REVISTA DE LA
UNIÓN MATEMÁTICA ARGENTINA
Volumen 45, Número 1, 2004, Páginas 25–41

WAVELET APPROXIMATION METHODS IN IMAGE AND SIGNAL COMPRESSION

GUSTAVO GARRIGÓS AND EUGENIO HERNÁNDEZ *

Abstract. The material in this paper comes from various conferences given by the authors. We start with a brief survey of harmonic analysis methods in linear and non-linear approximation related to signal compression. Special emphasis is made on wavelet-based methods and some of the mathematical theory of wavelets behind them. We also present recent results of the authors concerning non-linear approximation in sequence spaces and the validity of Jackson and Bernstein inequalities in general smoothness spaces.

1. Introduction

Real world images can be mathematically described in various ways [1]. A particularly simple model considers (analog) images as non-negative functions of two variables $f(x, y)$ supported in the unit square $[0, 1]^2$, which physically may be interpreted as light intensity fields upon a given screen. Precise mathematical expressions of images are sometimes known (e.g. in fractal type designs), although often this is not the case (as in most pictures from the real world). Analog images must be "digitized" in order to be stored and manipulated by computers. We briefly describe how to produce a digital version of $f(x, y)$ (we follow [4, p. 324]): a measuring device (a photometer) averages the light intensity over small squares (of side length 2^{-M}) distributed dyadically along the picture frame $[0, 1]^2$. So if M is large (typically, $M \geq 8$), we can codify the image as a sequence of 2^{2M} coefficients:

$$p_{\mathbf{k}} = p_{\mathbf{k}}^{(M)} = \frac{1}{|I_{M,\mathbf{k}}|} \int\int_{I_{M,\mathbf{k}}} f(x, y)\, dxdy, \quad 0 \leq k_1, k_2 < 2^M, \tag{1.1}$$

where $I_{M,\mathbf{k}}$ denotes the dyadic square $[\frac{k_1}{2^M}, \frac{k_1+1}{2^M}] \times [\frac{k_2}{2^M}, \frac{k_2+1}{2^M}]$. These squares are usually called *pixels* (or *picture elements*) *located at positions* $2^{-M}\mathbf{k}$, and correspond in practice to the number of "dots" that form a computer screen. To each of them we associate a single number $p_{\mathbf{k}}$ (typically a rounded integer between 0 and 2^8),

* The material in this paper was presented by the authors at the *International Workshop in Classical Analysis and Applications* held in Yaoundé (Cameroon) in December 2001, and the *VII Encuentro Nacional de Analistas "Alberto P. Calderón"y Primera Reunión Hispano-Argentina de Análisis* held in Villa de Merlo (Argentina) in August 2004. The authors thank the people and institutions promoting and supporting these conferences. Authors supported by the European Network "HARP 2002-2006" and the Spanish project "MTM2004-0678, MEC, (Spain)". First author also supported by *Programa Ramón y Cajal*, MCyT (Spain).

2000 Mathematical Subject Classification: 42A16, 42C40, 41A15.

Key words and phrases: linear approximation, non-linear approximation, wavelet, Sobolev spaces, Besov space.

which represents the "color level" of the picture at that point. In this way we have converted $f(x,y)$ into a "digital image" $\{p_{\mathbf{k}}\}$, a sequence of "bits" which can be stored and processed by computers.

For the mathematical model, this process may be reversed. Given a sequence of bits $\{p_{\mathbf{k}}\}$, we construct a so-called *observed image* $f^{(o)}(x,y)$ by:

$$f^{(o)}(x,y) = \sum_{\mathbf{k}} p_{\mathbf{k}}\, \phi_{M,\mathbf{k}}(x,y), \tag{1.2}$$

where $\phi_{M,\mathbf{k}}(\mathbf{x}) = \phi(2^M\mathbf{x} - \mathbf{k})$ and the function ϕ may be simply chosen as $\chi_{[0,1]^2}$ or replaced by smoother versions such as splines or wavelet-type scaling functions. In general, when M is sufficiently large, $f^{(o)}(x,y)$ is an almost indistinguishable copy of $f(x,y)$, and thus can be identified for mathematical purposes with the original image. The compression problem then consists in representing $f^{(o)}(x,y)$ with much less than 2^{2M} coefficients without loosing the visual resemblance with the original image.

With this example in mind, we can describe a general mathematical setting for compression problems, which is based in classical approximation theory. We are given a general class of functions $\mathcal{F}$ (typically a Banach space), endowed with a metric $d_{\mathcal{F}}$, and an increasing sequence of subsets $D_N \subset \mathcal{F}, N = 1,2,\dots$ We define the *error of approximation of* $f \in \mathcal{F}$ *to* D_N by

$$\sigma(f, D_N)_{\mathcal{F}} \equiv \inf_{g \in D_N} d_{\mathcal{F}}(f,g), \qquad N = 1,2,\dots \tag{1.3}$$

Then, the following questions must be studied:

1. Decide, depending on applications, what metric $d_{\mathcal{F}}$ and what classes D_N are suitable in order to approximate functions in $\mathcal{F}$.
2. Find simple and fast algorithms to produce approximations $f_N \in D_N$ which are close to realize the infimum described in (1.3).
3. Investigate the rate of decay of the approximation error $\sigma(f, D_N)_{\mathcal{F}}$. More precisely, given a prescribed rate, say $N^{-\epsilon}$, determine the class of functions $f \in \mathcal{F}$ for which $\sigma(f, D_N)_{\mathcal{F}} \le N^{-\epsilon}$ for all $N = 1,2,\dots$

In the above example of images, one can take $\mathcal{F} = L^2\big([0,1]^2\big)$ and let D_N be a certain subset of functions with at most N non-null coefficients in the expansion (1.2) (or in a given orthogonal expansion). Then, when $N \ll 2^{2M}$ the best approximation f_N can be seen as a "compressed version" of the original image $f(x,y)$, from which we have removed the less essential information in order to speed up transmissions or reduce storage memory. Understanding the interplay between "quality" of the compressed signal and number of coefficients employed is the main point in this theory.

Of course, this setting of approximation can also be applied to other situations, such as the processing of other types of signals (music, digital TV,...) or the numerical solutions to PDE's. In this last case f_N is an approximation by a certain numerical method of the (unknown) solution f. In all these cases it is essential that the compressed signal f_N is a faithful representation of the original f, for which often we do not know a precise expression or this cannot be measured in the whole continuous range of space.

The purpose of this article is to give a brief introduction to harmonic analysis methods for compression problems. More precisely, we describe the so-called *linear and non-linear approximation methods*, both by means of wavelets and Fourier bases, showing the different roles played in each case by Sobolev and Besov spaces. In the wavelet case, the fact that these are unconditional bases for many function spaces allows to reduce matters to the study of sequence spaces. The first part of the paper is a survey of results from [15, 5, 7, 4], and the second part contains theorems from the recent papers [13, 9] and some other sources. Finally, we wish to cite [17] for a wider and deeper perspective on the mathematics underlying image processing.

2. Linear approximation on Hilbert spaces

Let $\{e_j : j = 1, 2, \dots\}$ be a fixed orthonormal basis of a Hilbert space $\mathcal{H}$. We select as approximating sets the linear subspaces $L_N = span\{e_1, \dots, e_N\}$. Then, the infimum defined by $\sigma(f, L_N)_{\mathcal{H}}$ as in (1.3) is attained by the orthogonal projection of f onto L_N, that is

$$f_N = \sum_{j=1}^{N} \langle f, e_j \rangle e_j . \tag{2.1}$$

This gives a precise estimate of the error of $f \in \mathcal{H}$:

$$\sigma(f, L_N)_{\mathcal{H}} = \|f - f_N\|_{\mathcal{H}} = \Big(\sum_{j=N+1}^{\infty} |\langle f, e_j \rangle|^2 \Big)^{1/2} . \tag{2.2}$$

Since $\sum_{j=1}^{\infty} |\langle f, e_j \rangle|^2 < \infty$, we always have $\lim_{N\to\infty} \sigma(f, L_N)_{\mathcal{H}} = 0$. The question now is to find subspaces of $\mathcal{H}$ for which the decay rate of $\sigma(f, L_N)_{\mathcal{H}}$ is prescribed.

