WAVELET METHODS IN IMAGE COMPRESSION

GUSTAVO GARRIGÓS [*]

ABSTRACT. We present in printed form the contents of a survey lecture given at the *International Workshop in Classical Analysis and Applications* held in Yaoundé in December 2001. We intend to introduce, for a general public of mathematicians and engineers, several approaches to image compression based on linear and non-linear approximation. We compare Fourier and wavelet-based methods, and present some of the mathematics behind them. In particular we show how Sobolev and Besov spaces arise naturally from this theory. The results are known in the literature, so we have selected a suitable bibliography for further reading into the subject.

1. Introduction

In this lecture we shall consider the following general problem arising in signal processing.

The signal compression problem. *Given a signal f, which typically belongs to a Hilbert space $\mathcal{H}$, find an "approximated representation" f_M of f with the following three conditions:*

1. *The signal f_M is given by a fixed number of coefficients M;*
2. *There is a known control of the error: $\varepsilon[M] = \|f - f_M\|$;*
3. *There is a fast algorithm to produce f_M.*

In applications, f_M can be seen as a "compressed version" of the original signal f, 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.

Some practical situations where this problem arises are, for instance, the coding of sound signals for cellular telephones or music for CD's, digitalizing pictures to store in computers, or compressing video sequences for real time transmission via Internet. In these cases, f could correspond to an electric

[*] The author thanks the people and institutions promoting and supporting the workshop, in particular, the Ministry of Research of Cameroon, the University of Yaoundé, CIMPA, and Universidad Autónoma de Madrid. Research partially supported by the European Network "HARP 2002-2006", and the Spanish projects "BMF2001-0189" and *Programa Ramón y Cajal*.

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

Key words and phrases: Non-linear approximation, wavelet, Besov space.

voltage (associated, e.g., with an acoustic pressure) varying in time $t > 0$, or to a light intensity field which varies at every point $(x, y) \in \mathbf{R}^2$. In mathematical problems, f could also be the (unknown) solution to a PDE and f_M an approximation by a certain numerical method. In all these cases it is important that the compressed signal f_M 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.

In this lecture we shall present two approaches to this problem based on the orthogonal structure of Hilbert spaces: the so-called *linear and non-linear approximation methods*. We shall describe these methods with Fourier and wavelet bases, and see the different roles played in each case by Sobolev and Besov spaces. We are mostly following the references [5, 1], with scattered results from other sources. We also refer to [6] for a much wider perspective on the mathematics underlying image processing.

2. Mathematical setting

A natural mathematical model for the previous problem considers the signal f as an element of a Hilbert space $\mathcal{H}$. Typically, $\mathcal{H}$ is taken to be $L^2(\mathbf{R}^d)$, $L^2[0, N]^d$ or $\ell^2(N^d)$, depending on the kind of application. For instance, images from real life may be seen as functions $f(x, y)$ which correspond to light intensity fields in $[0, 1]^2$. We explain briefly how these can be discretized to be manipulated and stored in computers (we follow [1, p. 324]).

Example 2.1. : A mathematical model for images.

First one notices that the intensity function $f(x, y)$ associated with an image cannot in general be measured at *all points* $0 \le x, y \le 1$. In practice, a measuring device (a photometer) averages the light intensity over small squares distributed dyadically along the picture frame $[0, 1]^2$. So if N is large (typically, $N \ge 8$), we can codify the image as a sequence of 2^{2N} coefficients:

$$p_{\mathbf{k}} = p_{\mathbf{k}}^{(N)} = \frac{1}{|I_{N,\mathbf{k}}|} \int\int_{I_{N,\mathbf{k}}} f(x, y)\, dxdy, \quad 0 \le k_1, k_2 < 2^N, \tag{2.2}$$

where $I_{N,\mathbf{k}}$ denotes the dyadic square $[\frac{k_1}{2^N}, \frac{k_1+1}{2^N}] \times [\frac{k_2}{2^N}, \frac{k_2+1}{2^N}]$. These squares are usually called *pixels* (or *picture elements*) *located at position* $2^{-N}\mathbf{k}$, and correspond in practice to the number of "dots" that form a computer screen. To each of them we must associate a single number $p_{\mathbf{k}}$ (typically between 0 and 2^8), which represents the "grey level" of the picture at that point. In this way we have converted the image into a sequence of "bits" which can be stored and processed by computers.

