The unreasonable effectiveness of Haar frames

Stéphane Jaffard, Hamid Krim

November 10, 2021

Abstract

Keywords: Haar basis, frames, wavelets, Hölder global and pointwise regularity, machine learning.

1 Introduction

When the first author started his PhD in 1986, wavelets had just been introduced; among the very few papers available on the subject stood *painless nonorthogonal expansions* by I. Daubechies, A. Grossmann and Y. Meyer [11], in which wavelet frames were introduced. These constructions were considered as a compromise between the continuous wavelet transform, which is a very flexible tool whose shape constraints are easily met, and orthonormal wavelet bases whose use is computationally less costly. The notion of a frame was not new: it had been introduced by R. J. Duffin and A. C. Schaeffer in 1952 as redundant systems which share most of the "good" stability properties displayed by orthonormal bases [16]. Recall that, if H is an Hilbert space, a frame is a sequence $(e_n)_{n\in\mathbb{N}}$ of element of H satisfying

$$\exists C, C' > 0, \quad \forall f \in H, \qquad C \parallel f \parallel^2 \leq \sum_n |\langle f|e_n\rangle|^2 \leq C' \parallel f \parallel^2 . \qquad (1)$$

This condition implies that the vectors (e_n) span E, and f can be reconstructed from the e_n in a stable way: There exists a *dual frame* $(g_n)_{n\in\mathbb{N}}$ such that the partial sums $\sum_{n\leq N}\langle f|g_n\rangle e_n$ converge to f in H when $N \to +\infty$. The article of Duffin and Shaeffer however had remained largely unnoticed (except for the notable exception of the book of R. Young on nonharmonic Fourier series [49] which exposed the basic properties of frames); in particular, this notion, though stated in a general Hilbert space setting, had not been used outside of this context. Strangely, for the first author, the immediate consequence of reading *painless nonorthogonal expansions* was to guide him to the beautiful paper of Duffin and Schaeffer, the consequence of which was an improvement on the conditions under which a family of complex exponentials $e^{i\lambda_n t}$ form a frame on $L^2([a,b])$, see [25]. Another consequence was an introduction to one paper by Alex Grosmann, which was not written in the tough language (at least for a young PhD student in mathematics) of theoretical physics. This quickly led to visits in Marseille to meet Alex, and to understand his views on the newborn field of wavelet analysis. The first meeting with Alex was disconcerting: one quickly realized that he was a very unorthodox scientist; his deep insights, often masked by distressingly simple statements and explanations, could easily be missed; however, once one became aware of this, and pondered and analyzed his words, especially the ones which sounded the most innocuous, then the magic worked, and opened the door to the extremely original world of a great scientist.

Painless nonorthogonal expansions quickly had a deep impact. With a 34 years delay, it played the seminal role that could have been played by the Duffin and Shaeffer paper: Properties of expansions on frames were investigated, see e.g. [4,7] and references therein for an account of very significant developments of frame theory, and its relevance, in both mathematical analysis and in signal and image processing. To illustrate this importance, we point out the necessity of using frames rather than bases, in the *phase retrieval problem*: Assume that a signal f (in practice the setting usually is image processing) is unknown, but the information that can be acquired is $|\langle g_n|f\rangle|$ for a particular set of functions g_n. What are the conditions on the g_n to recover $|f|$? If the g_n form an orthonormal basis, this is obviously impossible: all the $\sum \varepsilon_n \langle f|g_n\rangle g_n$ with $\varepsilon_n = \pm 1$ are possible solutions; therefore some representation redundancy is needed. We encourage the reader to check geometrically that, in $\mathbb{R}^2$, the beautiful frame supplied by three unit vectors with angles $2\pi/3$ between each other solves the problem.

Another important motivation for frames arises in settings where no orthonormal bases are available, as in time-frequency analysis: The famous Balian-Low theorem states that there exist no orthonormal bases of $L^2(\mathbb{R})$ of the form

$$g_{m,n}(t) = e^{2i\pi\alpha m t} g(t - \beta n), \qquad m, n, \in \mathbb{Z}, \tag{2}$$

where g is smooth and well localized, i.e. which satisfies

$$\int_{\mathbb{R}} x^2 |g(x)|^2 dx < \infty \quad \text{ and } \quad \int_{\mathbb{R}} \xi^2 |\hat{g}(\xi)|^2 d\xi < \infty. \tag{3}$$

For several years, time-frequency frames similar to those in (2) were therefore the only option to get around the Balian-Low obstruction, see e.g. [18]. Note that another option later appeared: Wilson and Malvar bases (roughly speaking, the complex exponentials in (2) are replaced by sines and cosines, and a proper choice of g, α and β yields orthonormal bases, see [9, 12, 39]). These, however, did not diminish the important role of frames in time-frequency analysis: The most adapted signal processing tool which first detected the gravitational wave was a *tight frame*, see Sec. 4, composed of the union of 16 Wilson orthonormal bases, where the generating function g is in the Schwartz class, and compactly supported in the Fourier domain, see [5, 41]. The use of a (very) redundant frame is motivated by the fact that a gravitational wave has a much more sparse representation in this system than when using one Wilson basis only. Our objective in this paper bears some similarity with this last point: We will show that a union of Haar bases has unexpected analysis properties which makes it fit for some Artificial Intelligence (AI) and learning questions. Some results proved in this paper have been announced in [28].

2 The Haar basis

Let φ be the characteristic function of $[0,1]$, which, more precisely, we define as

$$\begin{cases} \varphi(x) = 1 \quad \text{if } x \in (0,1), \\ \varphi(0) = \varphi(1) = 1/2, \\ \varphi(x) = 0 \quad \text{else;} \end{cases} \tag{4}$$

and let

$$\forall x \in \mathbb{R}, \qquad \psi(x) = \varphi(2x) - \varphi(2x-1); \tag{5}$$

This (slightly unusual) definition for φ (and hence ψ) is motivated by the fact that we will consider pointwise values of partial sums of Haar series, and it is important that every point is a Lebesgue point of these partial sums; hence the values chosen for φ at the end points of its support. The Haar basis on $\mathbb{R}$ is the orthonormal basis of $L^2(\mathbb{R})$ composed of the functions

$$\begin{cases} \varphi(x-k) \quad \text{for } k \in \mathbb{Z} \\ 2^{j/2}\psi(2^j x - k) \quad \text{for } j \geq 0 \text{ and } k \in \mathbb{Z}. \end{cases} \tag{6}$$

This system (or, more precisely, its restriction on $[0,1]$) was introduced by A. Haar in his PhD thesis in 1909 in order to answer a question raised by D. Hilbert; a distressing drawback of Fourier expansions had recently been uncovered: The Fourier series of a continuous function may diverge at some points. Is it the case for all orthonormal systems or is the trigonometric system pathological? A. Haar proved that, if f is a continuous function, then the partial sums of the Haar expansion of f converge uniformly. Ironically, an expansion using discontinuous "building blocks" behaves better in terms of regularity than the decomposition using the smooth trigonometric system; this counterintuitive achievement is the first "unreasonable effectiveness" of the Haar basis. Of course, we cannot expect too much from decompositions on this system: The fact that it is composed of discontinuous functions obviously prevents it from being a basis of function spaces composed of continuous functions: indeed, the partial sums of the reconstruction have to belong to the function space considered. To be more precise, we recall which function spaces the Haar system is a basis for. We first provide the appropriate definition,

Definition 1 *Let E be a separable Banach space. A sequence $(e_n)_{n\in\mathbb{N}}$ of elements of E is an unconditional basis if it satisfies the following requirements:*

1. *$\forall f \in E$, there exists a unique sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ such that the partial sums $\sum_{n\leq N} a_n e_n$ converge to f, i.e.*

$$\left\| \sum_{n=1}^{N} a_n e_n - f \right\|_E \longrightarrow 0 \quad \text{when} \quad N \to +\infty. \tag{7}$$

2. *There exists $C > 0$ such that, for any sequence of real numbers $(a_n)_{n\in\mathbb{N}}$, for any sequence (ε_n) such that $|\varepsilon_n| \leq 1$, then*

$$\left\| \sum \varepsilon_n a_n e_n \right\|_E \leq C \left\| \sum a_n e_n \right\|_E. \tag{8}$$

