1996.

[22] M. Holschneider, *Wavelets. Analysis tool*, Oxford University Press, Oxford, 1995.

[23] S. G. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Trans. Patt. Anal. Mach. Intell. **11**(7) (1989), 674–693.

[24] S. Mallat, *A Wavelet Tour of Signal Processing.Third Edition: The Sparse Way*, Academic Press, San Diego, CA, 2008.

[25] S. Mallat and G. Peyré, *Orthogonal Bandlet Bases for Geometric Images Approximation*, Comm. Pure Appl. Math. **61**(9), 1173–1212.

[26] Y. Meyer, *Wavelets and operators*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1992.

[27] D. Taubman and M. Marcellin (Eds.), *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Springer, 2002.