For the theoretical model we still wish to see the image as a concrete function corresponding to the collection of observed pixels. One does this by considering a sequence $\{\phi_{N,\mathbf{k}}\}$ of real functions supported in $I_{N,\mathbf{k}}$, and constructing the so-called *observed image*:

$$f^{(o)}(\mathbf{x}) = \sum_{\mathbf{k}} p_{\mathbf{k}}\, \phi_{N,\mathbf{k}}(\mathbf{x}). \tag{2.3}$$

Typically, $\phi_{N,\mathbf{k}}(\mathbf{x}) = \phi(2^N\mathbf{x} - \mathbf{k})$, for a fixed function ϕ, which can be chosen simply as $\chi_{[0,1]^2}$, or replaced by smoother versions such as splines or "scaling functions". We observe that for large N, $f^{(o)}$ is an almost indistinguishable copy of f, and thus can be identified for mathematical purposes with the original image. The compression problem then consists in representing $f^{(o)}$ with a much smaller amount of coefficients without loosing the visual resemblance with the original f.

Continuing with the abstract framework, let $\mathcal{H}$ be a Hilbert space and $\{e_j\}_{j=1}^{\infty}$ a fixed *orthonormal basis.* Then, every $f \in \mathcal{H}$ can be expanded as $f = \sum_{j=1}^{\infty} c_j e_j$, where $c_j = \langle f, e_j\rangle$ are the *basis coefficients.* There are two standard ways of constructing an approximating signal f_M, by keeping only M basis coefficients:

1. *The linear approximation method.* Corresponds to minimizing the functional

$$\|f - \sum_{j=1}^{M} a_j e_j\|, \quad \text{over all} \quad a_1, \dots, a_M \in \mathbf{C}.$$

For an orthonormal basis $\{e_j\}$, the minimizer is given by $f_M = \sum_{j=1}^{M} \langle f, e_j\rangle e_j$, i.e., f_M is the orthogonal projection of f over $V_M := \mathrm{span}\{e_1, \dots, e_M\}$.

2. *The non-linear approximation method.* Corresponds to minimizing

$$\|f - \sum_{\lambda\in\Lambda} a_\lambda e_\lambda\|, \quad \text{over all} \quad a_\lambda \in \mathbf{C}$$

and all sets of indices $\Lambda \subset \mathbf{Z}_+$ such that $\mathrm{Card}\,\Lambda \le M$. When $\{e_j\}$ is an orthonormal basis, the minimizer is given by $f_M = \sum_{k=1}^{M} \langle f, e_{j_k}\rangle e_{j_k}$, where we have rearranged non-increasingly the coefficients

$$|\langle\, f\,,\, e_{j_1}\,\rangle| \ge |\langle\, f\,,\, e_{j_2}\,\rangle| \ge \dots$$

The three main questions which one poses at this point are the following:

A. Find "natural bases" to represent a typical applied signal, with the property that the approximation error is as small as possible.
B. Characterize in such bases the set of all signals with a given rate of approximation, say $\varepsilon[M] = O(M^{-\alpha})$.
C. Find efficient algorithms to compute numerically f_M, or store it with minimal information.

We shall discuss about the first two problems in the next sections, referring for the last one to the specialized literature. More precisely, we present the situation with Fourier and wavelet bases, compare their different roles, and derive the function spaces naturally related to question (B).

3. Linear Fourier approximation and Sobolev spaces

The results in this section are taken from [5, Ch.9.1]. To make simpler the exposition, we assume $d = 1$. Consider the Hilbert space $\mathcal{H} = L^2[0, N]$, and its associated Fourier basis

$$e_m(t) = \tfrac{1}{\sqrt{N}}\, e^{2\pi i m t/N}, \quad m \in \mathbf{Z}.$$

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

$$f(t) = \sum_{m\in\mathbf{Z}} \langle\, f \,,\, e_m \,\rangle\, e_m(t) = \sum_{m\in\mathbf{Z}} \tfrac{1}{\sqrt{N}}\, \hat{f}(\tfrac{2\pi m}{N})\, e_m(t),$$

with convergence in the L^2-sense. The last identity makes sense when we extend f to be zero outside $[0, N]$ and denote the *Fourier transform* by:

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

The *best linear approximation for* f with M coefficients is given by

$$f_M(t) = \sum_{|m|\leq M} \tfrac{1}{\sqrt{N}}\, \hat{f}(\tfrac{2\pi m}{N})\, e_m(t),$$

and the error of approximation by

$$\varepsilon[M] = \|f - f_M\| = \left(\frac{1}{N} \sum_{|m|>M} \left| \hat{f}(\tfrac{2\pi m}{N}) \right|^2 \right)^{\frac{1}{2}}. \tag{3.1}$$

We concentrate now in problem (B) from the previous section. That is, *characterize the set of all signals* $f \in L^2[0, N]$ *for which the error of approximation decays like* $\varepsilon[M] = O(M^{-\alpha})$. Taking into account the expression for the error in (3.1), we see that the smallness of $\varepsilon[M]$ is closely related to the fast decay of $\hat{f}(\omega)$ when $|\omega| \to \infty$. This asymptotic decay is in turn equivalent to regularity of the original f, due to the well-known identity

$$\widehat{f^{(k)}}(\omega) = \int_{-\infty}^{\infty} f^{(k)}(t)\, e^{-i\omega t}\, dt = (i\omega)^k\, \hat{f}(\omega). \tag{3.2}$$

These facts lead to a natural candidate for the study of Linear Fourier Approximation: the family of *Sobolev spaces*

$$H^\alpha(\mathbf{R}) = \left\{ f \in L^2(\mathbf{R}) \quad : \quad \int_{\mathbf{R}} |\omega|^{2\alpha} |\hat{f}(\omega)|^2\, d\omega < \infty \right\}.$$

The next theorem shows that this is indeed the case, providing a characterization in terms of the error, not exactly of the form $O(M^{-\alpha})$, but with an ℓ^2-variant of it.

THEOREM 3.3. : Linear Approximation spaces.

If $\alpha > 0$ and $f \in L^2_c(0,N)$, then

$$f \in H^\alpha(\mathbf{R}) \quad \Longleftrightarrow \quad \sum_{M=1}^{\infty} M^{2\alpha-1}\varepsilon[M]^2 < \infty.$$

In this case, $\varepsilon[M] = o(M^{-\alpha})$.

PROOF:

The idea of the proof is very simple, so we sketch it here:

$$\begin{aligned}
\sum_{M=1}^{\infty} M^{2\alpha-1}\varepsilon[M]^2 &= \sum_{M=1}^{\infty} M^{2\alpha-1}\Big(\frac{1}{N}\sum_{|m|>M}\Big|\hat{f}(\tfrac{2\pi m}{N})\Big|^2\Big) \\
&= \frac{1}{N}\sum_{m\in\mathbf{Z}}\Big(\sum_{M=1}^{|m|} M^{2\alpha-1}\Big)\Big|\hat{f}(\tfrac{2\pi m}{N})\Big|^2 \\
&= \frac{c_\alpha}{N}\sum_{m\in\mathbf{Z}}|m|^{2\alpha}\Big|\hat{f}(\tfrac{2\pi m}{N})\Big|^2 \sim \int_{\mathbf{R}}|\omega|^{2\alpha}|\hat{f}(\omega)|^2\,d\omega,
\end{aligned}$$

where the last step can be roughly justified by looking at the integral as a Riemann sum. We observe that a rigorous argument for this step requires some more effort, making essential use of the compact support condition for f in the statement of the theorem. □

We can conclude from the previous theorem that Linear Fourier Approximation is a good method to analyze signals with uniform smoothness at all points $t \in \mathbf{R}$. These can be coded using Fourier coefficients, and may be easily handled with the precise expression we have obtained for $\varepsilon[M]$. Observe that looking at the decay rate of the error of approximation we can estimate numerically the best α for which $f \in H^\alpha$. 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 $\mathbf{R}$ [5, 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. In fact, the simple example of the characteristic function of an interval $f = \chi_{[0,1/2]}$, which belongs to H^α for all $\alpha < \frac{1}{2}$, has an error of linear approximation of the order $\varepsilon[M] \sim M^{-1/2}$. Figure 9.1 in [5, p. 381] shows that this error is due to Gibbs oscillations near discontinuities, providing a very "fuzzy" representation of the original f. As we shall see below the representation of such signals can be largely improved by using non-linear approximation and wavelet bases.