The second requirement insures the numerical stability of the reconstruction of f using linear combinations of the e_n. A key consequence of (8) is that, if $f = \sum a_n e_n$, then the norm of f in E is equivalent to a quantity built on the $(a_n)_{n\in\mathbb{N}}$, and which actually only depends on the $|a_n|$. This means that E is isomorphic to a sequence space. In statistics, this key property is often referred to as the *multiplier property*. Note that it is only one of the two ingredients for an unconditional basis and, in particular, spaces that are not separable may satisfy the multiplier property (though they do not have unconditional bases). It is e.g. the case for the Hölder C^α spaces if a smooth orthonormal wavelet basis is used, i.e. a basis of the form (6) where φ and ψ are sufficiently smooth and well localized (since the C^α spaces are not separable, they cannot have an unconditional basis). These topics have played a central role at the beginning of wavelet theory: The first wavelet basis (before the term "wavelet" was coined), was introduced by J. O. Strömberg precisely in order to construct unconditional bases for the real Hardy spaces H^p [43]; several chapters of the seminal book [40] of Y. Meyer are devoted to obtaining equivalent sequence norms for large classes of function spaces, and the multiplier property plays a key role for wavelet methods in statistics, see e.g. [15] and refs. therein. Note that it is very specific to wavelets, and does not hold for other "classical" bases. In particular, in the periodic case, it has been known for a long time that the L^p or C^α norms cannot be characterized by a quantity bearing on the moduli of the Fourier coefficients (except in the Hilbert case, i.e. for L^2).

Coming back to the Haar basis, the subtle problem of determining which function spaces the Haar system is an unconditional basis for, has been completely settled by G. Bourdaud in [2] for Besov spaces, the definition of which we now recall. To that end, we use a smooth wavelet basis so that no restriction on the indices of function spaces are required (one can take the functions φ and ψ below in the Schwartz class, as shown by Y. Meyer, see [35]). A wavelet basis on $\mathbb{R}$ is an orthonormal basis of $L^2(\mathbb{R})$ which has the same algorithmic structure as the Haar system (6), but using functions φ and ψ which can be smooth and well localized. The orthonormal basis requirement implies that any function $f \in L^2(\mathbb{R})$ can be written

$$f(x) = \sum_{k\in\mathbb{Z}} c_k \, \varphi(x-k) + \sum_{j\geq 0} \sum_{k\in\mathbb{Z}} c_{j,k} \, \psi(2^j x - k), \tag{9}$$

where the wavelet coefficients of f are

$$c_{j,k} = 2^j \int_{\mathbb{R}} f(x)\psi(2^j x - k)dx \quad \text{ and } \quad c_k = \int_{\mathbb{R}} f(x)\varphi(x-k)dx. \tag{10}$$

We will use the fact that convergence also holds pointwise: If $f \in L^1(\mathbb{R})$, then the partial sums of f converge almost everywhere, and in particular at Lebesgue points of f, see [44, 47] (note that, in the case of the Haar basis, this is a direct consequence of the fact that the partial reconstruction of f up to the scale j is the piecewise constant function on dyadic intervals of length 2^{-j}, which takes for value the average of f on that interval). One of the equivalent definitions of Besov spaces is given by the following requirement.

Definition 2 *Let $\alpha \in \mathbb{R}$ and $p, q \in (0, \infty]$. A tempered distribution f belongs to the Besov space $B^{\alpha,q}_p(\mathbb{R})$ if and only if its wavelet coefficients on a wavelet* *basis in the Schwartz class satisfy the two conditions:* $(c_k) \in \ell^p$ *and*

$$\sum_{j\in\mathbb{Z}} \left(\sum_{k\in\mathbb{Z}} \left(2^{(\alpha-1/p)j} |c_{j,k}| \right)^p \right)^{q/p} \leq C, \tag{11}$$

using the usual convention for ℓ^∞ *when* p *or* q *is infinite.*

In particular, the global Hölder spaces $C^\alpha(\mathbb{R}) = B^{\alpha,\infty}_\infty(\mathbb{R})$ (for any $\alpha \in \mathbb{R}$), sometimes referred to as *Lipschitz spaces*, are characterized by the condition

$$(c_k) \in \ell^\infty \quad \text{and} \quad \exists C, \ \forall j,k \quad |c_{j,k}| \leq C 2^{-\alpha j}. \tag{12}$$

The characterization supplied by (11) (resp. (12)) can roughly be interpreted as follows: The fractional derivative of f of order α belongs to L^p (resp. L^∞), see [40].

Bourdaud's theorem states that the Haar system is an unconditional basis of the Besov space $B^{\alpha,q}_p(\mathbb{R})$ (and (11) holds) if and only if

$$\frac{1}{p} - 1 < \alpha < \min\left(\frac{1}{p}, 1\right).$$

Note that Bourdaud's result is much more general, and applies to large classes of wavelet bases (and it is actually expressed in the slightly different setting of *homogeneous Besov spaces*, a subtlety that we do not need to consider here). It is remarkable that this result is sharp, and the reason for which the Haar system is not a basis for larger values of α is the most obvious one: The Haar function ψ given by (5) no longer belongs to the corresponding space. Note that this trivial obstruction also prevents (11) to yield a wavelet characterization of Besov spaces in that case; indeed the Haar function only has one non-vanishing coefficient on the Haar basis, and therefore its coefficients obviously satisfy (11); therefore, if this characterization held for the Haar basis, it would follow that the Haar function belongs to the corresponding Besov space, which is not the case.

Despite the limitation due to its irregularity, the Haar system has been of constant use in signal and image processing. A purpose of this paper is to show that this limitation can be mitigated by using a frame instead of an orthonormal basis, thus taking advantage of redundancy (as in phase reconstruction, or gravitational wave detection). In order to more precisely address this question, we note that the information within the definition of an unconditional basis may be granularized into two different points:

1. The *analysis* problem : Is it possible to characterize the fact that a function belongs to a function space by a condition on the moduli of its coefficients on the analyzing system, as given, e.g. by (11) in the case of Besov spaces?
2. The *synthesis* problem : When does the partial sum reconstruction formula (7) (or (9) in the case of wavelet bases) hold in the corresponding function space?

It is clear that the second point cannot be improved by using a redundant system; indeed, as soon as the "building blocks" do not belong to the function space, the norms in (7) are not even defined. However, there may be some room for improvement concerning the first point: More information would be unveiled if (11) holds for the coefficients on a redundant system, and one might expect that it can be converted into some regularity information on the function allowing to go "beyond" the limitation of Bourdaud's theorem. Several results actually back this intuition.

The first one is the fact that $C^\alpha(\mathbb{R})$ regularity can be characterized with the help of the *continuous wavelet transform* even if a non-smooth wavelet is used: Let the wavelet ψ satisfy

$$|\psi(t)| \leq \frac{C}{1+|t|^2} \quad \text{and} \quad \int_{\mathbb{R}} \psi(t)dt = 0. \tag{13}$$

The continuous wavelet transform of a function $f \in L^\infty \cap L^1_{loc}$ is defined by

$$\forall a > 0 \text{ and } b \in \mathbb{R}, \qquad C_\psi(f,a,b) = \int_{\mathbb{R}} f(t)\, \psi\left(\frac{t-b}{a}\right) \frac{dt}{a},$$

see [20, 21]. A key point is that the function f can be reconstructed from its continuous wavelet transform using a reconstructing wavelet φ which may differ from ψ: under weak conditions on φ, one has

$$f(t) = C \int_0^\infty \int_{\mathbb{R}} C_\psi(f,a,b)\, \varphi\left(\frac{t-b}{a}\right) \frac{db\, da}{a^2};$$

In particular, one can use a C^1 compactly supported function φ, which allows to prove the following characterization which holds *without regularity assumption on* ψ, see e.g. [10, 22].