At this point it is convenient to work with sequences since $\mathcal{H}$ is isomorphic to the sequence space ℓ_2. We denote by $\mathbf{c} = (c_j)_{j=1}^{\infty}$ a sequence of complex numbers and $\mathbf{e}_j, j = 1, 2, \dots$, the canonical basis of ℓ_2, so that $\mathbf{c} = \sum_{j=1}^{\infty} c_j \mathbf{e}_j$. Given $s \in \mathbb{R}$, we define

$$\mathfrak{h}_2^s = \mathfrak{h}_2^s(\mathbb{N}) = \Big\{ \mathbf{c} = (c_j)_{j=1}^{\infty} \ : \ \|\mathbf{c}\|_{\mathfrak{h}_2^s} = \Big(\sum_{j=1}^{\infty} j^{2s} |c_j|^2 \Big)^{1/2} < \infty \Big\}$$

Notice that $\mathfrak{h}_2^0 = \ell_2$ and $\mathfrak{h}_2^s \subset \ell_2$ if $s \geq 0$.

Theorem 2.3. *Let $s > 0$.*

(a) If $\mathbf{c} \in \mathfrak{h}_2^s$ then $\sigma(\mathbf{c}, L_N)_{\ell_2} \leq N^{-s} \|\mathbf{c}\|_{\mathfrak{h}_2^s}$, $\forall\, N = 1, 2, \dots$

(b) If $\mathbf{a} \in L_N$, then $\|\mathbf{a}\|_{\mathfrak{h}_2^s} \leq N^s \|\mathbf{a}\|_{\ell_2}$, $\forall\, N = 1, 2, \dots$

(c) If $\mathbf{c} \in \ell_2$ and $\sigma(\mathbf{c}, L_N)_{\ell_2} \leq cN^{-s}$ for some $c > 0$ and all $N = 1, 2, \dots$, then $\mathbf{c} \in \mathfrak{h}_2^{s-\epsilon}$ for all $\epsilon > 0$.

Proof: (a) If $j \geq N + 1$, we have $\frac{1}{j} < \frac{1}{N}$. Thus,

$$\sigma(\mathbf{c}, L_N) = \Big(\sum_{j=N+1}^{\infty} |c_j|^2 \Big)^{1/2} = \Big(\sum_{j=N+1}^{\infty} j^{-2s} j^{2s} |c_j|^2 \Big)^{1/2} \leq N^{-s} \|\mathbf{c}\|_{\mathfrak{h}_2^s} .$$

(b) Let $\mathbf{a} = \sum_{j=1}^{N} a_j \mathbf{e}_j \in L_N$. Since $j \le N$ we obtain

$$\|\mathbf{a}\|_{\mathfrak{h}_2^s} = \Big(\sum_{j=1}^{N} j^{2s} |a_j|^2 \Big)^{1/2} \le N^s \Big(\sum_{j=1}^{N} |a_j|^2 \Big)^{1/2} = N^s \|\mathbf{a}\|_{\ell_2} .$$

(c) For $\mathbf{c} = (c_j)_{j=1}^{\infty} \in \ell_2$ write $\mathbf{a}^n = \sum_{j=1}^{2^n} c_j \mathbf{e}_j \in L_{2^n}$. Then,

$$\mathbf{c} = \sum_{j=1}^{\infty} c_j \mathbf{e}_j = \mathbf{a}^0 + \sum_{n=1}^{\infty} (\mathbf{a}^n - \mathbf{a}^{n-1}) .$$

Using the triangle inequality and (b) we deduce

$$\|\mathbf{c}\|_{\mathfrak{h}_2^{s-\epsilon}} \le \|\mathbf{a}^0\|_{\mathfrak{h}_2^{s-\epsilon}} + \sum_{n=1}^{\infty} \|\mathbf{a}^n - \mathbf{a}^{n-1}\|_{\mathfrak{h}_2^{s-\epsilon}} \le |c_1| + \sum_{n=1}^{\infty} 2^{n(s-\epsilon)} \|\mathbf{a}^n - \mathbf{a}^{n-1}\|_{\ell_2} .$$

Since

$$\|\mathbf{a}^n - \mathbf{a}^{n-1}\|_{\ell_2} \le \|\mathbf{a}^n - \mathbf{c}\|_{\ell_2} + \|\mathbf{c} - \mathbf{a}^{n-1}\|_{\ell_2} = \sigma(\mathbf{c}, L_{2^n})_{\ell_2} + \sigma(\mathbf{c}, L_{2^{n-1}})_{\ell_2}$$

we use our hypothesis to obtain

$$\|\mathbf{c}\|_{\mathfrak{h}_2^{s-\epsilon}} \le |c_1| + \sum_{n=1}^{\infty} 2^{n(s-\epsilon)} [c 2^{-ns} + c 2^{-(n-1)s}] = |c_1| + c' \sum_{n=1}^{\infty} 2^{-n\epsilon} < \infty .$$

$\square$

Inequalities of the type (a) and (b) in Theorem 2.3 are usually called of Jackson and Bernstein type. They were first proved by these authors in the context of approximation of continuous periodic functions by trigonometric polynomials, but using L^∞ norms rather than L^2 norms (see [3] for details).

Consider now the particular case of the Hilbert space $\mathcal{H} = L^2[0,1]$, and its Fourier basis

$$e_m(t) = e^{2\pi i m t}, \quad m \in \mathbb{Z}.$$

This is an orthonormal basis for $L^2[0,1]$, so that any function f can be represented in terms of its *Fourier series*:

$$f(t) = \sum_{m \in \mathbb{Z}} \langle f , e_m \rangle \, e_m(t) = \sum_{m \in \mathbb{Z}} \hat{f}(m) \, e_m(t),$$

with convergence in the L^2-sense. In the last identity we denote by $\hat{f}$ the *Fourier transform of* f:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) \, e^{-2\pi i \omega t} \, dt,$$

where f has been extended to be zero outside $[0,1]$. Using the Fourier basis, the space $\mathfrak{h}_2^s$, $s > 0$, defined above corresponds to the *fractional (periodic) Sobolev space*

$$H_2^s[0,1] = \Big\{ f \in L^2[0,1] \ : \ |f|_{H_2^s} \equiv \big(\sum_{m \in \mathbb{Z}} |m|^{2s} |\hat{f}(m)|^2 \big)^{1/2} < \infty \Big\}. \tag{2.4}$$

When $s = k \in \mathbb{N}$, it is easy to see that $H_2^k[0,1]$ is the classical Sobolev space of functions whose derivatives $f', \dots, f^{(k)}$ belong to $L^2[0,1]$. The following restatement of Theorem 2.3 establishes a link between the smoothness of a function f and the rate of decay of its Fourier approximation:

Theorem 2.5. *Let $s > 0$ and $f \in L^2[0,1]$.*

(a) If $f \in H_2^s[0,1]$, then $\sigma(f, L_N)_{L^2} \le N^{-s}|f|_{H_2^s[0,1]}$, $\forall\, N = 1, 2, \ldots$

(b) If $\sigma(f, L_N)_{L^2} \le cN^{-s}$ for all $N = 1, 2, \ldots$, then $f \in H_2^{s-\epsilon}[0,1]$ for all $\epsilon \in (0, s)$.

Remark 2.6. We remark that $H_2^s[0,1]$ can actually be characterized as the subspace of all $f \in L^2[0,1]$ for which

$$\sum_{N=1}^{\infty} \Big(N^s \sigma(f, L_N)_{L^2} \Big)^2 \frac{1}{N} < \infty .$$

Details can be found in [15, Ch.9.1].

We can conclude from the previous theorem that Linear Fourier Approximation is a good method to analyze signals with "uniform smoothness". These can be coded using Fourier coefficients, and may be easily handled with the precise expression for the error of approximation. When dealing with digital signals, one can estimate numerically the best s for which $f \in H_2^s$ by looking at the decay rate of successive errors of approximation. Examples of signals to which this technique can be applied are, for instance, audio recordings, which are only perceived in a limited range of low frequency harmonics (typically, smaller than 20 kHz), and therefore have a reasonably high uniform smoothness over $\mathbb{R}$ [15, p. 49]. Linear Fourier Approximation is however a bad model for images, since a single discontinuity at a point will turn in a low exponent of global smoothness. For instance, if $f = \chi_{[a,b]}$ is the characteristic function of an interval in $\mathbb{R}$, then the error decay is like $N^{-\frac{1}{2}}$ (see [15, p. 380]). The representation of such signals can be largely improved by using non-linear approximation and wavelet bases.

3. Wavelet bases and local regularity

Before continuing with more modern approaches to signal compression, we spend some time describing the main features of wavelet bases. We say that a function $\psi \in L^2(\mathbb{R})$ is an *orthonormal wavelet* whenever the system formed by translating and dilating this function

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k), \quad j, k \in \mathbb{Z}, \tag{3.1}$$

is an orthonormal basis for $L^2(\mathbb{R})$.