4. Wavelet bases and local regularity

Before passing to a more modern approach to the problem in terms of *wavelet bases*, we spend some time describing the main features of these. We say that a function $\psi \in L^2(\mathbf{R})$ is an *orthonormal wavelet* whenever the system formed by translating and dilating this function

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k), \quad j,k \in \mathbf{Z}, \tag{4.1}$$

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

Example 4.2. 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*:

$$\operatorname{Supp}\psi_{j,k} \subset \left[\tfrac{k}{2^j}, \tfrac{k+1}{2^j}\right] \quad \text{and} \quad \int_{-\infty}^{\infty}\psi_{j,k}(t)dt = \int_{-\infty}^{\infty}\psi(t)dt = 0. \tag{4.3}$$

It is easy to verify from these two facts that $\{\psi_{j,k}\}$ is actually an orthonormal basis of $L^2(\mathbf{R})$. Indeed, the orthonormality follows from the nesting property of dyadic intervals, while the completeness is a consequence of the L^2-density of the set

$$\left\{f = \textstyle\sum_{\text{finite}} a_I\chi_I \;:\; I \text{ dyadic }, \int f = 0\right\}.$$

This elementary example already illustrates the "zoom property" that makes of wavelet bases *excellent detectors of local singularities*. Roughly speaking, $2^{\frac{j}{2}}\langle f, \psi_{j,k}\rangle$ 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 singularities. We recall that singularities carry essential information of signals in many applied problems, such as the presence of edges in images. This makes of 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 presented below. The statement is a simplified version of Theorem 9.7 in the first edition of [5].

Theorem 4.4. : Wavelet characterization of local smoothness.

1. Let $f \in L^2(\mathbf{R})$ be a signal with local smoothness $Lip_\alpha(t_0)$, that is

$$|f(t) - f(t_0)| \leq C_{t_0} \, |t - t_0|^\alpha, \quad t \in \mathbf{R}. \tag{4.5}$$

Then, the wavelet coefficients decay as:

$$|\langle \, f \, , \, \psi_{j,k} \, \rangle| \, \lesssim \, 2^{-j(\alpha + \frac{1}{2})} \, (1 + |2^j t_0 - k|^\alpha). \tag{4.6}$$

2. Conversely, if for some $\varepsilon > 0$ it holds

$$|\langle \, f \, , \, \psi_{j,k} \, \rangle| \, \lesssim \, 2^{-j(\alpha + \frac{1}{2})} \, (1 + |2^j t_0 - k|^{\alpha - \varepsilon}),$$

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:

$$\begin{aligned}
|\langle \, f \, , \, \psi_{j,k} \, \rangle| &= \Big| \int (f(t) - f(t_0)) \, \psi_{j,k}(t) \, dt \Big| \\
&\lesssim \int |t - t_0|^\alpha \, |\psi_{j,k}(t)| \, dt \\
&= 2^{-\frac{j}{2}} \int |2^{-j} t + 2^{-j} k - t_0|^\alpha \, |\psi(t)| \, dt \\
&\lesssim 2^{-j(\frac{1}{2} + \alpha)} \textstyle\int |t|^\alpha \, |\psi(t)| \, dt \; + \; 2^{-\frac{j}{2}} \, |2^{-j} k - t_0|^\alpha.
\end{aligned}$$

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 (4.5) (with a Taylor polynomial, rather than $f(t_0)$) and more moment conditions in the wavelet

$$\int_{-\infty}^{\infty} t^\ell \, \psi(t) \, dt = 0, \quad \ell = 0, 1, \ldots, n - 1.$$

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 [5, p. 173]). □

We point out that the construction of wavelet bases with good properties for applications is still a vast field of research. Polynomial spline wavelets generalizing the Haar case were introduced by Battle and Lemarié, but they failed to have compact support. The now most popular Daubechies' wavelets exist for all smoothness degrees and have minimal compact support. Most important for applications are the wavelet bases in $L^2[0, N]$, formed by restricting $L^2(\mathbf{R})$ wavelets to $[0, N]$, and adding some special "boundary wavelets" which keep the vanishing moments at the extreme points $t = 0, N$. The construction and numerical implementation of algorithms with wavelet bases owes much to the contribution of S. Mallat and his *Multiresolution Analysis*. For more detailed discussion on all this part we refer to [5, Ch. 7], or to the text [4].