Proposition 2.1 *Let $f \in L^\infty \cap L^1_{loc}$. Let $\alpha \in (0,1)$. Then $f \in C^\alpha(\mathbb{R})$ if and only if its Haar-wavelet continuous wavelet transform satisfies*

$$\exists C > 0, \quad \forall (a,b) \in \mathbb{R}^{*,+} \times \mathbb{R}, \qquad |C_\psi(f,a,b)| \leq C a^\alpha.$$

Another motivation of this paper is found in the theory of approximation: Let $\alpha \in]0,1[$ and let $f \in L^\infty([0,1])$. Denote by f_n the best L^∞ approximation of f by piecewise constant functions on the intervals $[k/n, (k+1)/n]$; the error $\| f - f_n \|_\infty$ decays as $n^{-\alpha}$ if and only if $f \in C^\alpha([0,1])$. This is remarkable, since the approximants f_n are not even continuous, and it is due to the fact that the discontinuities of the f_n do not fall at the same points but are interlaced; of course, no such result holds if one considers nested sets of discontinuities, as in e.g. dyadic approximation, i.e. if one restricts to the f_{2^j}, see e.g. Chap. 12 Sec. 2 of [14]. Note that, for the sake of simplicity, we stated this problem within the framework of L^∞ approximation of C^α functions using piecewise constant functions. However, the general case of L^p approximation of functions in a Besov space using splines of arbitrary order is studied in [14].

Before considering the problem of regularity characterization using Haar frames, we first recall motivations that recently arose in AI and learning using Haar systems.

3 Relevance of frames and Haar decompositions in Machine Learning

In addition to their established and useful role in mathematical analysis, frame decompositions and their associated extensions are widely sought in applied sciences. These decompositions often are a preferred and adapted tool to zoom in and unveil useful features/properties which are invariably critical to not only uncovering algorithmic solutions to many data-driven problems, but to also simplifying the ensuing computational challenges. As noted earlier, data scale-based information which has long played a key role in signal processing and statistics, has recently emerged in the so-called Machine Learning (primarily Non-linear and Neural Network-based), yielding creative and promising exploitation of wavelet-based information.

Convolutional Neural Networks (CNN), see e.g. [33], broadly referred to as Deep Learning (DL) is a more refined and better performing Neural Network, consisting of layers of banks of linear filters. The output of each of these filters (i.e. the convolution of a signal and the impulse response of the filter, also reflected by the parameters of the latter) is non-linearly transformed. Their parameters (linear weights of each of the filters) together with the layer biases are iteratively adapted by the gradient of a global output objective appropriately adjusted to the respective layers. The non-linear transformation is aimed at capturing higher order information modes present in the data. While a notion of progressive scale in the analysis may be present, the filter parameters are unlike those of a Multi-Resolution analysis in wavelet analysis where the basis functions are canonical and known a priori. While DL is predominantly used in inference applications where class labels are learned/approximated during training, a larger class of applications have been explored. Its pervasive success has, however, been primarily empirical in the computing and engineering sciences, and the mathematical theory has much left to achieve.

While regularity and smoothness issues are not of obvious relevance in DL primary inference applications, they are of great importance in generative and reconstruction (e.g., super-resolution) problems of interest. The learned filters in the respective layers of DL, provide an approximation of the input data/function at the associated scale. The corresponding non-linear activation functions may be interpreted as a further polynomial enrichment [42] at the varying scales of the learning process. Other important applications include the so-called super-resolution problem of increasing the resolution quality of images invoked the importance of judiciously selecting (albeit still an open problem) the proper scales to improve the performance. The DL layer-wise filter bank structure effectively provides a frame-like representation whose redundancy is pre-determined by the chosen number of filters and layers, and further refined by the gradient-based adaptation [46].

3.1 Towards a Mathematical Formalism of DL

The pursuit of a formalism of DL has led Cheng et al. [6] to provide an alternative interpretation of CNN as a wavelet-function-based multi-scale frame optimization of data (referred to as Scattering Networks). The computational efficiency of this systematic and optimal representation selected from a wavelet frame representation (for a selected wavelet function) was shown to provide a near-translation-invariance and a viable inference framework. In contrast to CNN, the a priori choice of a canonical analyzing wavelet function, and its resulting non-optimal feature/structure matching as in CNN, limited its performance. Subsequent and recent data-driven and optimal over-complete representation [37, 38] using Dictionary learning, ultimately proved to be very competitive with CNN with additional robustness to so-called adversarial noise. This re-interpretation of CNN sought for a large set of data classes, an optimal selection of data-driven atoms in a frame using the LASSO (Least Absolute Shrinkage and Selection Operator) algorithm [45]. This consisted of securing for a given function a sparse set of atoms from an over-complete set of functions according to an optimal L^2 reconstruction error. This was hierarchically pursued in each layer of a network, and in contrast to the CNN set prior choice of filters/atoms, the number of atoms is determined by the LASSO optimization criterion. The sparsity constraint on the representation coefficients also impacted the robustness of the network [38].

3.2 Graph-Based Deep Learning

The notion of CNN was extended to data-graph structures [3] where the Laplacian played a central role in the deep structure representation of a graph. Starting with the fact that a complex exponential is an eigenfunction of a Laplacian, it is argued that eigenvectors of a graph Laplacian can be used to construct an orthonormal basis for the space of functions defined on the vertices of a graph. Much like an inner product may be defined on such a space, a *graph convolution* may be defined. Using the basic duality property of a convolution and product in a space and its Fourier dual, they propose to exploit the Eigen spectrum of the graph Laplacian to efficiently obtain the so-called graph convolution, and hence yield a formalism for Graph-based DL. In a similar way Li et al. [36], constructed a computationally more efficient Graph-based DL using their so-called Haar Convolution for which a Haar decomposition is performed at a much lower computational cost including the inversion. The implications of the present paper for CNN and Deep Learning will be developed in [32]. In addition to the noted computational advantage, an increasing number of other fast frame-based graph approaches are being explored [48,50]. A particular future research interest would seek to preserve this orthonormal Haar representation efficiency on Graph DL to a Haar frame representation, and thereby inject the analysis with the properties developed in this paper.

The regularity results of a Haar system in this paper present a new potential for further refining Graph DL and for additionally providing a new perspective on computational impacts, such as invariance to translation which is critical to data-inherent transient changes.

4 Haar frame

We first define the frame that we will use. The Haar basis (9) is an orthogonal basis of $L^2(\mathbb{R})$. We add to it the two orthogonal bases obtained by shifting the elements of the Haar basis by 1/3 and 2/3. This means that our analysing system is composed of the

$$I_k(x) = \varphi\left(x - \frac{k}{3}\right), \text{ and } H_{j,k}(x) = \psi\left(2^j x - \frac{k}{3}\right), \qquad j, k \in \mathbb{Z}. \tag{14}$$

After a correct normalization, the I_k together with the $2^{j/2}H_{j,k}$ form a union of three orthonormal bases; therefore, they constitue a *tight frame*, i.e. a frame for which the inequalities in (1) are equalities; more precisely, it satisfies

$$\forall f \in L^2(\mathbb{R}), \qquad \| f \|_{L^2} = \frac{1}{3}\left(\sum_k |\langle f|I_k\rangle|^2 + \sum_{j,k} |\langle f|2^{j/2}H_{j,k}\rangle|^2\right). \tag{15}$$

A similar wavelet frame-driven approach was first exploited to optimize the translation invariance of an orthonormal wavelet basis representation of a function [23, 31]. Indeed, one well documented drawback of using an orthonormal wavelet basis in signal an image processing is that it does not supply a translation invariant representation, but is dependent on a particular discrete dyadic grid which is chosen. This drawback can be mitigated by oversampling this dyadic grid, thus replacing the initial orthonormal basis by a finite union of orthonormal bases. The choice supplied by (14) corresponds to an equally spaced oversampling by a factor 3. Applying the so-called Multi-scale representation (a wavelet basis representation over a number of scales) in many practical scenarios such as signal detection in Radar scenario, or reconstruction of a function/signal in noise (i.e. denoising) as well as parameter estimation in communication, is highly dependent on the translation-invariance of the transformation. As an example, a continuous wavelet transform or a wavelet frame (i.e. redundant) representation of a signal guarantees that a time delay estimation or a detection of a very short transient will be successfully achieved. While an orthogonal wavelet representation carries several important statistical properties (such as non-correlated coefficients) as well as parsimony, it may turn out to be unable to detect a short transient or estimate translation-sensitive parameters. Towards mitigating such limitation, an algorithm seeking a translation invariant wavelet representation of a given function was proposed, by searching for the best samples to prune along the tree-based structure of the frame, thus selecting at each scale the even/odd elements for orthogonalization. This exploitation of a redundant wavelet representation was similarly and later illustrated in [8] as what may be viewed as an ”averaging procedure” for improved denoising, and referred to as cycle-spinning. The temporal delay of a signal, ideally irrelevant to a successful reconstruction of a signal, is thus central to a translation-invariant representation, and its reflection in the proper coefficients thus becomes essential.