The most classical example is the *Haar wavelet*, given by

$$\psi(x) = \chi_{[0,\frac{1}{2})} - \chi_{[\frac{1}{2},1)} = \begin{cases} 1, & \text{if } 0 \le x < \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \le x < 1. \end{cases}$$

The two main properties of the Haar system $\{\psi_{j,k}\}$, which are shared by most wavelet systems, are *time localization* and *vanishing moments*:

$$\operatorname{Supp} \psi_{j,k} \subset \left[\tfrac{k}{2^j}, \tfrac{k+1}{2^j}\right] \qquad \text{and} \qquad \int_{-\infty}^{\infty} \psi_{j,k}(t)dt = 0 = \int_{-\infty}^{\infty} \psi(t)dt. \tag{3.2}$$

It is easy to verify from these two facts that $\{\psi_{j,k}\}$ is actually an orthonormal basis of $L^2(\mathbb{R})$. We point out, however, that constructing orthonormal wavelets which are smooth and have a good decay is typically a difficult question (see [16, 11]).

If one is interested in signal analysis, this elementary example already illustrates a main feature of wavelet bases: they are excellent detectors of *local singularities*. Roughly speaking, the quantity $2^{\frac{j}{2}}\langle f\,,\,\psi_{j,k}\rangle$ subtracts the means of f over the left-half and right-half parts of the dyadic interval $I_{j,k}$. If f is very smooth, so that $f(t)\sim f(t_0)$ in a small interval around t_0, then $2^{\frac{j}{2}}\langle f\,,\,\psi_{j,k}\rangle\sim 0$ for $\psi_{j,k}$'s supported very close to t_0. On the other hand, if f has a jump at t_0, then $|2^{\frac{j}{2}}\langle f\,,\,\psi_{j,k}\rangle|$ has the size of the jump for $\psi_{j,k}$'s with a small support containing the singular point t_0.

This *zoom property* is common to all wavelet systems, constituting a major difference with Fourier systems for the detection of local singularities. We recall that singularities carry essential information of signals in many applied problems, such as the presence of edges in images. This makes wavelet bases very good tools for image processing, in detriment of Fourier bases. A general theorem which presents with more rigor the above arguments is given below. The statement is a simplified version of Theorem 9.7 in the first edition of [15].

Theorem 3.3. *Let $\alpha\in(0,1]$.*

1. *Let $f\in L^2(\mathbb{R})$ be a signal with local smoothness $Lip_\alpha(t_0)$, that is*
$$|f(t)-f(t_0)|\leq C_{t_0}\,|t-t_0|^\alpha,\quad t\in\mathbb{R}. \tag{3.4}$$
Then, the wavelet coefficients decay as:
$$|\langle f\,,\,\psi_{j,k}\rangle|\lesssim 2^{-j(\alpha+\frac12)}\,(1+|2^jt_0-k|^\alpha). \tag{3.5}$$

2. *Conversely, if for some $\varepsilon>0$ it holds*
$$|\langle f\,,\,\psi_{j,k}\rangle|\lesssim 2^{-j(\alpha+\frac12)}\,(1+|2^jt_0-k|^{\alpha-\varepsilon}),$$
then f belongs to $Lip_\alpha(t_0)$.

Proof: We sketch the proof of the first part, since it shows in a very transparent way the role of vanishing moments for the detection of local singularities:
$$\begin{aligned}
|\langle f\,,\,\psi_{j,k}\rangle| &= \Big|\int (f(t)-f(t_0))\,\psi_{j,k}(t)\,dt\Big| \\
&\lesssim \int |t-t_0|^\alpha\,|\psi_{j,k}(t)|\,dt \\
&= 2^{-\frac{j}{2}}\int |2^{-j}t+2^{-j}k-t_0|^\alpha\,|\psi(t)|\,dt \\
&\lesssim 2^{-j(\frac12+\alpha)}\textstyle\int |t|^\alpha\,|\psi(t)|\,dt\;+\;2^{-\frac{j}{2}}\,|2^{-j}k-t_0|^\alpha.
\end{aligned}$$
We observe that this proof is valid for $0<\alpha\leq 1$, while for $n-1<\alpha\leq n$ one needs a slightly different definition in (3.4) (with a Taylor polynomial of degree $n-1$, rather than just $f(t_0)$) and also more moment conditions in the wavelet
$$\int_{-\infty}^{\infty} t^\ell\,\psi(t)\,dt=0,\quad \ell=0,1,\dots,n-1.$$

Finally, about the second part, we just mention that it is a deeper theorem of S. Jaffard, where finer techniques in Harmonic Analysis involving Littlewood-Paley theory must be used (see [15, p. 173]). □

Before concluding this section we would like to mention how to construct *wavelet bases in d-dimensions.* Let $\mathcal{D}$ be the set of all dyadic cubes in $\mathbb{R}^d$ of the form $I_{j,\mathbf{k}} = 2^{-j}([0,1]^d + k)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}^d$; observe that $|I_{j,\mathbf{k}}| = 2^{-jd}$. A collection of functions $\Psi = \{\psi^1, \ldots, \psi^L\} \subset L^2(\mathbb{R}^d)$ is a *wavelet family* if the set

$$\{\psi^\ell_{I_{j,\mathbf{k}}}(x) = 2^{jd/2}\psi^\ell(2^j x - k) : I_{j,\mathbf{k}} \in \mathcal{D},\, \ell = 1, 2, \ldots, L\}$$

is an orthonormal basis for $L^2(\mathbb{R}^d)$. The reason to consider systems with finitely many generators is because in $\mathbb{R}^d$ one typically needs $L = 2^d - 1$ functions ψ^ℓ in order to obtain an orthonormal basis for $L^2(\mathbb{R}^d)$ (this is always the case when the ψ^ℓ's have sufficiently good decay). For instance, the *Haar wavelet family in* $\mathbb{R}^2$, $\Psi_H = \{\psi^1, \psi^2, \psi^3\}$, takes the form

$$\psi^1(x,y) = \phi_H(x)\psi_H(y), \quad \psi^2(x,y) = \psi_H(x)\phi_H(y), \quad \psi^3(x,y) = \psi_H(x)\psi_H(y),$$

where $\phi_H = \chi_{[0,1)}$ and $\psi_H = \chi_{[0,1/2)} - \chi_{[1/2,1)}$ are the Haar scaling function and wavelet respectively. We refer to [15, Ch. 7.7] for other examples of wavelet families in d-dimensions constructed as tensor products of 1-dimensional wavelets.

4. **Wavelet bases, Lebesgue and Sobolev spaces**

Under certain decay and smoothness conditions, wavelet bases provide characterizations of classical Lebesgue and Sobolev spaces. For $p \in (1, \infty)$ (and wavelets with enough decay) it is known that

$$\|f\|_{L_p(\mathbb{R}^d)} \approx \Big\| \Big[\sum_{\ell=1}^{L} \sum_{I \in \mathcal{D}} \big(|I|^{-\frac{1}{2}} |\langle f\,,\, \psi^\ell_I \rangle| \chi_I(\cdot)\big)^2 \Big]^{1/2} \Big\|_{L_p(\mathbb{R}^d)}. \tag{4.1}$$

Equivalence (4.1) reduces to Plancherel theorem for $p = 2$ and it is essentially a Littlewood-Paley characterization of Lebesgue spaces for $p \neq 2$. Conditions on $\Psi = \{\psi^1, \ldots, \psi^L\}$ for which (4.1) holds can be found, for instance, in [16, 11] or [20]. A collection $\Psi = \{\psi^1, \ldots, \psi^L\}$ of functions in $L_2(\mathbb{R}^d)$ for which (4.1) holds will be called *admissible wavelet family* for $L_p(\mathbb{R})$.

When $p \in (1, \infty)$ and $s > 0$, Sobolev spaces on $\mathbb{R}^d$ can be defined in analogy to (2.4) as:

$$H^s_p(\mathbb{R}^d) = \Big\{ f \in L_p(\mathbb{R}^d) \;:\; \mathcal{F}^{-1}[(1 + |\xi|^2)^{s/2} \hat{f}(\xi)] \in L^p(\mathbb{R}^d) \Big\}. \tag{4.2}$$

They also have a characterization using wavelets. When ψ^ℓ have enough smoothness and decay it is known that

$$|f|_{H^s_p(\mathbb{R}^d)} \approx \Big\| \Big[\sum_{\ell=1}^{L} \sum_{I \in \mathcal{D}} \big(|I|^{-\frac{s}{d}-\frac{1}{2}} |\langle f\,,\, \psi^\ell_I \rangle| \chi_I(\cdot)\big)^2 \Big]^{1/2} \Big\|_{L_p(\mathbb{R}^d)}. \tag{4.3}$$

We refer to [16, 11, 14] for conditions under which the collection Ψ is an admissible wavelet family for $H^s_p(\mathbb{R}^d)$.