Finally, we conclude this section with a note on *wavelet bases in d-dimensions*. These are defined as unions of systems

$$\psi_{j,\mathbf{k}}(\mathbf{x}) = 2^{\frac{jd}{2}}\psi(2^j\mathbf{x}-\mathbf{k}), \quad j\in\mathbf{Z},\ \mathbf{k}\in\mathbf{Z}^d,$$

with the particularity that for an orthonormal basis in $L^2(\mathbf{R}^d)$ one needs exactly 2^d-1 such functions $\psi\in L^2(\mathbf{R})$ (at least if we want them compactly supported). In practice, wavelets in d-dimensions are constructed as tensor products of 1-dimensional wavelets, and the numerical algorithms are very similar to these ones (see [5, Ch. 7.7]). As an exercise, the reader can amuse himself with the construction of a 2-dimensional Haar wavelet.

5. Non-linear wavelet approximation and Besov spaces

We come back to the problem of signal compression, this time using wavelet rather than Fourier bases, and concentrating in the non-linear approach (linear wavelet approximation produces similar results as Fourier methods, see [5, §9.1]). We let $\mathcal{H}=L^2(\mathbf{R})$, and consider an orthonormal basis of the form $\{\psi_{j,k}\}$ as in (4.1) above. Our goal is to profit from the "zoom property" of wavelets in order to reproduce the essential singularities of f in the compressed image f_M. Remember that singularities produce large wavelet coefficients, so we can expect a considerable improvement by using non-linear approximation.

Recall from §2 that, for a signal $f\in L^2(\mathbf{R})$, the non-linear approximation with M coefficients is given by

$$f_M = \sum_{\lambda\in\Lambda_M}\langle f\,,\,\psi_\lambda\rangle\psi_\lambda,$$

where $\Lambda_M=\big\{\lambda_1=(j_1,k_1),\ \ldots\ ,\lambda_M=(j_M,k_M)\big\}$ and we have sorted the coefficients non-increasingly:

$$\big|\langle f\,,\,\psi_{j_1,k_1}\rangle\big| \geq \big|\langle f\,,\,\psi_{j_2,k_2}\rangle\big| \geq \ldots$$

The error of approximation now takes the form

$$\varepsilon[M] = \|f-f_M\| = \Big(\sum_{\lambda\notin\Lambda_M}|\langle f\,,\,\psi_\lambda\rangle|^2\Big)^{\frac12}. \tag{5.1}$$

This quantity is certainly smaller than the linear error, but we must still see whether it produces a significant improvement. We wish to characterize the sets of signals for which $\varepsilon[M]\sim O(M^{-\alpha})$, and compare them with the corresponding ones in the linear method. Observe that the candidate spaces cannot correspond to fast asymptotic decays of $|\langle f\,,\,\psi_\lambda\rangle|$, when $|\lambda|\to\infty$, since non-linear errors *do not see the position of coefficients*. We should instead measure *sizes*, for which better quantifiers are the ℓ^p-norms: $\sum_\lambda |\langle f, \psi_\lambda \rangle|^p$, in our case with $0 < p < 2$. Indeed, let us define the following spaces:

$$\mathcal{B}_p = \mathcal{B}_p^{(\psi)} = \Big\{ f \in L^2(\mathbf{R}) \ : \ \|f\|_{\mathcal{B}_p} = \big(\sum_\lambda |\langle f, \psi_\lambda \rangle|^p \big)^{\frac{1}{p}} < \infty \Big\}. \tag{5.2}$$

Then, the following theorem tells us that these are the "right candidates" for non-linear approximation spaces.

Theorem 5.3. Non-linear approximation spaces.

1. If $f \in \mathcal{B}_p$ and $0 < p < 2$, then

$$\varepsilon[M] \lesssim \frac{\|f\|_{\mathcal{B}_p}}{M^{\frac{1}{p}-\frac{1}{2}}}. \tag{5.4}$$

2. Conversely, if $\varepsilon[M] = O(M^{-(\frac{1}{p}-\frac{1}{2})})$, then $f \in \mathcal{B}_{p+\varepsilon}$, for all $\varepsilon > 0$.