Finally, note that wavelet frames with similar properties have been developped independently of the wavelet theory by M. Frazier and B. Jawerth under the denomination of the ϕ-*transform*, cf. [19] and ref. therein.

4.1 Uniform regularity

The first problem we consider is the characterization of the uniform Hölder spaces $C^\alpha(\mathbb{R})$ on Haar frame coefficients

$$C_k = \int f(x) I_k(x) dx \quad \text{and} \quad c_{j,k} = 2^j \int f(x) H_{j,k}(x) dx. \tag{16}$$

Recall that these spaces coincide with the Besov spaces $B^{\alpha,\infty}_\infty(\mathbb{R})$, see [40]; if $0 < \alpha < 1$, an equivalent definition is given by

$$f \in L^\infty(\mathbb{R}) \quad \text{and} \quad \exists C, \ \forall x, y \in \mathbb{R}, \quad |f(x) - f(y)| \leq C|x - y|^\alpha, \tag{17}$$

thus defining a norm equivalent to the $C^\alpha(\mathbb{R})$ norm. We first note that a characterization of $C^\alpha(\mathbb{R})$ on the Haar basis coefficients cannot hold, because it would be satisfied by the Haar function itself (the coefficients of which all vanish except for one), and the Haar function is discontinuous, and therefore does not belong to $C^\alpha(\mathbb{R})$. Nonetheless, Theo. 1 below shows that such a characterization holds if using the coefficients on the Haar frame (14).

We will make the following regularity assumptions on the functions we will consider.

Definition 3 *Let f be a locally bounded function; f is Lebesgue-regular if every point is a Lebesgue point of f, i.e.*

$$\forall x \in \mathbb{R}, \qquad f(x) = \lim_{r \to 0} \frac{1}{2r} \int_{x-r}^{x+r} f(t) dt.$$

Note that this definition implicitly makes the assumption that functions are defined “point to point” and not “except for a set of vanishing measure”. Continuous functions are of course Lebesgue regular; but this class also allows for discontinuities. For instance assume that, at every point x, f has a right and a left limit at x, and that, at every discontinuity point x_0, f satisfies

$$f(x_0) = \frac{1}{2} \left(\lim_{x \to x_0^+} f(x) + \lim_{x \to x_0^-} f(x) \right);$$

then f is clearly Lebesgue regular. A key property that we will use is that the wavelet series of a Lebesgue-regular function f converges everywhere to f; this follows from the fact that, if $f \in L^1(\mathbb{R})$, then the partial sums of f converge at its Lebesgue points, see [44, 47].

We need to be more precise here concerning the meaning of the characterization of $C^\alpha(\mathbb{R})$ by (17). Let $f \in L^1_{loc} \cap L^\infty$; then its wavelet coefficients are well defined, and, if they satisfy (12), then the wavelet series of f converges uniformly towards a function g which coincides almost everywhere with f and satisfies (17). It follows that $f(x) = g(x)$ at every point x if f is defined "point by point" and is Lebesgue regular. These precautions, which may seem superfluous, are required if one wants to avoid absurd statements such as "the wavelet coefficients of the characteristic function of the rationals all vanish, and therefore it is a C^∞ function".

Theorem 1 *Let $\alpha \in (0,1)$. Let f be a Lebesgue regular function; $f \in C^\alpha(\mathbb{R})$ if and only if its Haar frame coefficients* (16) *satisfy*

$$\exists C, \quad \forall k, \quad |C_k| \le C \text{ and } \quad \forall j, k, \qquad |c_{j,k}| \le C 2^{-\alpha j}. \tag{18}$$

Remarks:

1. this is the same characterization as if we were using a smooth wavelet basis: Everything happens as if the Haar basis was composed of smooth wavelets.

2. It follows from the proof that the quantity

$$\sup_k |C_k| + \sup_{j,k} |2^{\alpha j} c_{j,k}|$$

supplies a norm which is equivalent to the C^α norm.

3. If f is a bounded locally integrable function, it follows that, if f satisfies (18), then f coincides a.e. with a C^α function (which, thus, is Lebesgue regular).

Proof of Theo. 1: Assume that $f \in C^\alpha(\mathbb{R})$. Then

$$|C_k| = \left| \int f(x) I_k(x) dx \right| \leq \parallel I_k \parallel_{L^1} \parallel f \parallel_{L^\infty} \leq \parallel f \parallel_{L^\infty},$$

hence the first statement in (18) holds. Let

$$I_{j,k} = \left[\frac{k}{3 \cdot 2^j}, \frac{k}{3 \cdot 2^j} + \frac{1}{2^j} \right] \tag{19}$$

be the support of $H_{j,k}$. we denote by $I^-_{j,k}$ the left half of $I_{j,k}$ and by $I^+_{j,k}$ the right half of $I_{j,k}$. Then

$$c_{j,k} = 2^j \int_{I_{j,k}} f(x) H_{j,k} dx = 2^j \int_{I^-_{j,k}} f(x) dx - 2^j \int_{I^+_{j,k}} f(x) dx = 2^j \int_{I^-_{j,k}} (f(x) - f(x + 2^{-j}) dx,$$

so that

$$|c_{j,k}| \leq 2^j \int_{I^-_{j,k}} |f(x) - f(x + 2^{-j})| dx \ \leq C 2^{-\alpha j},$$

hence the second statement in (18) holds.

Conversely, assume that (18) holds. Since f is Lebesgue regular, we can use the reconstruction formula for the Haar orthonormal wavelet basis only (which converges everywhere towards the pointwise value of f); thus

$$\forall x \in \mathbb{R}, \qquad |f(x)| = \sum_{k \in \mathbb{Z}} |c_k| \ |\varphi(x-k)| + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} |c_{j,k}| \ |\psi(2^j x - k)|,$$

$$\leq C + C \sum_{j=0}^{\infty} 2^{-\alpha j}$$

so that $f \in L^\infty$. Let us now estimate increments of f. We have three possible reconstruction formulas for f using any of the three orthonormal bases composing the tight frame; the idea of the proof is to use this extra flexibility. Let $x \neq y$ be given. Define J by

$$\frac{1}{4} 2^{-J} \leq |x - y| < \frac{1}{2} \cdot 2^{-J}.$$

Consider now the intervals $I_{J,k}$. Since these intervals are of length 2^{-J} and deduce from each other by a shift of $\frac{1}{3} \cdot 2^{-J}$, at least one of them contains both points x and y; we denote it by I_{J,k_J}. We now use either the Haar basis or one of its two "sisters" shifted by 1/3, the choice being driven by the fact that the interval I_{J,k_J} that we picked is the support of an element $H_{J,k}$ of the chosen basis. This implies that, for all generations $j < J$, either the support of an $H_{j,k}$ of this basis does not contain x and y, or x and y are in the same "half" of the support of $H_{j,k}$. Therefore, let us use the reconstruction formula using this orthonormal basis (and let us denote its elements by φ_k and $\psi_{j,k}$); since it converges everywhere towards the pointwise value of f, we get

$$f(x) - f(y) = \sum_k C_k(\varphi_k(x) - \varphi_k(y)) + \sum_{j \le J} \sum_k c_{j,k}(\psi_{j,k}(x) - \psi_{j,k}(y)) + \cdots$$

$$\cdots + \sum_{j > J} \sum_k c_{j,k}(\psi_{j,k}(x) - \psi_{j,k}(y)).$$

Because of our choice of the basis, it follows that

$$\forall k, \quad \varphi_k(x) = \varphi_k(y) \qquad \text{and} \qquad \forall j \le J, \, \forall k, \quad \psi_{j,k}(x) = \psi_{j,k}(y);$$

therefore

$$|f(x) - f(y)| \le \sum_{j > J} \sum_k |c_{j,k}||\psi_{j,k}(x) - \psi_{j,k}(y)|.$$

At each generation j, at most two terms bring a contribution; using (18), we get

$$|f(x) - f(y)| \le \sum_{j > J} C 2^{-\alpha j} \le C 2^{-\alpha J} \le C|x - y|^\alpha,$$

so that Theo. 1 holds.