The characterizations (4.1) and (4.3) allow us to work with sequence spaces when we study the problem of approximation. When $p > 0$ and $s \in \mathbb{R}$ we define the sequence space $\mathfrak{h}_p^s = \mathfrak{h}_s^p(\mathcal{D})$ as the set of all sequences of complex numbers $\mathbf{c} = \{c_I\}_{I\in\mathcal{D}}$ for which

$$\|\mathbf{c}\|_{\mathfrak{h}_p^s} = \Big\| \Big[\sum_{I\in\mathcal{D}} \big(|I|^{-\frac{s}{d}-\frac{1}{2}} |c_I| \chi_I(\cdot) \big)^2 \Big]^{1/2} \Big\|_{L_p(\mathbb{R}^d)} < \infty. \tag{4.4}$$

With this notation, for admissible wavelets, we have

$$|f|_{H_p^s(\mathbb{R}^d)} \approx \sum_{\ell=1}^{L} \big\| \{ \langle f , \psi_I^\ell \rangle \}_{I\in\mathcal{D}} \big\|_{\mathfrak{h}_p^s}, \quad \forall\, p \in (1,\infty),\ \forall\, s \geq 0\,. \tag{4.5}$$

5. **Linear multiresolution approximation**

Using the characterization of Lebesgue and Sobolev spaces in section 4 we can describe very simply the linear approximation with wavelet bases when the errors are measured in the norm of $L_p(\mathbb{R}^d)$ and $p \in (1,\infty)$ is fixed. If $\Psi = \{\psi^1, \dots, \psi^L\}$ is an admissible wavelet family, then a natural choice for the linear approximating subspaces is:

$$V_n = \overline{\operatorname{span}}_{L^p(\mathbb{R}^d)} \Big\{ \psi_{j,\mathbf{k}}^\ell \ :\ j \leq n,\ \mathbf{k} \in \mathbb{Z}^d,\ \ell = 1, \dots, L \Big\}, \qquad n = 0, 1, 2 \dots,$$

Observe that in V_n are only involved wavelets ψ_I^ℓ with "resolutions" $|I| \geq 2^{-nd}$. In wavelet theory, the collection of spaces $\{V_n\}_{n\in\mathbb{Z}}$ is called a "multiresolution analysis", so we shall refer to this kind of approximation as *linear multiresolution approximation*.

As in Hilbert spaces, the orthogonal projection into V_n gives a good candidate for best approximation of $f \in L_p(\mathbb{R}^d)$. More precisely, let

$$P_n(f) = \sum_{\ell=1}^{L} \sum_{j\leq n} \sum_{\mathbf{k}\in\mathbb{Z}^d} \langle f , \psi_{j,\mathbf{k}}^\ell \rangle \psi_{j,\mathbf{k}}^\ell\,. \tag{5.1}$$

When $p \in (1,\infty)$, the characterization of Lebesgue spaces given in (4.1) guarantees the existence of a constant $A_p > 0$ such that

$$\|P_n(f)\|_{L_p(\mathbb{R}^d)} \leq A_p\, \|f\|_{L_p(\mathbb{R}^d)}, \quad \forall\, n = 0, 1, 2, \dots,\ \forall\, f \in L_p(\mathbb{R}^d),$$

and moreover that the series in (5.1) converges in $L^p(\mathbb{R}^d)$. We denote the errors of approximation of f by $P_n(f)$ and V_n by

$$E(P_n(f))_{L_p} = \|f - P_n(f)\|_{L_p} \quad \text{ and } \quad \sigma(f, V_n)_{L_p} = \inf_{g\in V_n} \|f - g\|_{L_p}\,.$$

Then we claim that there exists $c_p > 0$ such that

$$\sigma(f, V_n)_{L_p(\mathbb{R}^d)} \leq E(P_n(f))_{L_p(\mathbb{R}^d)} \leq c_p\, \sigma(f, V_n)_{L_p(\mathbb{R}^d)} \tag{5.2}$$

for all $n = 0, 1, 2, \dots$ and all $f \in L_p(\mathbb{R}^d)$. The left-hand inequality of (5.2) follows from $P_n f \in V_n$. To see the right-hand side, take any $\varepsilon > 0$ and choose a function

$g_{n,\varepsilon} \in V_n$ such that $\|f - g_{n,\varepsilon}\|_{L_p} \le \sigma(f, V_n)_{L_p} + \varepsilon$. Then

$$\begin{aligned}\|f - P_n(f)\|_{L_p(\mathbb{R}^d)} &\le \|f - g_{n,\varepsilon}\|_{L_p(\mathbb{R}^d)} + \|g_{n,\varepsilon} - P_n(f)\|_{L_p(\mathbb{R}^d)} \\ &\le \|f - g_{n,\varepsilon}\|_{L_p(\mathbb{R}^d)} + A_p \, \|g_{n,\varepsilon} - f\|_{L_p(\mathbb{R}^d)} \\ &\le (1 + A_p)\, (\sigma(f, V_n)_{L_p(\mathbb{R}^d)} + \varepsilon)\,,\end{aligned}$$

which proves the right-hand side. In view of (5.2) the compression problems stated in the introduction can be treated, for multiresolution approximation, using either $\sigma(f, V_n)$ or $E(P_n(f))$.

Using the terminology of sequence spaces in section 4, we can prove the following theorem in analogy to Theorem 2.3 for Hilbert spaces.

Theorem 5.3. *Let $1 < p < \infty$ and $s > 0$.*

(a) If $\mathbf{c} \in \mathfrak{h}_p^s$ then $\sigma(\mathbf{c}, V_n)_{\mathfrak{h}_p^0} \le E(P_n(\mathbf{c}))_{\mathfrak{h}_p^0} \le 2^{-ns} \|\mathbf{c}\|_{\mathfrak{h}_p^s}$, $\forall\, n = 0, 1, 2 \ldots$

(b) If $\mathbf{a} \in V_n$ then $\|\mathbf{a}\|_{\mathfrak{h}_p^s} \le 2^{ns} \|\mathbf{a}\|_{\mathfrak{h}_p^0}$, $\forall\, n = 0, 1, 2, \ldots$

(c) If $\mathbf{c} \in \mathfrak{h}_p^0$ and $\sigma(\mathbf{c}, V_n)_{\mathfrak{h}_p} \le C 2^{-ns}$ for all $n = 0, 1, 2, \ldots,$ then $\mathbf{c} \in \mathfrak{h}_p^{s-\epsilon}$ for all $0 < \epsilon < s$.

Proof: (a) This follows easily from

$$\begin{aligned}\|\mathbf{c} - P_n(\mathbf{c})\|_{\mathfrak{h}_p^0} &= \Big\| \Big[\sum_{|I| < 2^{-nd}} \big(|I|^{-\frac{1}{2}} |c_I| \chi_I(\cdot) \big)^2 \Big]^{1/2} \Big\|_{L_p(\mathbb{R}^d)} \\ &\le 2^{-ns} \Big\| \Big[\sum_{|I| < 2^{-nd}} \big(|I|^{-\frac{s}{d} - \frac{1}{2}} |c_I| \chi_I(\cdot) \big)^2 \Big]^{1/2} \Big\|_{L_p(\mathbb{R}^d)}.\end{aligned}$$

(b) If $\mathbf{a} = \sum_{|I| \ge 2^{-nd}} a_I \mathbf{e}_I \in V_n$, the we have

$$\|\mathbf{a}\|_{\mathfrak{h}_p^s} = \Big\| \Big[\sum_{|I| \ge 2^{-nd}} \big(|I|^{-\frac{s}{d} - \frac{1}{2}} |c_I| \chi_I(\cdot) \big)^2 \Big]^{1/2} \Big\|_{L_p(\mathbb{R}^d)} \le 2^{ns} \|\mathbf{a}\|_{\mathfrak{h}_p^0}\,.$$

(c) For $\mathbf{c} = (c_I) \in \mathfrak{h}_p^0$ write $\mathbf{a}^n = P_n(\mathbf{c}) = \sum_{|I| \ge 2^{-nd}} c_I \mathbf{e}_I \in V_n$, $n = 0, 1, 2, \ldots$ Then, it is easily seen that

$$\mathbf{c} = \mathbf{a}^0 + \sum_{n=1}^{\infty} (\mathbf{a}^n - \mathbf{a}^{n-1})\,,$$

at least with pointwise convergence. The proof now follows the same lines as in part (c) of Theorem 2.3 with obvious modifications (that includes the use of (5.2)). □

The theorem we just proved together with the equivalences in (4.5) allow us to conclude the following result:

Theorem 5.4. *If $p \in (1, \infty)$ and $s > 0$, then there exists $c > 0$ such that for all $f \in H_p^s(\mathbb{R}^d)$*

$$\sigma(f, V_n)_{L_p(\mathbb{R}^d)} \le c\, 2^{-ns} |f|_{H_p^s(\mathbb{R}^d)}, \qquad n = 0, 1, 2, \ldots$$

Conversely, if $\sigma(f, V_n)_{L_p(\mathbb{R}^d)} \le c\, 2^{-ns}$ for some $c > 0$ and all $n = 0, 1, 2, \ldots,$ then we deduce $f \in H_p^{s-\varepsilon}(\mathbb{R}^d)$ for all $0 < \varepsilon < s$.