Proof: The proof is quite elementary and illustrates the link between non-increasing rearrangements and ℓ^p-norms. For the first part just observe that

$$\begin{aligned} k\,|\langle f, \psi_{\lambda_k} \rangle|^p &= |\langle f, \psi_{\lambda_k} \rangle|^p + \ \dots \ + |\langle f, \psi_{\lambda_k} \rangle|^p \\ &\leq |\langle f, \psi_{\lambda_1} \rangle|^p + \ \dots \ + |\langle f, \psi_{\lambda_k} \rangle|^p, \end{aligned}$$

and therefore $|\langle f, \psi_{\lambda_k} \rangle| \leq \|f\|_{\mathcal{B}_p} k^{-\frac{1}{p}}$. Substituting this estimate in the error (5.1) and summing the series leads directly to (5.4). The converse is analogous, since from $\varepsilon[M] = O(M^{-(\frac{1}{p}-\frac{1}{2})})$ it follows

$$M^{-(\frac{1}{p}-\frac{1}{2})} \gtrsim \big(\sum_{M<k\leq 2M} |\langle f, \psi_{\lambda_k} \rangle|^2 \big)^{\frac{1}{2}} \geq M^{\frac{1}{2}}\,|\langle f, \psi_{\lambda_{2M}} \rangle|.$$

Thus $|\langle f, \psi_{\lambda_{2M}} \rangle| \lesssim M^{-\frac{1}{p}}$, implying that the coefficients $\{\langle f, \psi_\lambda \rangle\}$ belong to $\ell^{p+\varepsilon}$, for all $\varepsilon > 0$. $\square$

The reader should observe that, so far, the previous construction of $\mathcal{B}_p$ applies to any orthonormal basis $\{\psi_\lambda\}$ in $\mathcal{H}$. The difficult question now becomes, for each such basis, *to identify $\mathcal{B}_p$ among the classical families of smoothness spaces in $\mathbf{R}^d$*. Observe that this is already a non-trivial question for the Fourier basis, where the only extensively studied case seems to be the *Wiener algebra*, corresponding to $p = 1$. This is however typically formed by nowhere Lipschitz functions (such as $\sum_k k^{-2} e^{i2^k t}$), so it is not a good model for signals arising in applications.

With wavelet bases the situation is surprisingly much better, since the associated $\mathcal{B}_p$'s *coincide with a subfamily of the classical Besov smoothness spaces.* Besov spaces were introduced in the 60's as generalizations of Lipschitz spaces, where the smoothness is measured with L^p-versions of the *modulus of continuity.* More precisely, the Besov space $B^\alpha_q(L^p)$ is defined as the subspace of

$L^p(\mathbf{R}^d)$ such that

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

The characterization of Besov spaces in terms of wavelets was first obtained by Y. Meyer in the mid 80's, and requires deeper techniques from Harmonic Analysis. The theorem we state below, characterizing $B^\alpha_p(L^p)$ in terms of $\mathcal{B}_p$ for indices of integrability p below 1, was independently obtained by DeVore and Popov using techniques from Approximation Theory. We refer to [2, § 30], and also to work of the author [3], where some extensions to higher dimensions are given.

Theorem 5.5. *Let $0 < p < 2$ and α defined by $\frac{\alpha}{d} = \frac{1}{p} - \frac{1}{2}$. Then, for all sufficiently "regular" wavelet bases in $L^2(\mathbf{R}^d)$ the approximation space $\mathcal{B}_p$ coincides with the Besov space $B^\alpha_p(L^p)$, with equivalence of norms*

$$\|f\|_{\mathcal{B}_p} \lesssim \|f\|_p + |f|_{B^\alpha_p(L^p)} \lesssim \|f\|_{\mathcal{B}_p}.$$

The previous theorem is a nice theoretical result, which would not be of much use if $\mathcal{B}_p$ did not contain the typical signals arising in applications. Remember this is exactly the problem we had with Sobolev classes H^α, whose uniform way of measuring smoothness excludes many discontinuous signals (such as images) except for very low α's. This produces low compression rates, which make the linear approximation method not very suitable for image processing. The situation turns out to be much better with Besov spaces due to the "zooming property" of wavelets. The discontinuities of signals are concentrated in few wavelet coefficients, which do not affect the ℓ^p-summability of the total series, and hence the total Besov smoothness. A precise statement of this is given in our last theorem, which has been taken from §9.2.2 in the first edition of [5].

Theorem 5.6. : Piecewise regularity.