As pointed out to us by Albert Cohen, the argument of selecting the "right" interval which includes both x and y, is similar to the one developed in the *mixing lemma*, used in order to derive optimal order of approximation of functions in a Besov space by sequences f_n of piecewise constant functions on the intervals $[k/n, (k+1)/n]$ see e.g. Chap. 12 Sec. 2 of [14]. Nonetheless, note that the approximants supplied by the three shifted Haar systems supply a much more economical sequence, and allows to obtain the same order of approximation characterization result, as we shall now show.

We now circle back to the problem that we raised at the end of Sec. 2. We will check that the order of approximation supplied by the piecewise constant functions between dyadic intervals and dyadic intervals shifted by $1/3$ is as good as if we used the whole sequence of piecewise constant functions on the intervals $[k/n, (k+1)/n]$. Let us be more precise. Let $f \in L^\infty$ and denote by $P_j(f)$ its partial reconstruction at scale 2^{-j} using the Haar basis and let $Q_j = P_{j+1} - P_j$. If $f \in C^\alpha$, then it follows from Theo. 1 that its Haar basis coefficients satisfy

$|c_{j,k}| \leq C2^{-\alpha j}$. Since $Q_j(f) = \sum_k c_{j,k}\psi_{j,k}$, it follows that $|Q_j(f)(x)| \leq C2^{-\alpha j}$; Therefore, since

$$f - P_j(f) = \sum_{l>j} Q_j(f),$$

it follows that

$$\| f - P_j(f) \|_\infty \leq C2^{-\alpha j}. \tag{20}$$

Of course this condition is not sufficient by itself, as shown by picking for f the Haar wavelet which satisfies (20) but has no positive Hölder regularity. We now check that it is sufficient to add the same requirement on the projections on the shifted Haar systems in order to obtain a necessary and sufficient condition. Let $P_j^i(f)$ $(i = 1, 2, 3)$, denote (for $i = 1$) the projection of f on the Haar system at scale j (i.e. $P_j^1 = P_j$) and (for $j = 2, 3$) the projection of f on the translates respectively by 1/3 and -1/3 of the Haar system at scale j. We shall prove the following result.

Proposition 4.1 *Let $f \in L^\infty$). The condition*

$$\exists C > 0, \quad \forall i = 1, 2, 3, \quad \forall j \geq 0, \qquad \| f - P_j^i(f) \|_\infty \leq C2^{-\alpha j}. \tag{21}$$

is equivalent to $f \in C^\alpha$.

Proof: By (20), we already know that, if $f \in C^\alpha$, then $\| f - P_j^1(f) \|_\infty \leq C2^{-\alpha j}$; and, by translating the Haar system, the same result holds for $f - P_j^2(f)$ and $f - P_j^3(f)$. Conversely, assume that (21) holds. Note that

$$Q_j(f) = (f - P_j(f)) - (f - P_{j+1}(f)).$$

Therefore

$$\| Q_j^1(f)) \|_\infty \leq C2^{-\alpha j};$$

but, on the support of $\psi_{j,k}$, $|Q_j^1(f)(x)| = |c_{j,k}^1|$. Thus, we obtain that

$$\exists C > 0, \quad \forall j, \quad \forall k, \qquad |c_{j,k}^1| \leq C2^{-\alpha j}. \tag{22}$$

The same argument applies to $Q_j^2(f)$ and $Q_j^3(f)$, so that all the Haar frame coefficients of f satisfy (22); we can now apply Theo. 1 which implies that $f \in C^\alpha$.

4.2 Pointwise regularity: The Haar basis

We will now consider the problem of characterizing pointwise regularity. We start by recalling this notion, and the existing partial results if the Haar basis is used.

Definition 4 *Let $x_0 \in \mathbb{R}$ and $\alpha \geq 0$. A locally bounded function $f : \mathbb{R} \to \mathbb{R}$ belongs to $C^\alpha(x_0)$ if there exist $C, R > 0$ and a polynomial P_{x_0} with $deg(P_{x_0}) < \alpha$ such that*

$$\text{for a.e. } x \in B(x_0, R), \qquad |f(x) - P_{x_0}(x)| \leq C|x - x_0|^\alpha. \tag{23}$$

The pointwise Hölder exponent of f at x_0 is $h_f(x_0) = \sup\{\alpha : f \in C^\alpha(x_0)\}$.

The polynomial P is unique; it is called the *Taylor polynomial* of f at x_0. When using smooth wavelets, criteria based on the wavelet coefficients on an orthonormal wavelet basis allow to recover pointwise regularity; let us recall some notations. A dyadic interval is an interval of the form

$$\lambda (= \lambda_{j,k}) = \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right).$$

Let λ be a dyadic interval; 3λ denotes the interval of same center as λ and three times wider. Wavelets coefficients can therefore be indexed by dyadic intervals: We will write $c_\lambda := c_{j,k}$.

Definition 5 *Let f be a locally bounded function, and let the (φ_k) and $(\psi_{j,k})$ generate a smooth wavelet basis. The wavelet leaders of f are the quantities*

$$d_\lambda = \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|.$$

If the generating wavelet is smooth, then wavelet leaders allow to estimate pointwise Hölder exponents, see [24] for the initial 2-microlocal wavelet coefficients criterium and [27] for its reformulation in terms of wavelet leaders, which we now recall. We denote by $\lambda_j(x_0)$ the dyadic interval of width 2^{-j} which contains x_0.

Theorem 2 *Let $f \in C^\varepsilon(\mathbb{R})$ for an $\varepsilon > 0$. If the generating wavelets φ and ψ belong to $C^N(\mathbb{R})$, then*

$$\forall x_0 \in \mathbb{R} : \quad \text{if } h_f(x_0) < N, \text{ then} \qquad h_f(x_0) = \liminf_{j \to +\infty} \frac{\log\left(d_{\lambda_j(x_0)}\right)}{\log(2^{-j})}. \tag{24}$$

This question is more difficult to answer when using an irregular wavelet basis. To our knowledge, it was only tackled in [29] by B. Mandelbrot and the first author. Their motivation was the determination of the pointwise regularity of the Polya function, a famous example of a continuous "Peano-type" space-filling function. A remarkable property is that the coefficients of the Polya function on the Schauder basis (which consists of the primitives of the Haar basis) are given explicitly by Bernoulli binomial coefficients. Thus, the study of its regularity can be reduced to the understanding of which regularity properties of a function can be derived from its decomposition on a wavelet basis of limited smoothness; indeed, though it is not exactly a wavelet basis, the Schauder basis has the same algorithmic form as a wavelet basis, and the same type of regularity results, such as Theo. 2, apply, as long as the regularity exponents involved are below the regularity of the elements of the Schauder basis (which are continuous piecewise linear functions and therefore are Lipschitz functions), see [29]. Of course, no such result can hold for regularity exponents higher than the regularity of the wavelet, for the same reasons as in the uniform regularity case: analyzing one element of the basis using the same basis could not, by construction, detect the irregularity of this element at some points. Let us first recall the results concerning pointwise regularity results for the Haar basis. We will need the following notions. Recall that a *dyadic rational* is a point of the form $k/2^j$, for $j,\, k \in \mathbb{Z}$.

Definition 6 *Let $x \in \mathbb{R}$. The rate of approximation of x by dyadic rationals is*

$$r(x) = \limsup_{j \to +\infty} \frac{\log\left(dist(x, 2^{-j}\mathbb{Z})\right)}{\log(2^{-j})}. \tag{25}$$

Every x satisfies $r(x) \geq 1$, and almost every x satisfies $r(x) = 1$. We recall the following result of [29].

Proposition 4.2 *Let f be a locally bounded function, and $x_0 \in \mathbb{R}$. If $f \in C^\alpha(x_0)$ for $\alpha < 1$, then its wavelet leaders on the Haar basis satisfy*

$$|d_{\lambda_j(x_0)}| \leq C 2^{-\alpha j}. \tag{26}$$

Conversely, if (26) holds and if the Haar basis coefficients $c_{j,k}$ satisfy the uniform decay assumption

$$\exists \varepsilon, C > 0: \qquad |c_{j,k}| \leq C 2^{-\varepsilon j}, \tag{27}$$

then

$$h_f(x_0) \geq \frac{\alpha}{r(x_0)}.$$

This result yields the best possible pointwise regularity that can be inferred from the knowledge of the size of the Haar basis coefficients. Since $r(x) = 1$ a.e., if (27) holds, then this result yields the exact pointwise regularity at almost every point. Besides dyadic points (where it is clear that decay estimates on the Haar coefficients cannot allow to estimate pointwise regularity), the estimate of $h_f(x_0)$ yielded by Prop. 4.2 deteriorates if $r(x_0)$ is large, i.e. if there exists a sequence of scales such that x_0 is "very close" to dyadic points at these scales. This also indicates that estimates on the Haar frame coefficients might not suffer from this drawback; indeed, if x_0 is "close" to dyadic points at certain scales, it will be "far" from the dyadic points shifted by 1/3 (at the same scales). Before formalizing this idea (see Theo. 3 below) we first consider a simple example where the limitations of Prop. 4.2 can be observed.