Finally observe that, when the wavelets ψ^ℓ are compactly supported and f has support in $[0,1]^d$ we can write

$$f = P_0 f + \sum_{\ell=1}^{L} \sum_{j\geq 1} \sum_{|\mathbf{k}|\leq c2^j} \langle\, f\,,\, \psi^\ell_{j,\mathbf{k}}\,\rangle \psi^\ell_{j,\mathbf{k}}.$$

Therefore, the approximation from V_n only takes about $N = 2^{nd}$ coefficients from the wavelet expansion (together with $P_0 f$), with a decay error of the order of $N^{-s/d}$. Thus, the use of wavelet bases for linear approximation does not seem to improve the results we obtained with the Fourier basis in §2 (see [15, p.381] for explicit examples). In particular, if $f = \chi_R$ with $R = [a,b]\times[c,d] \subset \mathbb{R}^2$, it is not difficult to see that $f \in H^s_2(\mathbb{R}^2)$ if and only if $s < 1/2$. Hence the linear error of approximation in $L_2(\mathbb{R}^2)$ for this function decays like $N^{-(\frac{1}{4}-\varepsilon)}$ for any $\varepsilon > 0$. [1]

6. **Besov spaces and wavelets**

In the same way as Sobolev spaces arise naturally from linear approximation methods, non-linear approximation leads to a different type of smoothness classes: the *Besov spaces*. In the classical setting these can be defined as follows. We use the notation $\Delta_h f(x) = f(x+h) - f(x)$ and $\Delta^k_h f = \Delta_h(\Delta^{k-1}_h f)$.

Definition 6.1. *If $\alpha > 0$ and $0 < \tau, q \leq \infty$, the Besov space $B^\alpha_{\tau,q}(\mathbb{R}^d)$ is the set of all $f \in L_\tau(\mathbb{R}^d)$ for which*

$$|f|_{B^\alpha_{\tau,q}} := \sum_{i=1}^{d} \Big[\int_0^\infty \Big(\frac{\|\Delta^{[\alpha]+1}_{t\mathbf{e}_i} f\|_\tau}{t^\alpha} \Big)^q \frac{dt}{t} \Big]^{\frac{1}{q}} < \infty.$$

There exist other definitions of Besov spaces using Littlewood-Paley theory, which coincide with the previous one at least when $\alpha > d(1/\tau - 1)_+$ (see [18]), as well as extensions of these definitions to all $\alpha \in \mathbb{R}$ and even to anisotropic smoothness parameters. We observe that in non-linear approximation it will be common to use Besov spaces with indices $\tau < 1$.

It is well known in wavelet theory that, at least when $\alpha > d(1/\tau - 1)_+$, under sufficient smoothness and decay in the wavelet family Ψ one has

$$|f|_{B^\alpha_{\tau,q}} \approx \Big[\sum_{\ell=1}^{L} \sum_{j\in\mathbb{Z}} \Big(\sum_{k\in\mathbb{Z}^d} \big(|I_{j,\mathbf{k}}|^{-\frac{\alpha}{d}+\frac{1}{\tau}-\frac{1}{2}} |\langle\, f\,,\, \psi^\ell_{I_{j,\mathbf{k}}}\,\rangle| \big)^\tau \Big)^{\frac{q}{\tau}} \Big]^{\frac{1}{q}} < \infty\,. \tag{6.2}$$

We refer to [16, 2, 10] for conditions under which such characterizations hold.

As in the case of Lebesgue and Sobolev spaces, (6.2) will allow us to work with sequence spaces to study the problems of approximation. When $\alpha \in \mathbb{R}$ and $0 < \tau, q \leq \infty$, define the sequence space $\mathfrak{b}^\alpha_{\tau,q} = \mathfrak{b}^\alpha_{\tau,q}(\mathcal{D})$ as the set of all complex sequences

[1] Actually, in this example, one can remove the "ε" (see exercise 9.5 of [15, p.431].

$\mathbf{c} = \{c_I\}_{I\in\mathcal{D}}$ such that

$$\|\mathbf{c}\|_{\mathfrak{b}^\alpha_{\tau,q}} \equiv \Big[\sum_{j\in\mathbb{Z}}\Big(\sum_{|I|=2^{-jd}} \big(|I|^{-\frac{\alpha}{d}+\frac{1}{\tau}-\frac{1}{2}}|c_I|\big)^\tau\Big)^{\frac{q}{\tau}}\Big]^{\frac{1}{q}} < \infty\,, \tag{6.3}$$

so that, for admissible wavelets,

$$|f|_{B^\alpha_{\tau,q}(\mathbb{R}^d)} \approx \sum_{\ell=1}^{L} \Big\| \{\langle\, f\,,\,\psi^\ell_I\,\rangle\}_{I\in\mathcal{D}}\Big\|_{\mathfrak{b}^\alpha_{\tau,q}}. \tag{6.4}$$

It is interesting to observe that, when $\tau = q$ these spaces are isomorphic to $\ell_\tau(\mathcal{D})$. In fact the quasi-norm of $\mathfrak{b}^\alpha_{\tau,\tau}$ can be written as

$$\|\mathbf{c}\|_{\mathfrak{b}^\alpha_{\tau,\tau}} = \Big[\sum_{I\in\mathcal{D}} \big(|I|^{-\frac{\alpha}{d}+\frac{1}{\tau}-\frac{1}{2}}|c_I|\big)^\tau\Big]^{\frac{1}{\tau}} = \Big[\sum_{I\in\mathcal{D}} \|c_I\mathbf{e}_I\|^\tau_{\mathfrak{b}^\alpha_{\tau,\tau}}\Big]^{\frac{1}{\tau}}. \tag{6.5}$$

Finally we point out a main difference between Sobolev and Besov spaces. When $\alpha \in [\frac{1}{2}, \frac{d}{2(d-1)})$, the space $B^\alpha_{\tau,\tau}(\mathbb{R}^d)$ with $\frac{1}{\tau} = \frac{\alpha}{d} + \frac{1}{2}$ contains characteristic functions of sets while this cannot belong to $H^\alpha_2(\mathbb{R}^d)$ (see §7 bellow for details). This makes Besov spaces more suitable than Sobolev spaces for image processing.

7. **Non-linear wavelet approximation**

We return to the problem of wavelet approximation, this time using a non-linear approach. We fix $p \in (1,\infty)$ and a wavelet family $\Psi = \{\psi^1,\ldots,\psi^L\}$ which we assume admissible for $L_p(\mathbb{R}^d)$. For simplicity we just let $L = 1$. We define the following approximating sets

$$\Sigma_N = \Big\{ f = \sum_{I\in\Lambda} c_I\psi_I \;:\; \Lambda\subset\mathcal{D} \text{ with Card } \Lambda \le N,\; c_I\in\mathbb{C}\Big\}, \quad N = 0,1,\ldots$$

Observe that $\Sigma_N + \Sigma_N \subset \Sigma_{2N}$, and that Σ_N is not a linear space. This kind of non-linear approximation is called *N-term approximation*.

An algorithm to produce good approximations by Σ_N is to consider the N basis coefficients of f with largest absolute values. More precisely, if $f = \sum_{I\in\mathcal{D}} \langle\, f\,,\,\psi_I\,\rangle\psi_I \in L_p(\mathbb{R}^d)$ and we consider the non-increasing rearrangement

$$\|\langle\, f\,,\,\psi_{I_1}\,\rangle\psi_{I_1}\|_{L_p(\mathbb{R}^d)} \ge \|\langle\, f\,,\,\psi_{I_2}\,\rangle\psi_{I_2}\|_{L_p(\mathbb{R}^d)} \ge \ldots, \tag{7.1}$$

then the so-called **greedy algorithm** is defined by

$$f \mapsto G_N(f) = \sum_{i=1}^{N} \langle\, f\,,\,\psi_{I_i}\,\rangle\psi_{I_i} \in \Sigma_N\,. \tag{7.2}$$

A theorem due to V. Temlyakov [19] shows that, when $p \in (1,\infty)$, there exists $c_p > 0$ such that

$$\|f - G_N(f)\|_{L_p(\mathbb{R}^d)} \le c_p\,\sigma(f,\Sigma_N)_{L_p}, \quad \forall\, N = 1,2,\ldots,\ \forall\, f\in L_p(\mathbb{R}^d). \tag{7.3}$$

Thus, for the purposes of N-term approximation in $L^p(\mathbb{R}^d)$ we may use indistinctly $\sigma(f,\Sigma_N)_{L_p}$ or $E(G_N(f))_{L_p} := \|f - G_N(f)\|_{L_p(\mathbb{R}^d)}$. An elementary proof of (7.3) using the associated sequence spaces $\mathfrak{h}_p^0$ can be found in [12, 9] (see Theorem 2.1 and Proposition 3.2 in [9]).