1. Let $t_0 = 0 < t_1 < \ldots < t_r = N$, be a finite partition of $[0, N]$. If $f \in L^\infty[0, N]$ belongs to Lip $_\alpha(t_{i-1}, t_i)$, for all $i = 1, \ldots, r$, then

$$f \in \mathcal{B}_p, \quad \textit{for all } p > \frac{1}{\alpha + \frac{1}{2}}.$$

2. Let $[0, N]^d = \Omega_1 \cup \ldots \cup \Omega_r$ be a finite partition of a d-dimensional cube with domains Ω_i having a "regular" border. If $f \in L^\infty[0, N]^d$ belongs to Lip $_\alpha(\Omega_i)$, for all $i = 1, \ldots, r$, then

$$f \in \mathcal{B}_p, \quad \textit{for all } p > \max\Big\{ \frac{2(d-1)}{d}, \frac{1}{\frac{\alpha}{d} + \frac{1}{2}} \Big\}.$$

Proof: The theorem can be proved directly from the "smoothness characterization" of $\mathcal{B}_p$ in Theorem 5.5. A more illustrative proof in 1-dimension, which shows the role of wavelet coefficients, is as follows. For "good points" around which f is Lipschitz, we can bound the coefficients by

$$|\langle f, \psi_{j,k} \rangle| \lesssim 2^{-j(\alpha+\frac{1}{2})}, \tag{5.7}$$

as in Theorem 4.4. For "bad points" the most we can say is

$$|\langle f, \psi_{j,k} \rangle| \lesssim 2^{-\frac{j}{2}}.$$

Now, at each resolution level j, there are at most a constant number of dyadic intervals $I_{j,k}$ which contain "bad points". Thus, the bad part of the ℓ^p-series defining $\mathcal{B}_p$ gives:

$$\mathcal{S}_B = \sum_{j=0}^{\infty} \sum_{k \text{ finite}} |\langle f, \psi_{j,k} \rangle|^p \lesssim \sum_{j=0}^{\infty} 2^{-\frac{jp}{2}} < \infty.$$

For the rest of coefficients we use the "good estimate" in (5.7) to obtain

$$\mathcal{S}_G = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j N} |\langle f, \psi_{j,k} \rangle|^p \lesssim \sum_{j=0}^{\infty} 2^j N \, 2^{-j(\alpha+\frac{1}{2})p},$$

which is finite exactly when $p > \frac{1}{\alpha+\frac{1}{2}}$. The d-dimensional case follows similarly, except that the number of "bad coefficients" at each resolution level j is now of the order of $2^{j(d-1)}$. Observe that to have such a control for "bad coefficients" we need to assume some "regularity" of the borders $\partial\Omega_i$, avoiding domains with fractal behavior. □

6. Conclusions and further results

We have presented a very brief sketch of the mathematics behind two natural methods for image processing, so it is time to compare practically the results derived with each of them. As explained in §2, we shall regard an image as a function $f \in L^2[0,1]^2$. We also know from §2 that linear approximation keeps the *first M elements* in the basis expansion. When using wavelets, which remember are supported in small dyadic squares $I_{j,\mathbf{k}}$ over $[0,1]^2$, it is natural to take $M = 2^{2m}$, since we want the same resolution over the whole image. Thus the *best linear approximant* is:

$$f_M = \sum_{j=0}^{m} \sum_{0 \le |\mathbf{k}| < 2^j} \langle f, \psi_{j,\mathbf{k}} \rangle \, \psi_{j,\mathbf{k}},$$

from which 2-dimensional versions of §3 produce a *linear error*:

$$\|f - f_M\| \;\lesssim\; M^{-\frac{\alpha}{2}} \, \|f\|_{H^\alpha}, \quad \forall\, f \in H^\alpha.$$

From §5, the *best non-linear approximant*, $\tilde{f}_M$, produces instead:

$$\|f - \tilde{f}_M\| \;\lesssim\; M^{-\frac{\alpha}{2}}\,\|f\|_{\mathcal{B}_p}, \quad \forall\, f \in \mathcal{B}_p, \quad \text{if } \tfrac{\alpha}{2} = \tfrac{1}{p} - \tfrac{1}{2}.$$

The improvement of non-linear approximation over linear approximation can now be seen in two ways:

1. With the same rate of approximation $M^{-\frac{\alpha}{2}}$, there are more images in $\mathcal{B}_p = B_p^\alpha(L_p)$ than there are in H^α. In fact, when $\alpha > \frac{1}{2}$ the former contains discontinuous signals that cannot belong to the latter.