4.3 The Haar-Weierstrass function

Let R be the 2-periodic function defined by

$$\begin{cases} R(x) = 0 & \text{if } x \in \mathbb{Z}, \\ R(x) = 1 & \text{if } x \in (2k, 2k+1) \\ R(x) = -1 & \text{if } x \in (2k+1, 2k+2) \end{cases} \tag{28}$$

Definition 7 *Let $\beta > 0$; The Haar-Weierstrass functions $\mathcal{H}_\beta$ are defined by*

$$\forall x \in \mathbb{R}, \qquad \mathcal{H}_\beta(x) = \sum_{j=1}^{\infty} 2^{-\beta j} R(2^j x) \tag{29}$$

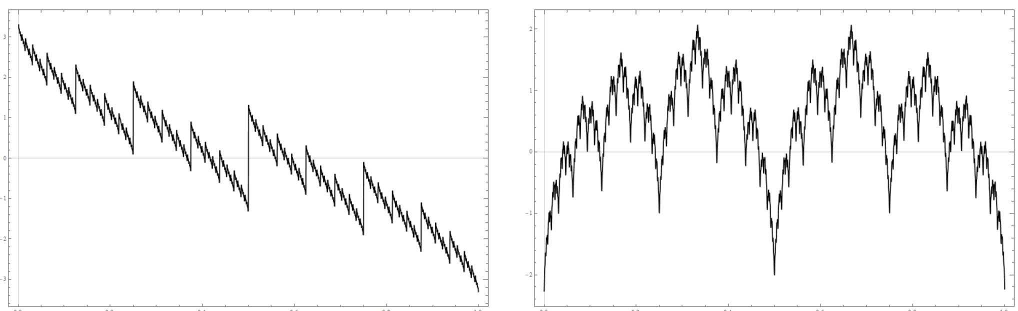
$$= \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}} 2^{-\beta(j+1)} \psi_{j,k}(x). \tag{30}$$

This is a "poor's man" version of the Weierstrass functions, where the smooth sine (or cosine) function is replaced by the saw-tooth function R. Note that its Haar basis coefficients have the same amplitude at a given scale, so that it is impossible to infer from their size some variations on the pointwise regularity of $\mathcal{H}_\beta$. Nonetheles, we will see that its Hölder exponent is an extremely irregular function which takes all values between from 0 to β on any interval of arbitrary small length.

First, note that the series (29) is normally convergent on $\mathbb{R}$. It follows that $\mathcal{H}_\beta$ is continuous at the points which are not dyadic rationals; and, additionally, if $r_{j,k} = k/2^j$ (with k odd) is a dyadic rational in $(0,1]$, then $\mathcal{H}_\beta$ has a right and a left limit at $r_{j,k}$. The amplitude of the discontinuities at these points will yield an upper bound of the Hölder exponent, using the following lemma of [26].

Lemma 4.3 *Let $f \in L^\infty_{loc}$ be such that f is Lebesgue-regular and has everywhere a right and a left limit. Let $x \in \mathbb{R}$, and let $x_n \to x$ be a sequence of discontinuity points of f; denote by $\Delta_f(x_n)$ the jump of f at x_n. Then*

$$h_f(x_0) \leq \liminf_{s \to x_0} \frac{\log(\Delta_f(x_n))}{\log(|x - x_n|)}. \qquad (31)$$



Haar-Weierstrass vs. Meyer-Weierstrass functions

Credit: Guillaume Saes

The wavelet series (30) are displayed for $\beta = 1/2$, using Haar's wavelet (left) and Meyer's wavelet (right). Since Meyer's wavelet is C^∞, it follows that the Meyer-Weierstrass function is very similar to a Weierstrass function. In particular, its Hölder exponent takes everywhere the value $H = 1/2$. This is in sharp contrast with the Haar-Weierstrass function which has discontinuities at dyadic points, and whose Hölder exponent takes all possible values between 0 and 1/2.

Let us compute the amplitude $\Delta(r_{j,k})$ of the jump (i.e. the difference between these limits) at $r_{j,k}$: The first of the $R(2^l x)$ which has a discontinuity at $r_{j,k}$ is $R(2^j x)$ and its jump is negative, of amplitude $-2 \cdot 2^{-\beta j}$ and the next ones, for $l > j$ have a positive jump, of amplitude $2 \cdot 2^{-\beta l}$. It follows that the jump at $r_{j,k}$ is

$$\Delta(r_{j,k}) = C_\beta \cdot 2^{-\beta j}, \qquad \text{where} \qquad C_\beta = \frac{2^{1-\beta} - 1}{1 - 2^{-\beta}} \qquad (32)$$

so that $C_\beta \neq 0$ if $\beta \neq 1$, which we assume.

It follows from Def. 6, (31) and (32) that the Hölder exponent of $\mathcal{H}_\beta$ satisfies

$$\forall x \in \mathbb{R}, \qquad h_{\mathcal{H}_\beta}(x) \leq \frac{\beta}{\alpha_2(x)}.$$

On other hand, the second part of Prop. 4.2 yields that

$$\forall x \in \mathbb{R}, \qquad h_{\mathcal{H}_\beta}(x) \geq \frac{\beta}{\alpha_2(x)}.$$

We have thus obtained the following result.

Proposition 4.4 *Let $\beta > 0$ such that $\beta \neq 1$. The Hölder exponent of $\mathcal{H}_\beta$ is*

$$\forall x \in \mathbb{R}, \qquad h_{\mathcal{H}_\beta}(x) = \frac{\beta}{\alpha_2(x)}. \tag{33}$$

Remark: This implies that $\mathcal{H}_\beta$ is a multifractal function, i.e. the *equi-hölder sets*

$$E_H = \{x : h_{\mathcal{H}_\beta}(x) = H\}$$

are everywhere dense fractal sets of Hausdorff dimension $D_{\mathcal{H}_\beta}(H) = H/\beta$ for $H \in [0, \beta]$ (and they are empty if $H \notin [0, \beta]$). This is a direct consequence of the fact that points which satisfy $r(x) = R$ have Hausdorff dimension $1/R$, see e.g. [17] (or [26] where similar functions are studied).

4.4 Pointwise regularity: The Haar frame

We will now show that, in contradistinction with Prop. 4.2, the Haar frame coefficients allow to recover the pointwise Hölder exponent everywhere. Let us first introduce some notations. We index the elements of the Haar frame $H_{j,k}$ by their support $\lambda = I_{j,k}$, see (19) (note that, though the length of λ is 2^{-j}, it is not necessarily a dyadic interval). The corresponding Haar coefficient is

$$c_\lambda = 2^j \int f(x) H_{j,k}(x) dx.$$

Haar leaders also are indexed by dyadic intervals and defined by

$$d_\lambda = \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|.$$

We will prove the following result.

Theorem 3 *Let f be a locally bounded function, and let $\alpha > 0$. If $f \in C^\alpha(x_0)$, and if the Taylor polynomial of f at x_0 is constant, then its wavelet leaders on the Haar frame satisfy*

$$d_{\lambda_j(x_0)} \leq C 2^{-\alpha j}. \tag{34}$$

Conversely, if (34) *holds and if the Haar frame coefficients* $c_{j,k}$ *satisfy the uniform decay assumption*

$$\exists \varepsilon, C > 0: \qquad |c_{j,k}| \leq C 2^{-\varepsilon j}, \tag{35}$$

then

$$\exists C: \quad \textit{if } |x - x_0| \leq 1/2, \qquad |f(x) - f(x_0)| \leq C|x - x_0|^\alpha \log(|x - x_0|). \tag{36}$$

Remarks: The restriction on the Taylor polynomial is automatically satisfied if $0 < \alpha < 1$. It is easy to check that it is also satisfied by a class of functions which plays an important role in multifractal analysis: The distribution functions of singular measures (with no restriction on α). The first statement of the theorem is a classical result, and we recall its proof for completeness, see e.g. [29], or, more recently [30] where it used in the more general context of pointwise $T^p_\alpha(x_0)$ regularity: Its purpose was to obtain a pointwise irregularity criterium based on the continuous (Haar) wavelet transform for the Brjuno function. The Haar wavelet is a natural choice in this case because integrals of the Brjuno function on certain intervals with rational ends have an (almost) explicit form, so that, for an appropriate positioning of the support of the Haar wavelet, the orders of magnitude of the Haar coefficients are known.