Taking into account these facts, we wish to prove Jackson and Bernstein type inequalities for non-linear approximation with Σ_N. The proofs are somewhat more difficult than in the linear case. All the results will be stated and proved in the setting of sequence spaces. A key step will be the following proposition, whose proof can be found in [12, 9].

Proposition 7.4. *Let $p \in (1,\infty)$. Then there exists a constant $c_p > 0$ so that for all finite sets $\Gamma \subset \mathcal{D}$*

$$c_p^{-1} (Card\,\Gamma)^{\frac{1}{p}} \leq \Big\| \sum_{I\in\Gamma} \frac{\mathbf{e}_I}{\|\mathbf{e}_I\|_{\mathfrak{h}_p^0}} \Big\|_{\mathfrak{h}_p^0} \leq c_p (Card\,\Gamma)^{\frac{1}{p}} . \tag{7.5}$$

Proposition 7.6. Jackson's inequality

Let $1 < p < \infty$ and $0 < \tau < p$. Choose $\alpha > 0$ such that $\frac{\alpha}{d} = \frac{1}{\tau} - \frac{1}{p}$. Then, for all $\mathbf{s} \in \mathfrak{b}^{\alpha}_{\tau,\tau}$ we have,

$$\sigma(\mathbf{s}, \Sigma_N)_{\mathfrak{h}_p^0} \leq E(G_N(\mathbf{s}))_{\mathfrak{h}_p^0} \leq c\, N^{-\frac{\alpha}{d}} \|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}} , \quad N = 1, 2, 3, \ldots$$

Proof: Consider the non-increasing order

$$\|s_{I_1}\mathbf{e}_{I_1}\|_{\mathfrak{h}_p^0} \geq \|s_{I_2}\mathbf{e}_{I_2}\|_{\mathfrak{h}_p^0} \geq \|s_{I_3}\mathbf{e}_{I_3}\|_{\mathfrak{h}_p^0} \geq \ldots \tag{7.7}$$

By the triangle inequality we can write

$$E(G_N(\mathbf{s}))_{\mathfrak{h}_p^0} = \|\mathbf{s} - G_N(\mathbf{s})\|_{\mathfrak{h}_p^0} \leq \sum_{j=0}^{\infty} \Big\| \sum_{N2^j < k \leq N2^{j+1}} s_{I_k}\mathbf{e}_{I_k} \Big\|_{\mathfrak{h}_p^0} . \tag{7.8}$$

The non-increasing ordering (7.7) implies that $\|s_{I_k}\mathbf{e}_{I_k}\|_{\mathfrak{h}_p^0} \leq \|s_{I_{N2^j}}\mathbf{e}_{I_{N2^j}}\|_{\mathfrak{h}_p^0}$ for all k such that $N2^j < k \leq N2^{j+1}$. Hence,

$$\Big\| \sum_{N2^j < k \leq N2^{j+1}} s_{I_k}\mathbf{e}_{I_k} \Big\|_{\mathfrak{h}_p^0} \leq \Big\| s_{I_{N2^j}}\mathbf{e}_{I_{N2^j}} \Big\|_{\mathfrak{h}_p^0} \Big\| \sum_{N2^j < k \leq N2^{j+1}} \frac{\mathbf{e}_{I_k}}{\|\mathbf{e}_{I_k}\|_{\mathfrak{h}_p^0}} \Big\|_{\mathfrak{h}_p^0} .$$

Using the right hand side inequality in Proposition 7.4 we obtain

$$\Big\| \sum_{N2^j < k \leq N2^{j+1}} s_{I_k}\mathbf{e}_{I_k} \Big\|_{\mathfrak{h}_p^0} \leq c_p \Big\| s_{I_{N2^j}}\mathbf{e}_{I_{N2^j}} \Big\|_{\mathfrak{h}_p^0} (N2^j)^{\frac{1}{p}}$$

Substituting this result in (7.8) we deduce

$$E(G_N(\mathbf{s}))_{\mathfrak{h}_p^0} \leq c_p \sum_{j=0}^{\infty} \|s_{I_{N2^j}}\mathbf{e}_{I_{N2^j}}\|_{\mathfrak{h}_p^0} (N2^j)^{\frac{1}{p}} \tag{7.9}$$

Now, observe that by the non-increasing ordering in (7.7) we have

$$\|s_{I_{N2^j}}\mathbf{e}_{I_{N2^j}}\|_{\mathfrak{h}_p^0} \leq \Big(\frac{1}{N2^j} \sum_{k=1}^{N2^j} \|s_{I_k}\mathbf{e}_{I_k}\|^{\tau}_{\mathfrak{h}_p^0} \Big)^{\frac{1}{\tau}},$$

and also that for all $I \in \mathcal{D}$

$$\|s_I\mathbf{e}_I\|_{\mathfrak{h}_p^0} = |I|^{\frac{1}{p}-\frac{1}{2}} |s_I| = |I|^{-\frac{\alpha}{d}+\frac{1}{\tau}-\frac{1}{2}} |s_I| = \|s_I\mathbf{e}_I\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}} . \tag{7.10}$$

Thus, using the observation in (6.5) we can write

$$\|s_{I_{N2^j}}\mathbf{e}_{I_{N2^j}}\|_{\mathfrak{h}_p^0} \leq (N2^j)^{-\frac{1}{\tau}}\Big(\sum_{k=1}^{\infty}\|s_{I_k}\mathbf{e}_{I_k}\|^{\tau}_{\mathfrak{b}^{\alpha}_{\tau,\tau}}\Big)^{\frac{1}{\tau}} = (N2^j)^{-\frac{1}{\tau}}\|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}}\,.$$

Substituting in (7.9) we obtain,

$$E(G_N(\mathbf{s}))_{\mathfrak{h}_p^0} \leq\ c_p\,\|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}} N^{\frac{1}{p}-\frac{1}{\tau}}\sum_{j=0}^{\infty}2^{-j\frac{\alpha}{d}}\ =\ c'\,N^{-\frac{\alpha}{d}}\|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}}\,,$$

which establishes the proposition. □

Our next result is a Bernstein type inequality in the context of non-linear approximation with Σ_N.

Proposition 7.11. Bernstein's inequality
Let $1 < p < \infty$ and $0 < \tau < p$. Choose $\alpha > 0$ such that $\frac{\alpha}{d} = \frac{1}{\tau} - \frac{1}{p}$. Then,

$$\|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}} \leq CN^{\frac{\alpha}{d}}\|s\|_{\mathfrak{h}_p^0}$$

for all $\mathbf{s} \in \Sigma_N$ and all $N = 1, 2, 3, \ldots$

Proof: Let $\mathbf{s} = \sum_{k=1}^{N} s_{I_k}\mathbf{e}_{I_k} \in \Sigma_N$, where we can assume that

$$\|s_{I_1}\mathbf{e}_{I_1}\|_{\mathfrak{h}_p^0} \geq \|s_{I_2}\mathbf{e}_{I_2}\|_{\mathfrak{h}_p^0} \geq \cdots \geq \|s_{I_N}\mathbf{e}_{I_N}\|_{\mathfrak{h}_p^0}\,. \tag{7.12}$$

By (6.5) and (7.10) we can write

$$\|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}} = \Big(\sum_{k=1}^{N}\|s_{I_k}\mathbf{e}_{I_k}\|^{\tau}_{\mathfrak{b}^{\alpha}_{\tau,\tau}}\Big)^{\frac{1}{\tau}} = \Big(\sum_{k=1}^{N}\|s_{I_k}\mathbf{e}_{I_k}\|^{\tau}_{\mathfrak{h}_p^0}\Big)^{\frac{1}{\tau}}\,.$$