2. For a given image f, the best non-linear approximation rate is much faster than its linear counterpart. For instance, a typical image with border $\partial\Omega$, which we can think as belonging to $f \in \mathrm{Lip}_1(\mathbf{R}^2 \setminus \partial\Omega)$, will satisfy the best estimates

$$f \in H^\alpha, \quad \forall\, \alpha < \tfrac{1}{2} \;\Longrightarrow\; \varepsilon_{\text{LIN}}[M] \lesssim M^{-\frac{1}{4}+\varepsilon}$$

$$f \in \mathcal{B}_p, \quad \forall\, p > 1 \;\Longrightarrow\; \varepsilon_{\text{N-L}}[M] \lesssim M^{-\frac{1}{2}+\varepsilon}.$$

A well-known picture that illustrates this fact is called *Lena*, and can be seen in [5, Fig. 9.3]. Compressing by $\frac{1}{16}$ the original picture of 256^2 pixels, non-linear approximation reduces by one third the linear error, and produces in addition a more than acceptable reproduction of the original f. For other examples, in [1] the best Besov exponents from 24 pictures of the *Kodak Photo Sampler* are numerically estimated with non-linear errors, resulting in all cases within the range $0.3 \le \alpha \le 0.7$. Thus, it seems reasonable that typical smoothness lies below 1, which is important because of the d-dimensional constraints of Theorem 5.6. Unfortunately, due to these constraints higher smoothness does not necessarily improve the non-linear rate.

We do not wish to conclude this exposition without mentioning some further advances in this field which are under current investigation:

1. The study of *bounded variation spaces*, $BV(I)$, as larger classes of spaces with norms that better adapt to human visual perception. In 2-dimensions, they are related with classical Besov's by:

$$\mathcal{B}_1 = B_1^1(L_1) \;\subset\; BV[0,1]^2 \;\subset\; B_\infty^1(L_1),$$

with errors of approximation characterized by:

$$\varepsilon_{\text{LIN}}[M] = O(M^{-\frac{1}{4}}) \quad \text{ and } \quad \varepsilon_{\text{N-L}}[M] = O(M^{-\frac{1}{2}}).$$

The new space BV contains functions discontinuous along curves, while $\mathcal{B}_1$ does not. An excellent introduction to these spaces is given in the survey paper [6].

2. The study of *adaptive basis algorithms*, which optimize the non-linear approximation of a particular signal among a whole "dictionary of bases". One can also compare wavelet non-linear approximation with *spline adaptive*

approximation. The latter partitions the image with narrow triangles elongated along the edges, so as to minimize the overlap of the spline supports with the border lines (see [5, Fig. 9.6]). This produces, for functions such as $f = \chi_\Omega$, errors of approximation of the order $\varepsilon_{\rm spl}[M] = O(M^{-1})$, compared with $\varepsilon_{\rm N-L}[M] = O(M^{-\frac{1}{2}})$ using standard wavelet bases. The spline triangularization, however, has a high computational cost, so much research is employed trying to adapt wavelet bases to specific contours of images (*wedgelets*).

3. Finally, approximation techniques are also valid for the *estimation of signals in additive noises.* These models consider signals as random vectors $X = f + W$, where the (unknown) signal f is corrupted by a Gaussian noise. An approach for noise-removal consists in minimizing $E[\|X - f_M\|^2]$, with non-linear techniques similar to the ones presented above. We refer again to Chapter 10 of [5], or the survey paper [6] for a wider introduction on these questions.

References

[1] 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.*

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

[3] G. Garrigós, R. Hochmuth and A. Tabacco, "Wavelet characterizations for anisotropic Besov spaces: cases $0 < p < 1$". Preprint, 2001. Available at *http://www.uam.es/gustavo.garrigos.*

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

[5] S. Mallat, *A wavelet tour of signal processing.* Academic Press Inc., San Diego, Second Edition 1999. Figures available at *http://www.cmap.polytechnique.fr/∼mallat.*

[6] Y. Meyer, "Wavelets and functions with bounded variation from image processing to pure mathematics". Rend. Mat. Acc. Lincei s. 9, fasc. spec. 2000, 77–105.

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

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