Proof: Assume that $f \in C^\alpha(x_0)$ for $\alpha < 1$. Let $j \geq 0$ be given, $j' \geq j$ and $\lambda' \subset 3\lambda_j(x_0)$; since $H_{j',k'}$ has a vanishing integral,

$$c_{j',k'} = 2^{j'} \int_{I_{j',k'}} f(x) H_{j',k'} dx = 2^{j'} \int_{I_{j',k'}} (f(x) - f(x_0)) H_{j',k'} dx,$$

so that

$$|c_{j',k'}| \leq 2^{j'} \int_{I_{j',k'}} |f(x) - f(x_0)| dx \leq C 2^{j'} \int_{I_{j',k'}} |x - x_0|^\alpha dx \ .$$

Since x and x_0 belong to the support of $I_{j',k'}$, hence to $3\lambda_j(x_0)$, $|x - x_0| \leq 2^{-j}$ and it follows that $|c_{j',k'}| \leq C 2^{-\alpha j}$; and therefore (34) holds.

Suppose now that (34) and (35) hold. Because of Theo. 1, f belongs to $C^\varepsilon(\mathbb{R})$ and therefore the wavelet series of f converges to the pointwise value of f everywhere. Define j by

$$\frac{1}{4} 2^{-j} \leq |x - x_0| < \frac{1}{2} \cdot 2^{-j}. \tag{37}$$

At least one of the three Haar bases is such that, at the generation j, x and x_0 belong to the same interval I_{j,k_j}. As in the proof of uniform regularity, in order to estimate increments of f we use this Haar basis for the reconstruction formula:

$$f(x)-f(x_0)=\sum_k C_k(\varphi_k(x)-\varphi_k(x_0))+\sum_{j'\le j}\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0))+\cdots$$

$$\cdots+\sum_{j'>j}\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0)).$$

The terms for $j'<j$ vanish because x and x_0 belong to the same (right or left) half of the support of $\psi_{j',k'}$ (or of φ_k). As regards the terms for $j'\ge j$, we first assume that $j'\le [Aj]$, where A is a (large) constant, that will be fixed later. Each sum

$$\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0))$$

contains at most two nonvanishing terms: the ones such that x or x_0 belong to $I_{j',k'}$; but, in that case (34) implies that $|c_{j',k'}|\le C2^{-\alpha j}$. Therefore

$$\left|\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0))\right|\le C2^{-\alpha j},$$

and

$$\sum_{j'=j}^{[Aj]}\left|\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0))\right|\le Cj2^{-\alpha j}. \tag{38}$$

Assume now that $j'>j$. Because of the localization of the Haar basis, (35) implies that

$$\left|\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0))\right|\le C2^{-\varepsilon j},$$

and

$$\sum_{j'>[Aj]}\left|\sum_{k'} c_{j',k'}(\psi_{j',k'}(x)-\psi_{j',k'}(x_0))\right|\le C2^{-\varepsilon Aj}. \tag{39}$$

We pick A such that $\varepsilon A=\alpha$; (37) implies that $j\le C|\log(|x-x_0])|$, so that (36) follows from (38) and (39).

It follows from Theo. 2 extends to the Haar frame setting: If f satisfied (35), and if If the generating wavelets φ and ψ belong to $C^N(\mathbb{R})$, then

$$h_f(x_0)=\liminf_{j\to+\infty}\ \frac{\log\left(d_{\lambda_j(x_0)}\right)}{\log(2^{-j})},$$

if the right-hand side is less than 1.

5 Concluding remarks and open problems

If a regular oversampling is not required, the results we obtained also hold for less redundant frames, and the shift by $1/3$ can be replaced by other rational shifts that are not dyadic (a dyadic shift would lead to no new Haar system elements for j large enough). Indeed the key argument in the proofs of Theorems 1 and 2 and Prop. 4.1 is that, if x and y are such that $|x-y| \sim 2^{-j}$, then one can find a translated dyadic interval of length close to 2^{-j} which contains both x and y. This is clearly possible using only the Haar basis and one translate by a rational $r = p/(2k+1)$. Indeed, if I is an interval of length l satisfying $2^{-j} \leq l < 2 \cdot 2^{-j}$; let m be defined by

$$\frac{p}{8(2k+1)} \leq 2^{-m} < \frac{p}{4(2k+1)}.$$

Then I clearly is included either in a dyadic interval of length 2^{-j+l} or in a interval of the same length obtained as a shift by r of a dyadic interval. Therefore, the proofs of Theorems 1 and 2 and Prop. 4.1 work in the same way for such unions of two orthonormal bases. The same conclusion also holds if adding additional translates by $r = q/(2k+1)$ for several values of q (and in particular all of them, if one is concerned by the requirement of a regular oversampling). Note that the question of finding appropriate irrational translations for which the above results would hold is an open problem.

Similarly, the above results extend to the several variable version of the Haar system, which is obtained by a tensor product construction: For $x = (x_1, \cdots, x_d)$, we define

$$\Phi(x) = \varphi(x_1) \cdots \varphi(x_d),$$

and

$$\Psi^{(i)}(x) = \psi_1(x_1) \cdots \psi_d(x_d),$$

where the ψ_l are either the one-variable functions φ or ψ, the choice $\varphi(x_1) \cdots \varphi(x_d)$ being excluded (so that there are $2^d - 1$ wavelets $\Psi^{(i)}$). Then the d-variables Haar basis is composed of the $\Phi(x-k)$ for $k \in \mathbb{Z}^d$ and the $2^{dj/2}\Psi^{(i)}(2^j x - k)$ for $j \geq 0$ and $k \in \mathbb{Z}^d$. A regularly spaced Haar frame is obtained by shifting this basis by the vectors $\sum \varepsilon_i e_i/3$ where $\varepsilon_i \in \{0, 1, 2\}$ and the e_i are the elements of the canonical basis of $\mathbb{R}^d$. This yields 3^d orthonormal bases, and the d-variables Haar frame is composed of the union of these bases. The proofs that we gave extend without difficulty to this setting; indeed, the key point is to notice that, if $|x-y| \sim 2^{-j}$, then one can find a translated dyadic cube of width close to 2^{-j}

which contains both x and y; it suffices to use, in each direction of the canonical basis, the translation supplied by the one-dimensional case for the corresponding coordinate of the segment $[x, y]$; this yields an interval of dyadic length in each variable; these lengths are not necessarily the same in each direction, but picking the largest one will yield a cube of dyadic width which includes the segment $[x, y]$, and thus has the required property.

Theorems 1 and 2 also extend to piecewise smooth wavelets, such as the "spline wavelets" constructed by G. Battle and P.-G. Lemarié, see [1, 34]; in that case, the same proofs allow to characterize global and pointwise regularity up to an order α given by the number of vanishing moments of the wavelet, which is larger than its uniform regularity (and was the natural bound for previous regularity results). This is possible because these wavelets are piecewise polynomials between integers. In contrast, an interesting open problem would be to determine if these results could be extended to e.g. Daubechies wavelets, the singularities of which are not located at integers, but on fractal sets, see [13].

Note that the extension of Prop. 4.1 for d-variables yields shifted cubes of dyadic width such that the piecewise constant functions of these cubes are sufficient to yield optimal rate of approximation which is more more sparse than if one used approximation by functions f_n which are piecewise constant on the cubes $\prod_{i=1}^{d} \left[\frac{k_i}{n}, \frac{k_i+1}{n}\right]$. Using only two shifted Haar bases instead of 3 yields a frame of redundancy 2^d in dimension d. This leaves open the question of minimal redundancy in dimension d that would be sufficient to derive the conclusions of Theorems 1 and 2 and Prop. 4.1.