The ordering (7.12) together with the left hand side inequality that appears in Proposition 7.4 gives

$$\begin{aligned}\|\mathbf{s}\|_{\mathfrak{b}^{\alpha}_{\tau,\tau}} &\leq c_p\Big[\sum_{k=1}^{N}\Big(k^{-\frac{1}{p}}\,\Big\|\sum_{\ell=1}^{k}s_{I_\ell}\mathbf{e}_{I_\ell}\Big\|_{\mathfrak{h}_p^0}\Big)^{\tau}\Big]^{\frac{1}{\tau}}\\ &\leq c_p\,\|\mathbf{s}\|_{\mathfrak{h}_p^0}\Big(\sum_{k=1}^{N}k^{-\frac{\tau}{p}}\Big)^{\frac{1}{\tau}}\ =\ c'N^{\frac{\alpha}{d}}\|\mathbf{s}\|_{\mathfrak{h}_p^0}\,.\end{aligned}$$

□

Proposition 7.13. *Let $1 < p < \infty$ and $\alpha > 0$. If for $\mathbf{s} \in \mathfrak{h}_p^0$ we have $\sigma(\mathbf{s}, \Sigma_N)_{\mathfrak{h}_p^0} \leq CN^{-\frac{\alpha}{d}}$ for all $N = 1, 2, \ldots,$ then $\mathbf{s} \in \mathfrak{b}^{\beta}_{t,t}$, for all $\beta \in (0, \alpha)$ provided $\frac{1}{t} = \frac{\beta}{d} + \frac{1}{p}$.*

Proof: For $\mathbf{s} \in \mathfrak{h}_p^0$ write $\mathbf{b}^n = G_{2^n}(\mathbf{s}) \in \Sigma_{2^n}$, $n = 0, 1, 2, \ldots$ Then,

$$\mathbf{s} = \mathbf{b}^0 + \sum_{k=1}^{\infty}(\mathbf{b}^n - \mathbf{b}^{n-1})\,.$$

The proof now follows the same lines as the proof of part (c) in Theorem 2.3 with some obvious modifications, which includes the use of (7.3). □

Corollary 7.14. *Let $1 < p < \infty$ and $0 < \tau < p$. Choose $\alpha > 0$ such that $\frac{\alpha}{d} = \frac{1}{\tau} - \frac{1}{p}$. Then, there exists $C > 0$ such that for all $f \in B^{\alpha}_{\tau,\tau}(\mathbb{R}^d)$,*

$$\sigma(f, \Sigma_N)_{L_p(\mathbb{R}^d)} \leq CN^{-\frac{\alpha}{d}} |f|_{B^{\alpha}_{\tau,\tau}(\mathbb{R}^d)}\,, \quad N = 1, 2, 3, \dots$$

Moreover, if for $f \in L_p(\mathbb{R}^d) \cap L_\tau(\mathbb{R}^d)$ and we have $\sigma(f, \Sigma_N)_{L_p(\mathbb{R}^d)} \leq CN^{-\frac{\alpha}{d}}$ for all $N = 1, 2, \dots,$ then $f \in B^{\beta}_{t,t}(\mathbb{R}^d)$ for all $\beta \in (0, \alpha)$ provided $\frac{1}{t} = \frac{\beta}{d} + \frac{1}{p}$.

In particular, when $f = \chi_R$, for R a "rectangle" in $\mathbb{R}^d$, it is not difficult to see that $f \in B^{\alpha}_{\tau,\tau}(\mathbb{R}^d)$ when $d(\frac{1}{\tau} - 1)_+ < \alpha < \frac{1}{\tau}$. Therefore, for non-linear approximation in $L_2(\mathbb{R}^d)$ we have $\sigma(f, \Sigma_N)_{L_2(\mathbb{R}^d)} \lesssim N^{-\alpha/d}$ for all $\alpha \in (0, \frac{d}{2(d-1)})$. When $d = 1$, this improves dramatically the power $N^{-1/2}$ obtained in §2 and §5 with linear approximation. When $d = 2$, the non-linear approximation in $L_2(\mathbb{R}^2)$ gives the bound $N^{-(\frac{1}{2}-\varepsilon)}$, compared to $N^{-\frac{1}{4}}$ obtained in §5 for linear approximation. Also, in this particular example, and more generally for all functions with *bounded variation*[1] $f \in BV(\mathbb{R}^2)$, it can be proved that $\sigma(f, \Sigma_N)_{L_2(\mathbb{R})} \lesssim N^{-\frac{1}{2}} |f|_{BV(\mathbb{R}^2)}$ (see [15, p.402] and [6]).

8. **Further results**

In the previous sections we have measured the error of approximation in the norm of $L_p(\mathbb{R}^d)$ for $p \in (1, \infty)$. Nevertheless, sometimes it is interesting to use other norms which keep more features of a given image. One case which has been proposed by other authors is the use of Sobolev or Besov norms (see [5, Ch. 4] or [12]).

In our paper [9] we do something more general by studying the approximation problems associated with the family of *Triebel-Lizorkin spaces* $F^s_{p,q}(\mathbb{R}^d)$. These can be defined in terms of Littlewood-Paley theory as follows. Let $\varphi \in \mathcal{S}(\mathbb{R}^d)$ be such that $\varphi(x) = 1$ when $|x| \leq 1$ and $\varphi(x) = 0$ when $|x| \geq 3/2$. For $k \in \mathbb{Z}$, define $\varphi_k(x) = \varphi(2^{-k}x) - \varphi(2^{-(k-1)}x)$, so that $\sum_{k=-\infty}^{\infty} \varphi_k(x) = 1$ for all $x \in \mathbb{R}^d \setminus \{0\}$.

Definition 8.1. *For $s \in \mathbb{R}, 0 < p < \infty$ and $0 < r \leq \infty$ the* **Triebel-Lizorkin** *space $F^s_{p,r}(\mathbb{R}^d)$ is defined as the set of all $f \in L_p(\mathbb{R}^d)$ such that*

$$|f|_{F^s_{p,r}} = \Big\| \Big(\sum_{j=-\infty}^{\infty} |2^{js} |(\varphi_j \widehat{f})^\vee|^r \Big)^{1/r} \Big\|_{L_p} < \infty\,. \tag{8.2}$$

There is a characterization of the Triebel-Lizorkin spaces using wavelets. Under suitable decay and smoothness conditions on a wavelet family $\Psi = \{\psi_1, \dots, \psi_L\}$ (see [14]), a function $f \in L_p(\mathbb{R}^d)$ belongs to $F^s_{p,r}$ if and only if

$$\Big\| \Big[\sum_{\ell=1}^{L} \sum_{I \in \mathcal{D}} \big(|I|^{-\frac{s}{d}-\frac{1}{2}} |\langle\, f\,,\, \psi^\ell_I \,\rangle| \chi_I(\cdot) \big)^r \Big]^{1/r} \Big\|_{L_p} < \infty\,, \tag{8.3}$$

and $|f|_{F^s_{p,r}}$ is equivalent to the expression in (8.3). We remark that when $1 < p < \infty$, $F^0_{p,2} = L_p$ and $F^s_{p,2} = H^s_p$, while for any p we have $F^s_{p,p} = B^s_{p,p}$.

[1] This space satisfies $BV(\mathbb{R}^2) \subset B^{1-\varepsilon}_{1,1}(\mathbb{R}^2)$, $\forall\, \varepsilon > 0$.

As in the case of Besov spaces, (8.3) will allow us to work with sequence spaces when studying problems of approximation. For $s \in \mathbb{R}, 0 < p < \infty$ and $0 < r \leq \infty$, we define the sequence space $\mathfrak{f}^s_{p,r} \equiv \mathfrak{f}^s_{p,r}(\mathcal{D})$ as the set of complex sequences $\mathbf{c} = \{c_I\}_{I \in \mathcal{D}}$ such that

$$\|\mathbf{s}\|_{\mathfrak{f}^s_{p,r}} = \Big\| \Big[\sum_{I \in \mathcal{D}} \big(|I|^{-\frac{s}{d} - \frac{1}{2}} |c_I| \chi_I(\cdot) \big)^r \Big]^{1/r} \Big\|_{L_p} < \infty , \tag{8.4}$$

so that

$$|f|_{F^s_{p,r}} \approx \sum_{\ell=1}^{L} \|(\langle f \, , \, \psi^\ell_I \rangle)\|_{\mathfrak{f}^s_{p,r}} . \tag{8.5}$$

We present now some of the results about N-term approximation for the spaces $\mathfrak{f}^s_{p,r}$ which appear in [9].