References

[1] G. Battle. A block spin construction of ondelettes, Part II: The QFT connection. *Comm. Math. Phys.*, 114:93–102, 1988. 25

[2] G. Bourdaud. Ondelettes et espaces de Besov. *Revista Matemática Iberoamericana*, 11(3):477–512, 1995. 5

[3] J. Bruna, W. Zaremba, A. Szlam, and Y. Lecun. Spectral networks and locally connected networks on graphs. *ICLR 2014*, 27, 2016. 10

[4] D. Larson C. Heil, P. Jorgensen. *Wavelets, Frames, and Operator Theory.* Contemporary Mathematics, A. M. S., 2004. 2

[5] E. Chassande-Mottin, S. Jaffard, and Y. Meyer. Des ondelettes pour détecter les ondes gravitationnelles. *Gazette de la S.M.F.*, 148:61–64, 2016. 3

[6] Xiuyuan Cheng, Xu Chen, and Stéphane Mallat. Deep Haar scattering networks. *Information and Inference: A Journal of the IMA*, 5:105–133, 2016. 9

[7] O. Christensen. *An introduction to Frames and Riesz Bases.* Birkhäuser, 2003. 2

[8] R. Coifman and D. Donoho. Translation-invariant denoising. In *Wavelets and Statistics, Antoniadis A., Oppenheim G. (eds) Lect. Not. Stat.*, volume 103, pages 121–150, Lyon, France, 1995. 12

[9] R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *C. R. Acad Sci. Paris, Sér I, Math.*, 312:259–261, 1991. 3

[10] I. Daubechies. *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992. 7

[11] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys*, 27:1271–1283, 1996. 1

[12] I. Daubechies, S. Jaffard, and J.L. Journé. A simple Wilson orthonormal basis with exponential decay. *SIAM J. Math. Anal.*, 22:554–572, 1991. 3

[13] I. Daubechies and J. C. Lagarias. On the thermodynamic formalism for multifractal functions. *Reviews in Mathematical Physics*, 6:1033–1070, 1994. 25

[14] R.A. DeVore and G.G. Lorentz. *Constructive Approximation.* Springer-Verlag, New York, 1993. 8, 15

[15] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia. *J. Roy. Statist. Soc.*, B 57(2):301–369, 1995. 5

[16] R. J. Duffin and A. C. Schaeffer. Painless nonorthogonal expansions. *Trans. Amer. Math. Soc.*, 72:341–366, 1952. 1

[17] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* John Wiley & Sons, West Sussex, England, 1993. 21

[18] P. Flandrin. *Explorations in Time-Frequency Analysis.* Cambridge University Press, 2018. 3

[19] M. Frazier and B. Jawerth. A discrete transform and decompositions of distribution spaces. *J. Funct. Anal.*, 93:34–170, 1990. 12

[20] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math.*, 15:723–736, 1984. 7

[21] A. Grossmann, J. Morlet, and T. Paul. Transforms associated to square integrable representations I. *J. Math. Phys. (reedited in : Fundamental Papers in Wavelet Theory, C. Heil et D. Walnut, eds.)*, 26:2473–2479, 1985. 7

[22] M. Holschneider and Ph. Tchamitchian. Pointwise analysis of Riemann's "non differentiable" function. *Inventiones Mathematicae*, 105:157–176, 1991. 7

[23] H. Carfantan J.-C. Pesquet, H. Krim and J. G. Proakis. Estimation of noisy signals using time-invariant wavelet packets. In *Proc. ASILOMAR Conf.*, pages 31–34, 1993. 11

[24] S. Jaffard. Exposants de Hölder en des points donnés et coefficients d'ondelettes,. *Compt. Rend. Acad. Sci.*, 308:79–81, 1989. 17

[25] S. Jaffard. A density criterion for frames of complex exponentials. *Michig. Math. J.*, 38:339–348, 1991. 2

[26] S. Jaffard. Old friends revisited: the multifractal nature of some classical functions. *Journal of Fourier Analysis and Applications*, 3(1):1–22, 1997. 20, 21

[27] S. Jaffard. Wavelet techniques in multifractal analysis. In *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus and M. van Frankenhuijsen, Eds., Proc. Symposia in Pure Mathematics*, volume 72(2), pages 91–152. AMS, 2004. 17

[28] S. Jaffard and H. Krim. Regularity properties of Haar frames. *Compt. Rend. Acad. Sci., to appear*, 2021. 3

[29] S. Jaffard and B. Mandelbrot. Local regularity of nonsmooth wavelet expansions and application to Polya's function. *Adv. Math.*, 120:265–282, 1996. 18, 22

[30] S. Jaffard and B. Martin. Multifractal analysis of the Brjuno function. *Invent. Math.*, 212:109–132, 2018. 22

[31] H. Krim J.C. Pesquet and H. Carfantan. Time-invariant orthonormal wavelet representations. *IEEE Transactions on Signal Processing*, 44:1964 – 1970, 1996. 11

[32] H. Krim, S. Jaffard, S. Roheda, S. Mahdizadehaghdam, and A. Panahi. On stabilizing generative adversarial networks. *In preparation*, 2021. 10

[33] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015. 8

[34] P.-G. Lemarié. Ondelettes à localisation exponentielle. *Journal de Mathématiques Pures et Appliquées*, 67:227–236, 1988. 25

[35] P.-G. Lemarié and Y. Meyer. Ondelettes et bases hilbertiennes. *Rev. Mat. Iberoamer.*, 1:1–18, 1986. 5

[36] M. Li, Z. Ma, Y-G Wang, and X. Zhuang. Haar transform for graph neural networks. *arxives*, 3, 2019. 10

[37] S. Mahdizadehaghdam, L. Dai, H. Krim, E. Skau, and H. Wang. Image classification: A hierarchical dictionary learning approach. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2597–2601, 2017. 9

[38] S. Mahdizadehaghdam, A. Panahi, H. Krim, and L. Dai. Deep dictionary learning: A parametric network approach. In *IEEE Transactions on Image Processing*, 2019. 9, 10

[39] H.S. Malvar. Lapped transforms for efficient transform/subband coding. *IEEE Trans. Acoust., Speech, Signal Process.*, 38(6):969–978, 1990. 3

[40] Y. Meyer. *Ondelettes et Opérateurs.* Hermann, Paris, 1990. English translation, *Wavelets and operators*, Cambridge University Press, 1992. 5, 6, 12

[41] V. Necula, S. Klimenko, and G. Mitselmakher. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Journal of Physics: conference series*, 363:012–032, 2012. 3

[42] S. Roheda and H. Krim. Conquering the cnn over-parameterization dilemma: A volterra filtering approach for action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34. AAAI, 2019. 9

[43] J.O. Strömberg. A modified franklin system and higher order spline systems on R^n as unconditional bases for Hardy spaces. In W.Beckner, editor, *Conference in honor of Antoni Zygmund, Vol. 2,*, pages 475–493. Wadsworth Math series, 1983. 5

[44] T. Tao. On the almost everywhere convergence of wavelet summation methods. *Appl. Comput. Harmon. Anal.*, 3(4), 1996. 5, 13

[45] R Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58-1:267–288, 1996. 10

[46] Kenneth Tran, Ashkan Panahi, Aniruddha Adiga, Wesam A. Sakla, and Hamid Krim. Nonlinear multi-scale super-resolution using deep learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, Brighton, United Kingdom, May 12-17, 2019*, pages 3182–3186. IEEE, 2019. 9

[47] G.G. Walter. Pointwise convergence of wavelet expansions. *J. Approx. Theory*, 80:108–118, 1995. 5, 13

[48] Y. G. Wang and X Zhuang. Tight framelets on graphs for multiscale data analysis. In *In Wavelets and Sparsity XVIII*, volume 11138, pages 100–111. SPIE, 2019. 10

[49] R. Young. *An introduction to nonharmonic Fourier series.* Academic Press, 1980. 1

[50] X. Zheng, B. Zhou, Y. G. Wang, and X. Zhuang. Decimated framelet system on graphs and fast g-framelet transforms. *arXiv:2012.06922, 2020b*, 2020. 10

Addresses:

Stéphane Jaffard: Univ Paris Est Creteil, CNRS, LAMA, F-94010 Creteil, France

Univ Gustave Eiffel, LAMA, F-77447 Marne-la-Vallée, France

`jaffard@u-pec.fr`

Hamid Krim: Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA

`ahk@ncsu.edu`