Theorem 8.6. Jackson's inequality
Let $s, \alpha \in \mathbb{R}$, $0 < p, \tau, q < \infty$ and $0 < r \leq \infty$ be so that

$$\max\{\tau, q\} < p \quad and \quad \frac{\alpha}{d} - \frac{1}{\tau} = \frac{s}{d} - \frac{1}{p}. \tag{8.7}$$

Then, for all $\mathbf{s} \in \mathfrak{b}^\alpha_{\tau,q}$ we have,

$$\sigma(\mathbf{s}, \Sigma_N)_{\mathfrak{f}^s_{p,r}} \leq C N^{-(\frac{1}{\tau \vee q} - \frac{1}{p})} \|\mathbf{s}\|_{\mathfrak{b}^\alpha_{\tau,q}} , \quad N = 1, 2, 3, \ldots \tag{8.8}$$

For the proof of this result see Theorem 4.3 in [9]. We remark that if an inequality of the type

$$\sigma(\mathbf{s}, \Sigma_N)_{\mathfrak{f}^s_{p,r}} \leq C N^{-\varepsilon} \|\mathbf{s}\|_{\mathfrak{b}^\alpha_{\tau,q}} , \quad N = 1, 2, 3, \ldots \tag{8.9}$$

holds, then the restrictions on the indices in (8.7) must necessarily hold and moreover $\varepsilon \leq \frac{1}{\tau \vee q} - \frac{1}{p}$. In this sense, we say that Theorem 8.6 is sharp (see Propositions 4.1 and 4.2 in [9]).

Theorem 8.10. Bernstein's inequality
Let $s, \alpha \in \mathbb{R}$, $0 < p, \tau, q < \infty$ and $0 < r \leq \infty$ be so that

$$\min\{\tau, q\} < p \quad and \quad \frac{\alpha}{d} - \frac{1}{\tau} = \frac{s}{d} - \frac{1}{p}. \tag{8.11}$$

Then, there exists C, $0 < C < \infty$ so that

$$\|\mathbf{s}\|_{\mathfrak{b}^\alpha_{\tau,q}} \leq C N^{(\frac{1}{\tau \wedge q} - \frac{1}{p})} \|s\|_{\mathfrak{f}^s_{p,r}} \tag{8.12}$$

for all $\mathbf{s} \in \Sigma_N$ and all $N = 1, 2, 3, \ldots$

We refer to section 5 in [9] for a proof of this theorem and similar examples illustrating the sharpness of the exponents.

Using these Jackson and Bernstein type inequalities for the spaces $\mathfrak{f}^s_{p,r}$ it is possible to obtain an analogous result to the one stated in Proposition 7.13, replacing $\mathfrak{h}^0_p$ by $\mathfrak{f}^s_{p,r}$. We omit the statement since actually a sharper version appears in [9, §6].

Let us finally mention that the decay of $\sigma(\mathbf{s},\Sigma_N)_{\mathfrak{f}^s_{p,r}}$ can be precisely quantified when one uses ℓ^q norms. Namely, given a metric space $\mathcal{F}$ we can define the *approximation spaces of order* $\gamma>0$ *and* $0<q\le\infty$, $A^\gamma_q(\mathcal{F})$, by

$$A^\gamma_q(\mathcal{F}) = \Big\{ \mathbf{s}\in\mathcal{F} \ : \ |\mathbf{s}|_{A^\gamma_q(\mathcal{F})} := \big(\sum_{N=1}^\infty [N^\gamma\sigma(\mathbf{s},\Sigma_N)_{\mathcal{F}}]^q \frac{1}{N}\big)^{\frac{1}{q}} < \infty \Big\}$$

Then, there is a well known procedure to derive identities for $A^\gamma_q(\mathcal{F})$ from inequalities of Jackson and Bernstein type as in Theorems 8.6 and 8.10 (for $\tau=q$) [8]. We refer to the recent paper [13] for a simple restatement of this general procedure which does not make use of real interpolation. As a consequence of these considerations one obtains

Corollary 8.13.

$$A^\gamma_q(\mathfrak{f}^s_{p,r}) = \mathfrak{b}^{s+\gamma}_{q,q}, \quad \textit{when} \quad \frac{\gamma}{d} = \frac{1}{\tau}-\frac{1}{p} > 0\,.$$

We refer to section 6 in [9] for more details in these types of identities.

Remark 1. Since $\mathfrak{b}^s_{p,p}=\mathfrak{f}^s_{p,p}$, the results stated in this section for Triebel-Lizorkin sequence spaces give results for non-linear approximation in the Besov space $B^s_{p,p}$.

Remark 2. The main ingredient in the proof of Theorems 8.6 and 8.10 is the fact that for the sequence space $\mathfrak{f}^s_{p,r}$ it holds a similar estimate to (7.5). More precisely, a sequence space $\mathfrak{f}$ with a quasi-norm $\|\ \|_{\mathfrak{f}}$ is called a *p-space* if there exits $c>0$ such that

$$c^{-1}\,(Card\,\Gamma)^{\frac{1}{p}} \ \le\ \Big\|\sum_{I\in\Gamma}\frac{\mathbf{e}_I}{\|\mathbf{e}_I\|_{\mathfrak{f}}}\Big\|_{\mathfrak{f}} \ \le\ c\,(Card\,\Gamma)^{\frac{1}{p}}\,.$$

for all finite sets $\Gamma\subset\mathcal{D}$. For a proof of the fact that $\mathfrak{f}^s_{p,r}$ are p-spaces see Proposition 3.2 in [9]. Jackson's and Bernstein's inequalities can be obtained similarly for p-spaces (see [9] and [13] for details).

References

[1] M. Barnsley and L. Hurd, *Fractal image compression.* AK Peters Ltd. Wellesley, Massachusetts (1993).

[2] G. Bourdaud, "Ondelettes et espaces de Besov", Rev. Mat. Iberoamericana, Vol. 11, No. 3, (1995), 477-512.

[3] E. W. Cheney, "Introduction to Approximation Theory", McGraw-Hill, 1966.

[4] A. Chambolle, R. DeVore, Nam-Yong Lee and B.J. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage". IEEE Trans. Image Processing **7** (3), 319-335, March 1998. Available at *http://www.math.purdue.edu/∼lucier.*

[5] A. Cohen, "Wavelet methods in numerical analysis". *Handbook of numerical analysis, Vol. VII*, 417–711. North-Holland, Amsterdam, 2000.

[6] A. Cohen, R. DeVore, P. Petrushev and H. Xu, "Nonlinear approximation and the space BV(R^2)". Amer. J. Math. **121** (1999), no. 3, 587–628.

[7] R.A. DeVore, *Nonlinear Approximation.* In *Acta Numerica* **7**, Cambridge University Press, 1998, 51-151.

[8] R.A. DeVore and V. Popov, "Interpolation spaces and nonlinear approximation", Function Spaces and Applications (Lund, 1986), Lecture Notes in Math., vol. 1302, Springer-Verlag, Berlin, 1988, 191-215.

[9] G. Garrigós and E. Hernández, "Sharp Jackson and Bernstein Inequalities for N-term Approximation in Sequence Spaces with Applications", Ind. Univ. Math. Journal, **53** (6) (2004), 1741-1764. Available at *http://www.uam.es/eugenio.hernandez*.

[10] G. Garrigós, R. Hochmuth and A. Tabacco, "Wavelet characterizations for anisotropic Besov spaces when $0 < p < 1$". Proc. Edin. Math. Soc. 47, (2004), 573-595. Available at *http://www.uam.es/gustavo.garrigos*.

[11] E. Hernández and G. Weiss, *A First Course on Wavelets*, CRC Press, Boca Raton, FL, (1996).

[12] C. Hsiao, B. Jawerth, B.J. Lucier and X.M. Yu, *Near optimal compression of almost optimal wavelet expansions*. Wavelets: mathematics and applications, 425–446, Stud. Adv. Math., CRC, Boca Raton, FL, 1994.

[13] G. Kerkyacharian and D. Picard, "Entropy, universal coding, approximation, and bases properties", Constr. Approx. 20 (2004), 1-37.

[14] G. Kyriazis, "Multilevel characterization of anisotropic function spaces", SIAM J. Math. Anal. 36, (2004), 441-462.

[15] S. Mallat, *A wavelet tour of signal processing*. Academic Press Inc., San Diego, Second Edition 1999.

[16] Y. Meyer, "Ondelettes". Hermann, 1990.

[17] Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equations*. The fifteenth Dean J. B. Lewis memorial lectures. University Lecture Series **22**. AMS, Providence (RI), 2001.

[18] J. Peetre, "New thoughts on Besov spaces", Math. Dept. Duke University, 1976.

[19] V. Temlyakov, "The best m-term approximation and greedy algorithms". Adv. Comput. Math. **8** (3), (1998), 249–265.

[20] P. Wojtaszczyk, "Wavelets as unconditional bases in $L_p(R)$". J. Fourier Anal. Appl. **5** (1999), no. 1, 73–85.

Departamento de Matemáticas C-XV, Universidad Autónoma de Madrid, 28049 Madrid, Spain.

E-mail address: `gustavo.garrigos@uam.es, eugenio.hernandez@uam.es`